TECHNOLOGIA

Traffic management and quality of experience

MC069

This document has been prepared for Ofcom

Jeremy Klein, Jonathan Freeman, Rob Morland and Stuart Revell
7 April 2011

Version 1

Contents

Abbreviations used

Abbreviation	Explanation
3G	3rd Generation Mobile
4G	4th Generation Mobile
ADSL	Asymmetric Digital Subscriber Line technology
ATM	Asynchronous Transfer Mode
BGP	Border Gateway Protocol
B-RAS	Broadband Remote Access Server
CAPEX	Capital expenditure
CATV	Cable Television
CDN	Content Distribution Network
CMTS	Cable Modem Termination System
CoS	Class of Service
DCKTN	Digital Communications Knowledge Transfer Network
DNS	Directory Name Server
DOCSIS	Data Over Cable Service Interface Specification
DPI	Deep Packet Inspection
DSL	Digital Subscriber Line
DSLAM	Digital Subscriber Line Access Multiplexer
DWDM	Dense Wavelength Division Multiplexing
FTP	File Transfer Protocol
GGSN	Gateway GPRS Support Node
GPRS	General packet radio service
GW	Gateway
HD	High Definition (Video)
HE	Head End
HLR	Home Location Register
HSPA	High Speed Packet Access
ICT	Information and Communication technologies
IGP	Internal Gateway Protocol
IP	Internet Protocol
IPTV	Internet Protocol television
ISO	International Standards Organisation
ISP	Internet service provider
IuB	UMTS interface between RNC with the Node B
LC	Local Centre
LCON	Local Centre Optical Node
LTE	Long Term Evolution
IuCs	UMTS interface between RNC and Circuit Switched network
IuPs	UMTS interface between RNC and Packet Switched network

IuR	UMTS interface between RNCs
MAC	Media Access Control
MMOG	Massively Multiplayer Online Game
MSC	Mobile Switching Centre
NGA	Next Generation Access
Node B	Base station
NTU	Network Termination Unit
ON	Optical Node
OPEX	Operating Expenditure
P2P	Peer to peer
PC	Personal Computer
PSTN	Public Switched Telephone Network
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RBS	Radio Base Station
RC	Regional Centre
RCON	Regional Centre Optical Node
RF	Radio Frequency
RNC	Radio Network Controller
SD	Standard Definition (Video)
SDH	Synchronous Digital Hierarchy
SGSN	GPRS Support Node
SNR	Signal to Noise Ratio
STB	Set Top Box
TCP	Transmission Control Protocol
TM	Traffic Management
UDP	User Datagram Protocol
Uu	UMTS interface between User Equipment and RBS / Node B
VLAN	Virtual LAN
VLR	Visitor Location Register
VoIP	Voice over IP
WCDMA	Wideband Code Division Multiple Access
WLAN	Wireless Local Area Network technology

Executive Summary

This study was commissioned by Ofcom to provide detail on the technical aspects of traffic management, to explore the effects of traffic management on consumers' quality of experience, and to examine ways of measuring and characterising traffic management and connection performance. The study was initiated independently of the Broadband Stakeholder Group's deliberations on a voluntary code of practice on traffic management transparency[1].

The role of traffic management

Traffic management is used by all UK ISPs. However, in the main, traffic management currently follows a 'fair use' paradigm which is intended only to limit 'excessive' or 'unfair' use by the heaviest users. Until recently, the users considered to generate the most traffic have been running peer to peer (P2P) applications, and ISPs have therefore mainly targeted P2P in their policies. However, currently the greatest traffic growth is in video streaming. Video is widely expected to grow further in the context of internet connected TVs and the launch of hybrid TV services. Without a response of some sort, there will be congestion and a reduction in many users' QoE. In the future there will certainly be other applications that will put pressure on internet capacity.

Traffic management is one possible response to these pressures but expanding capacity is an alternative. In practice, traffic management will probably be pursued alongside capacity expansion. Not all network operators face the same cost structure in expanding capacity so we can foresee that traffic management will develop unevenly between network types. Some ISPs will mainly increase capacity while others will respond through more use of traffic management.

We did not find a strong intention among ISPs to utilise either more traffic management, or more complex traffic management. ISPs told us that they recognised traffic management brings with it technical complexity, cost and market communication consequences. Accordingly, current traffic management approaches are generally the minimum necessary to prevent excessive users from degrading the experience of the majority.

However, for the reasons cited above, it is reasonable to assume that moves in the direction of both *more* traffic management and more complex traffic management will be inevitable overall. This will not be a dramatic increase; instead, traffic management is expected to evolve. We have suggested five possible scenarios for the future of traffic management reflecting different ISP strategies. These are:

1 Fair use
2 Traffic management evolves to facilitate congestion-sensitive traffic
3 Traffic management evolves in response to growth in video streaming
4 Traffic management evolves as a business/marketing tool
5 Managed services become the norm.

Traffic management has often been opposed on net neutrality grounds as being injurious to consumers' interests. An alternative view of traffic management is that it is a way to make the consumer experience more controlled and less subject to the vagaries of congestion. By treating different types of data differently, traffic management allows the performance of applications to be

[1] http://www.broadbanduk.org/content/view/479/7/

managed individually so that the most QoS sensitive applications receive the better QoS from the network. Whereas in an unmanaged situation, consumers would tend not to be able to understand and predict the factors that affect their experience, in a traffic managed situation there is potentially more certainty and more transparency, and a better overall quality of experience for the majority of customers.

The technologies of traffic management

ISPs differ in their traffic management implementations but the basic techniques of traffic management are straightforward. In all cases there is a traffic management *decision* which is then enacted in the network as an *intervention*. The decision can take account of the type of traffic, the user's profile and the cumulative usage relative to any caps or limits that are in place. The traffic management intervention can either be to modify the traffic priority or to change the bandwidth allocated (a guaranteed minimum or to impose a maximum speed cap). These two types of intervention affect different traffic protocols differently.

Traffic management technology is reasonably mature and there is no indication that disruptive technology or breakthroughs might occur. Current technology is adequate to identify traffic types. There is no sense that 'internet abusers' are winning the battle against traffic managers.

The challenge of transparency

The principle of transparency has been broadly accepted by the ISPs that implement traffic management[2], but achieving it is not necessarily straightforward. The challenges include those listed below.

- Traffic management is often non-deterministic. The amount of traffic management and its effects on users can differ according to the level of congestion on the network. Both the amount of traffic management and its impact depend on the level of traffic at the time. For example, on one day it could be that reducing the priority of a particular class of data packet would result in increased latency and jitter. On another day, with greater congestion, there could be more traffic management applied and the impact could be that packets are lost altogether. Because traffic management *policies* alone are insufficient to fully describe the effects of traffic management, full transparency would involve providing data that describe the effects of policies over time and therefore the resulting quality of experience for users.

- Lack of standard metrics. There are no standard industry-wide metrics for the measurement and characterisation of traffic management.

- Traceable measurement is not straightforward. Making traceable measurements can be costly and impose overheads on devices and communication channels.

- Apparent complexity of impact. The way in which traffic management works and its effects on the user experience can appear complex, and not easily communicated and understood.

However there are some aspects of traffic management that are helpful to transparency. These are set out below.

- Underlying simplicity. Despite the technical complexity of implementing traffic management, there are only two interventions available to ISPs, namely to change the priorities of data packets or to change the bandwidth (data rate) allocated to a class of

[2] Including the Broadband Stakeholders Group

traffic. This suggests that a common template for describing traffic management is in principle achievable.

- In-network measurement is possible. Data networks embody the potential for in-network measurements at nodes and interfaces, and some equipment is already able to produce certain traffic statistics. However parameters such as speed, latency and jitter only have full meaning at a consumer's connection and may not be able to be measured meaningfully from within a network.
- Commonality of user behaviour. The majority of consumers use a small number of applications (e.g. email, browsing, streaming video, VoIP) so that a 'key facts' summary of how a particular package would perform should meet most people's needs.

Approaches to transparency

Taking account of the traffic management scenarios, we suggest that transparency involves three factors which cannot always be simultaneously satisfied. These are:

- accuracy
- meaningfulness
- comparability.

Implicit traffic management

As well as explicit traffic management through packet prioritisation and bandwidth management there are forms of network design which affect traffic differentially and can also be regarded as a form of traffic management. The dimensioning of networks, the partitioning of access pipes and the use of CDNs all affect QoS, and can do so in ways that discriminate between traffic types. Some of these issues can be addressed through attempts to improve transparency, some will be resolved through market competition, but others may need regulation.

Recommendations

Our recommendations on ways of measuring and characterising traffic management and connection performance - and the relationships between them - are illustrated in Figure 1 overleaf. The boxes are numbered for ease of reference and discussed below.

1. Tariff package design

Current packages include some degree of traffic management. At present we do not see a need to intervene in the design of packages per se because the combination of transparency and market competition appears sufficient. The possibility of imposing minimum connection standards has been raised by Ofcom but traffic management appears unlikely to affect applications such as email and web browsing which would, presumably, be the core of a set of minimum standards.

An argument might be made for packages being designed to be more comparable. For example, some ISPs calculate cumulative volume over a month whereas others calculate volumes over periods of hours. While comparability is indeed a transparency objective, we do not consider that this should override allowing diversity in the packages offered in the market.

2. QoS Policy Form

Despite the technical complexity of traffic management we have found that basic techniques of traffic management are sufficiently bounded for a common template to be used to describe traffic management policies. The general approach is to identify each type of traffic that is treated separately in a particular policy, and then to describe (i) the time and extent to which data packets of that type are prioritised/de-prioritised – which can range from 'guaranteed' to 'blocked', and (ii) the bandwidth specifically allocated to that traffic type – which can range from a minimum 'guaranteed' rate to a maximum 'restricted' or 'capped' rate.

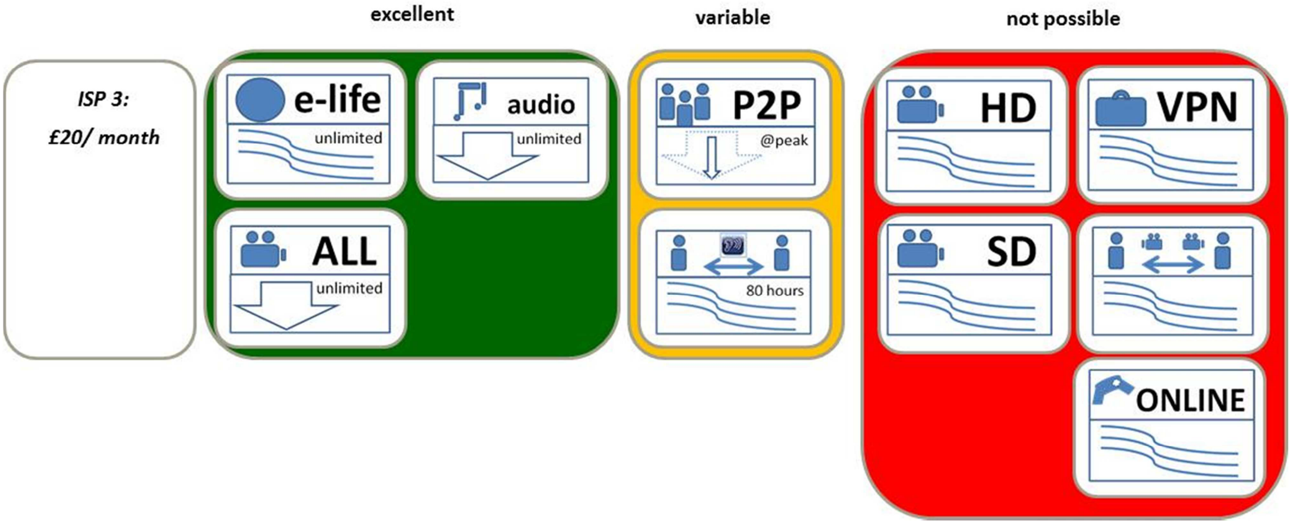
To aid meaningfulness, data rates or volume caps should be given in both bps and indicative units of consumption (e.g. the number of hours of video allowed).

Some information about traffic management cannot be specified purely with reference to an *ex ante* policy, for example, if the amount of traffic management and its effects are statistical (non-deterministic) in nature. In such cases either time series data or estimates should ideally be given in order to provide consumers with greater certainty. However the ability to do this has to be assessed on a case by case basis. Where traffic management is applied in the core, it may be difficult to identify parameters which can give a useful insight into the likely impact of traffic management on an individual customer.

The use of this template is applicable in all traffic management scenarios, though it is conceivable that in the more complex traffic management regimes envisaged in scenario 4, that the template could become large. In that case it might be preferable for consumers to use a 'wizard' though the underlying data in the QoS policy form would still be needed.

3. QoE Summary

To provide consumers with a more meaningful (but less accurate) description when choosing between packages there should be a visual representation of the quality of experience likely to be achieved. An example of such a representation is shown below. The translation between QoS and QoE will need to use standardised values (see box 6 below).

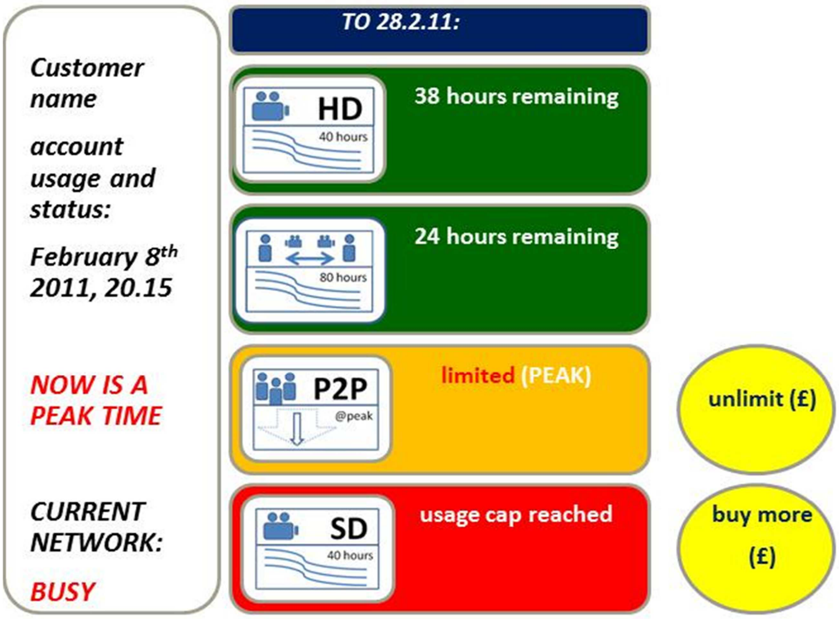


4. ISP-generated in-network measurements and status information (real time and historic)

Data networks have the capability to measure certain traffic statistics. Measurements of performance from within the network can use either software embedded in nodes and interface cards or extra equipment. These are collectively called ‘probes’. ISPs should be encouraged to measure performance and status information, and to provide it both in real time and as historic time series. Where new services are being launched, or there are no measurement data, then ISPs should construct a model or take a series of occasional measurements in order to provide estimates.

5. Real time connection status dashboard

Ideally, consumers should be able to see information about their connection and where they stand in relation to volume limits etc. in real time. One such possibility is illustrated below. Not all network architectures currently support this functionality and it would be costly for some ISPs to implement this. Consumers on tariff packages with simple, high, volume limits are in less need of this information than consumers with complex or low volume limits.



6. Standard QoE thresholds

The translation of QoS into QoE is a necessary step in producing the QoE Summary and will need to be done consistently across ISPs if the QoE Summary is to have value. While such a translation could be undertaken by Ofcom we think that industry bodies such as the Broadband Stakeholder Group should be in a position to agree standards. The translation will need to be updated regularly in line with technical developments and user trends.

7. SamKnows-type measurements

SamKnows[3]. uses monitoring facilities at a user's connection to measure performance and send data back to be aggregated. SamKnows currently only measures QoS for a few traffic types but in principle the same technique could be deployed on different traffic types to give a relative measure of traffic management within the network.

8. Wizard

At present most packages are sufficiently simple that using a 'wizard' to assist consumer decision-making is not essential. If complex packages emerge (such as those suggested in traffic management scenario 4) then wizards could be required. Assuming this is left to third parties, ISPs may be requested such organisations to produce data specifically designed to be input into a wizard.

[3] We will refer to the approach as SamKnows because it is the best example of this approach in the UK, though it is no doubt possible for similar techniques to be used by other companies

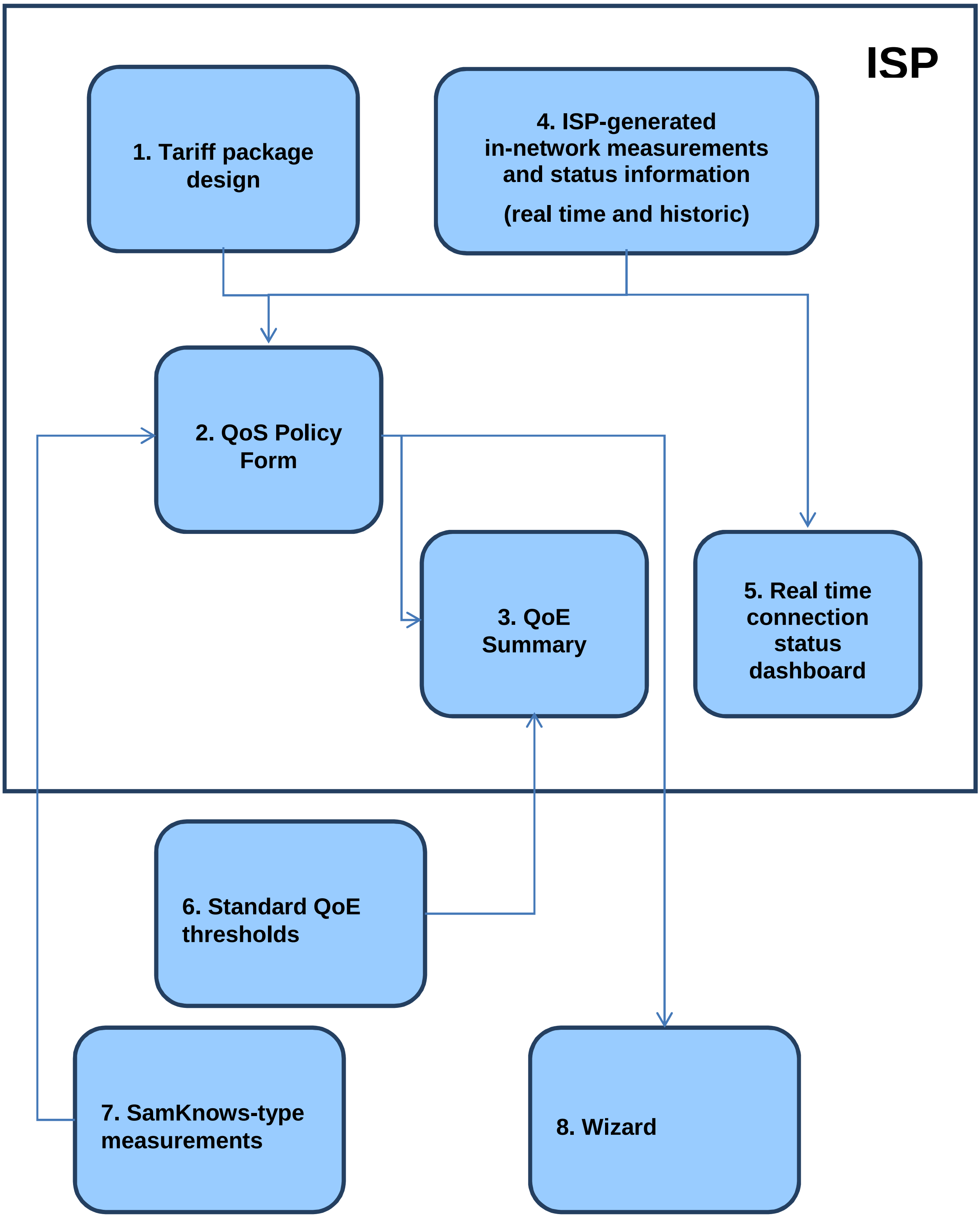

ISP
1. Tariff package design
4. ISP-generated in-network measurements and status information
(real time and historic)
2. QoS Policy Form
3. QoE Summary
5. Real time connection status dashboard
6. Standard QoE thresholds
7. SamKnows-type measurements
8. Wizard

Figure 1: Overall recommendations

1 Introduction

Traffic management has recently risen in importance. There are emerging ‘battle lines’ between those who oppose traffic management and believe that the internet should exhibit strict neutrality, and those who believe that the mix of traffic on the internet has made traffic management a commercial and technical necessity. This study was commissioned by Ofcom to provide detail on the technical aspects of traffic management, to explore the effects of traffic management on consumers’ quality of experience, and to examine ways of measuring and characterising traffic management and connection performance. The study was initiated independently of the Broadband Stakeholder Group’s deliberations on a voluntary code of practice on traffic management transparency[4].

The study has involved desk research and interviews with a representative selection of ISPs, technology providers and content providers. All interviews were conducted on a non-attributable basis. Where reference is made in this report to individual companies, the material has been drawn from sources already in the public domain.

Some respondents asked that their participation should not be disclosed, so we have decided not to list the companies we spoke to. However we are grateful for the assistance provided by all the companies that we contacted during the preparation of this report.

[4] http://www.broadbanduk.org/content/view/479/7/

2 Definitions

There are many definitions of traffic management in use. From a technical point of view there are many aspects of network design and operation that affect QoS, and even affect the QoS of one traffic type compared to another. We think the key to traffic management as envisaged in our brief (Appendix A) is that the discrimination between types of traffic needs to arise out of a purpose. Accordingly, for this project we have defined traffic management as "**purposeful discrimination in access to network resources on the basis of traffic type, origin, or destination"**.

The internet is a packet switched network in which packets are normally transmitted on a best efforts basis. When there is congestion (more traffic than can be handled), data may be delayed or lost. Without traffic management, different data packets are treated more or less equally. Traffic management is a collection of technologies and policies which lead to different types of traffic being treated differently. Under congested conditions traffic management would cause some data to have a greater chance of being delivered than others. Such discrimination between data types would probably affect users' experience; in the extreme some applications would not be able to function. Of course, congestion could also cause applications to fail, but the distinguishing feature of traffic management is that it involves *purposeful* discrimination.

Considerations of whether certain practices are – or are not – traffic management has led us to refine our definition by recognising several sub-divisions within traffic management. The terms describing the sub-divisions are not necessarily in wide use but do, we believe, help delineate the field.

Some traffic management is 'explicit'. **Explicit traffic management** within the open internet involves identifying the class of traffic involved and then allocating data bandwidth or packet priority on a discriminatory basis.

There can also be 'implicit' traffic management. **Implicit traffic management** within the open internet is said to occur when the design and provisioning of a network has the effect of discriminating between traffic classes. For example the capacity of the pipes between content providers and an internet gateway affects the likelihood of it becoming congested. The word 'purposeful' in our definition seeks to exclude situations where discrimination has arisen without a deliberate intention to favour one sort of traffic over another. One form of implicit traffic management is the use of content distribution networks (CDNs). CDNs have a network infrastructure and provide local caching and/or connections to major points of presence. Both these techniques result in improved user experiences with respect to the content carried over the CDN.

The definitions given above focus on the open internet. Recent developments in IPTV and other managed services has resulted in some of the physical infrastructure (mainly the access network) being shared between the internet and managed services. For example, an access link could be physically shared between internet traffic and a paid-for IPTV service such as BT Vision, Virgin Media Player, Sky Anytime Plus and the forthcoming YouView service. These arrangements are not generally regarded as a form of traffic management but we consider that they should be. Accordingly we have also considered what we call '**partitioning**' of bandwidth between the internet and other services (such as IPTV) which may use some of the same physical infrastructure.

There are two sorts of partitioning – ‘static partitioning’ and ‘dynamic partitioning’. **Static partitioning** is where links are shared on a relatively fixed basis. **Dynamic partitioning** is where the sharing is variable, such as when a user pays to watch a movie using a managed IPTV service. Such an arrangement could lead to bandwidth being taken out of a broadband link for the duration of the movie.

3 Technology overview of traffic management

3.1 Introduction

This chapter provides a technical overview of traffic management and its application to the UK internet.

3.2 The principles of traffic management

One challenge in any study of traffic management is that it can be described in many different ways. For example, traffic management can be viewed in terms of:

1. its application across different network types (e.g. fixed DSL, cable, mobile);
2. where it is controlled and enacted within the layers of the ISO 7-layer model of communications;
3. where it is controlled and enacted in the physical, geographic network (e.g. core network, access network);
4. the impact it has on different traffic types (P2P, web browsing, streaming video, etc.);
5. the impact it has on different users, or classes of user;
6. the type of traffic management intervention that is used, and the decision bases for enacting the intervention.

In order to manage the complexity of this topic, we have selected a number of 'views', which provide a good description of the types of traffic management in use and the effects these have on QoE for consumers. These views will form the foundation for our QoE work and will help Ofcom to understand the issues and determine policy.

Within the main body of the report we have provided two 'views':

- an Intervention View (point 6. above, at section 3.3);
- a Physical Network View (point 3. above, at section 3.4).

We believe that these two views will provide the understanding that this study needs. However, we have also included in the appendix two additional views which may be helpful to some readers:

- an ISO Model View (point 2. above, at Appendix D);
- a Network Type View (point 1. above, at Appendix E).

We believe that these four views together provide the detail required to understand how traffic management in the UK works and to inform the process of setting policy.

Sections 3.3 and 3.4 below describe the two views that underpin the remainder of the project work. Subsequent sections within this chapter look at the way in which traffic management develops as networks expand and mature (section 3.5) and future developments in traffic management (section 3.6).

The chapter concludes with a summary of the differences between traffic management application in fixed and mobile networks (section 3.7), and some observations on the status of traffic management technology today and in the future (section 3.8).

3.3 An Intervention View of Traffic Management

While the implementation of traffic management is far from trivial, there are relatively few underlying techniques available. All traffic management involves a *decision basis* and an *intervention*. For example, exceeding a monthly usage allowance is a decision basis, and the response of cutting data rate according to policy is an intervention.

ISPs use many criteria to decide on what traffic management to apply. However, our study has shown that there are three main inputs used by UK ISPs in reaching decisions on what traffic management to apply, though they can be used in combination, and more complex rules can be set. The three decision inputs are:

- the *user identity* (or profile), specifying a QoS package for that user;
- whether or not a *usage cap* has been exceeded (note that these caps are often set by the user's tariff);
- the particular *traffic type*.

With regard to interventions there are two main types.

- *Packet prioritisation*. Wherever queues occur in a network, higher priority traffic will get through whereas lower priority traffic may be delayed or suffer packet loss. This is typically applied today in the core network, but may in future migrate closer to the access network to increase the effectiveness of traffic management in maximising network utilisation but minimising the effect on most users.
- *Bandwidth allocation*. The bandwidth (or data rate) offered to a user or a type of traffic can be actively controlled. Users can be offered a minimum guaranteed rate or can be limited or capped at a maximum rate. In most cases this is applied at levels 2 and 3 in the scheduler or in the access network which is where most networks have the greatest constraints on bandwidth.

Figure 2 below shows the 6 traffic management interventions that exist in the 2 by 3 matrix which describes all combinations of intervention type and decision input. For ease of reference we have labelled the cells by their row and column number.

Decision Input **Intervention type**	**1. User identity**	**2. Usage cap**	**3. Traffic type**
A. Packet prioritisation	A1	A2	A3
B. Bandwidth allocation	B1	B2	B3

Figure 2: Matrix of traffic management approaches

We have reviewed all the traffic management interventions identified in our interviews and have confirmed that they do indeed fall into one or more of the cells in this matrix. Cell A2 is greyed out because we have found no evidence of ISPs currently using this type of intervention (implementing a usage cap via the packet prioritisation route). We found at least one example of an ISP using each other of the interventions identified in the matrix.

The location in the network of the decision and the intervention is not necessarily identical. For example, some traffic management decisions are made at management centres in the core network, but are implemented in the access network.

Similarly, traffic management decisions and implementations often span different layers in the ISO model. Many decisions get made at ISO Layers 3 (Network) or 4 (Transport) even if they are subsequently enacted at ISO Layer 2 (Data Link).

3.3.1 Intervention types

The two types of intervention have different characteristics and it is helpful to understand these when predicting the effects of traffic management on overall network traffic and on consumers' QoE. The main characteristics are listed in Table 1.

.

	Characteristic					
	Possible actions	**Currently applied in**	**ISO Model Level**	**Impact of negative intervention on data type**		**Comments**
Intervention type				**TCP/IP** **FTP**	**UDP** **RTP**	
A. Packet prioritisation	Prioritise or De-prioritise	Core network	Layers 3 and 4	Retransmission of packets	Data loss	TCP/IP traffic can be effectively managed by de-prioritising this traffic type
B. Bandwidth allocation	Guarantee or Cap	Access network	Layers 2 and 3	Reduced throughput (Service maintained, but at lower speed)	Reduced quality (Codec may drop to a lower rate)	Video is best managed by prioritising or giving guaranteed bandwidth in the access network

Table 1: Characteristics of intervention types

For each intervention type there is a positive action, which is generally beneficial for the traffic concerned, and a negative action, which may limit or constrain it. The positive approaches are sometimes used by ISPs to support 'up-selling' of consumers onto better packages (that offer greater throughput, or higher usage caps), or as the basis for managed services (which may guarantee performance for a particular traffic type or piece of content).

We explore the two types of intervention below.

Packet prioritisation (ISO layers 3 and 4)

Packet prioritisation makes use of the provision in packet headers of a field to indicate priority level or mark a packet as being of a particular traffic type. Protocols allow for end to end support for priority levels but this is not generally implemented across network boundaries. In practice, ISPs disregard the priority indicated on packets as they enter their networks and reset priorities to match their own traffic management policies.

Networks are made up of switching nodes, routing nodes and transmission links. Nodes switch/route packets from inputs to selected outputs. There are many complicated queuing algorithms used by ISPs to optimise traffic flow, but detailed knowledge of these is not needed in order to appreciate how ISPs manage traffic. The principle is that these nodes operate queues according to packet priority.

Within the ISO model it is important to appreciate that the Layer 3 (Network) does not guarantee the delivery of IP packets. Thus nodes are permitted to delay packets and drop packets that have queued too long. Low priority packets will tend to be delayed and/or dropped. Algorithms implemented at Layer 4 and above will determine whether packets have been lost and arrange re-transmission where necessary.

Different data types fare differently under packet de-prioritisation. In the case of TCP/IP traffic (e.g. P2P file sharing) on an IP network, flow control is implemented via a buffer at the receive end. If TCP/IP data is de-prioritised it suffers increased latency or data loss. High latency causes the flow control mechanisms at either end to reduce data rate. When the buffer is approaching capacity the receiver will tell the transmitter to stop sending data. Transmission errors are overcome by the receiver requesting re-transmission of any lost or corrupted packets. This works well when there is adequate network capacity and not too many packets are lost or corrupted. However, on a congested or poor-quality network, TCP/IP becomes increasingly inefficient and, in extreme conditions, contributes to further congestion through frequent attempted re-tries to send data.

UDP traffic (e.g. video streaming) would also suffer increased latency and data loss. Unlike TCP/IP packets, delayed or lost UDP packets are not re-transmitted. So, unlike TCP/IP, de-prioritisation of UDP packets will not lead to increased network load through attempted re-transmissions. However, depending on the application, users are likely to notice the effects of delayed or incomplete data, so packet prioritisation is not the best way to manage UDP traffic.

Bandwidth allocation (ISO layers 2 and 3)

Bandwidth allocation does not cause packet loss unless it reduces data rate to below that required for a particular application. Adequate data rate is particularly important for codecs, which necessarily operate in real time. Bandwidths can be set either for a connection as a whole or for individual traffic types separately.

The basic methodology for a telecommunication system works on the principle of taking a data stream from the upper layers of the ISO model and transmitting this over a physical interface using a modulated signal. The signals are modulated onto a physical layer medium transmitting and receiving the information sent in the frequency and time domain.

Networks are managed from two fundamental planes - data and control planes. The data contains the information being transmitted / received and some instructions on how this will be dealt with in the network. The control plane dictates how the network is managed including priorities for the information transmitted and potential bandwidth allocated in the pipes carrying the information.

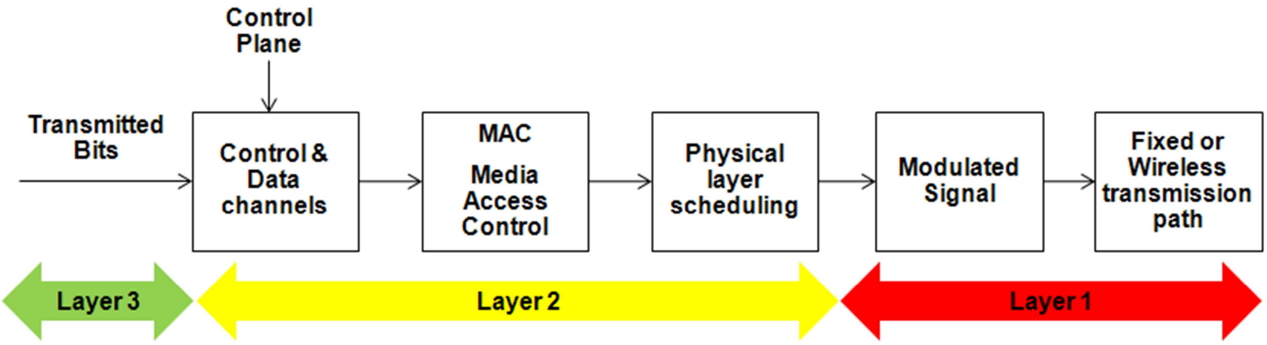


Figure 3 – Basic transmission path, functional blocks

Figure 3 shows the basic functional blocks of how an access node deals with the transmitted bits (data) and uses the control plane signals to create channels. The Media Access Control makes decisions on how the data should be scheduled into transmission medium resource. The physical layer scheduler will schedule information according to service priority, characteristic and available bandwidth determined by the link quality of the physical path, normally determined by the signal to noise ratio (SNR). The signal to noise ratio will determine how fast the link can be operated and is achieved by selecting a higher order modulation scheme which effectively allows more bits to be packed into the same physical allocation.

If no bandwidth control is required then the default mode will be to transmit the information in the highest order modulation scheme, therefore the fastest available bandwidth. In the case where a user is being bandwidth restricted then Layer 2 can control the actual amount of data scheduled to be transmitted over a fixed time period.

The bandwidth available (physical resource) and modulation schemes vary by access technology. For example, 3G wireless systems operate on 5MHz carriers which are shared by all users with a particular carrier/sector. The overall carrier is split into sub blocks which can be allocated depending on the characteristics of the different traffic types.

Cable uses multiple 8MHz channels and DSL networks have a frequency spectrum of approximately 1.1MHz. 3G and cable systems work on the principle of a shared medium and therefore can be used to allocate different amounts of bandwidth to different users. DSL's medium is a physical copper or aluminium wire which is not shared but suffers from high interference due to the poor RF quality of the cabling.

Different methodologies can be used to differentiate traffic and manage bandwidth from Layer 3 to 1 to set up multiple transmission pipes to an individual user. The scheduler will make decisions based on the available or restricted bandwidth.

The transmitted bits, at the layer 3 level, may have already been shaped or policed to determine the priority of traffic entering into the lower ISO layer transmission which will subsequently control the bandwidth (upper limit cap and lower limit guaranteed speed).

Traffic policing results in limiting to a maximum rate, excess traffic is dropped (or remarked) which results in peak traffic being smoothed. By contrast, traffic shaping holds packets back in a queue and then schedules for later transmission. The result of traffic shaping is a smoothed packet output rate.

The UMTS standard for example has four different classes, see Table 2 below.

	Conversational class	Streaming class	Interactive class	Background class
Fundamental characteristics	**Real Time**	**Real Time**	**Best Effort**	**Best Effort**
	Low delay guaranteed bit rate	Guaranteed bit rate	No guaranteed bit rate	-Destination is not expecting the data within a certain time
	- Preserve time relation (variation) between information entities of the stream	- Preserve time relation (variation) between information entities of the stream	- Request response pattern	
	- Conversational pattern (stringent and low delay)		-Preserve payload content	-Preserve payload content
Example of the application	Voice	Streaming video	Web browsing	emails

Table 2 - UMTS QoS Classes, main parameters

DOCSIS and DSL networks can utilise the QoS control defined by IEEE 802.1P, known as class of service (CoS). This is implemented as a 3-bit field called the Priority Code Point (PCP) which specifies a priority value of between 0 and 7 inclusive that can be used by Layer 2 QoS processes to differentiate and schedule traffic.

Table 3 below shows the QoS levels and traffic characteristics.

Network priority	Traffic characteristics	3 Bit PCP
0 (lowest)	Background	1
1	Best Effort	0
2	Excellent Effort	2
3	Critical Applications	3
4	Video, < 100 ms latency	4
5	Voice, < 10 ms latency	5
6	Internetwork Control	6
7 (highest)	Network Control	7

Table 3 - IEEE802.1P CoS network priority classes

Decisions at Layer 2 implementation can be determined by the Network Policy & Control strategy and the information can be programmed into nodes, transmitted over IP or sent separately over a control plane dynamically to change implementation. The IP (layer 3) may have been marked by a

DPI node or by the ISP traffic shaping/policing function for the layer 2 MAC to act upon accordingly.

3.3.2 Decision inputs

In principle traffic management decisions can take account of many different factors. In practice we found three main types of decision basis, though they can be used together. These are:

- the user's identity, providing reference to their tariff and account details. These might indicate such factors as the allowable bandwidth or how other services, such as managed IPTV, will co-exist with general internet activity
- usage caps – implemented in general by comparing the value of a counter with a pre-defined limit
- the type of traffic, identified typically using packet inspection technologies.

The characteristics of these different decision bases are shown in Table 4.

	Characteristics		
Decision input	**Input measurement or context variable**	**Types identified**	**Example actions**
1. User identity	User account	Consumer / business Tiers of tariff	Allocate bandwidth Prohibit or allow traffic
2. Usage cap	Packet counter	Download or upload amount per period	Warn user of approach to cap Limit bandwidth if cap exceeded
3. Traffic type	Packet inspection	P2P, VoIP, gaming Audio / video streaming	Prioritise or de-prioritise Limit or guarantee bandwidth

Table 4: Characteristics of decision inputs

The most controversial and potentially complex form of decision input is the traffic type. The majority of traffic type identification is initiated by inspecting packet headers and marking[5] them accordingly for transmission across the network. The inspection equipment will investigate the header information being transmitted across layer 3 and, based on criteria set in the Policy and Control node, will implement IP header manipulation. The complexity of this is determined by the ISP policy on traffic management. This ranges from simple blanket prioritisation, such as marking all P2P, through to complex bandwidth allocation by user and/or service intervening by changing packet headers and controlling pipe speeds.

The ISPs told us that they use established techniques to identify different types of traffic on their networks. Typically this is achieved by one or more of the following methods:

- association – by noting sending or receiving IP addresses, physical device types or the ports on which traffic is presented;

[5] Traffic derived from within the network could already be marked by the source according to the traffic management policy and therefore DPI techniques are not required. Traffic shaping and policing can therefore be applied based on the known marking.

- shallow inspection – looking at headers to identify data protocols;
- deep inspection – looking inside packets at the data payload;
- heuristic – looking at the pattern of traffic to determine its type.

In some cases users or content sources attempt to improve the throughput of their data by changing packet headers to disguise their data as another type, which they believe is subject to less management. ISPs told us that they were generally able to identify traffic types using other techniques and were quickly able to minimise the impact of such actions on their networks.

The equipment that identifies traffic is often referred to as a “DPI Box”, regardless of whether traffic is identified actually using deep packet inspection, by shallow inspection or using heuristic methods. ISPs told us that they currently have no need to look in detail at the data payload in order to make commercially- or technically-motivated traffic management decisions.

3.3.3 Interventions used by ISPs

We have analysed the interview responses and published policies from ISPs in order to infer the interventions in use, and we have matched these against the six traffic management interventions defined in Figure 1. One of these interventions (A2) was never used, leaving five remaining.

We found that the DSL ISPs are collectively[6] using all five of the interventions.

We found that mobile operators were using all the interventions except B3 (bandwidth allocation by traffic type). We believe that mobile operators currently have no need to intervene in this way since the bandwidth available in their radio access networks already provides a physical limit which does not need to be augmented. They maintain active control through application of the other interventions. A1 (packet prioritisation by user identity) is the mechanism used, we believe, to block VoIP in some mobile tariffs.

We found that cable providers were using all the interventions except A1 (packet prioritisation by user identity) and B3 (bandwidth allocation by traffic type). We believe that cable operators currently have no need for intervention A1 as they have sufficient bandwidth headroom in their access network that they don’t need to prioritise packets by user; prioritisation by traffic type (A3) provides the control they need. Their access network bandwidth headroom also means they don’t currently have a need to allocate bandwidth by traffic type, meaning that B3 is not required at this time.

We did not find that any of the five interventions in use were exclusive to a particular network type, or that they operated fundamentally differently when used by operators over mobile, DSL or cable networks. The absence of conflicting interventions and characteristics across the network types means that it would be possible to design a single, flexible, traffic management regulatory policy which is applicable to all types of network, and all ‘flavours’ of ISP. Provided that ISPs are allowed to select which of the five interventions they use, they will have the tools they need to manage traffic on their networks and to give their users an appropriate QoE

3.4 A Physical Network View of Traffic Management

To explain the concepts detailed in the previous sections the diagram below, Figure 4, shows how a system using packet inspection can inspect the IP traffic and mark packets accordingly for the network to deal with, driven by the Policy & Control strategy.

[6] This is not to imply that *any individual ADSL ISP* uses all five

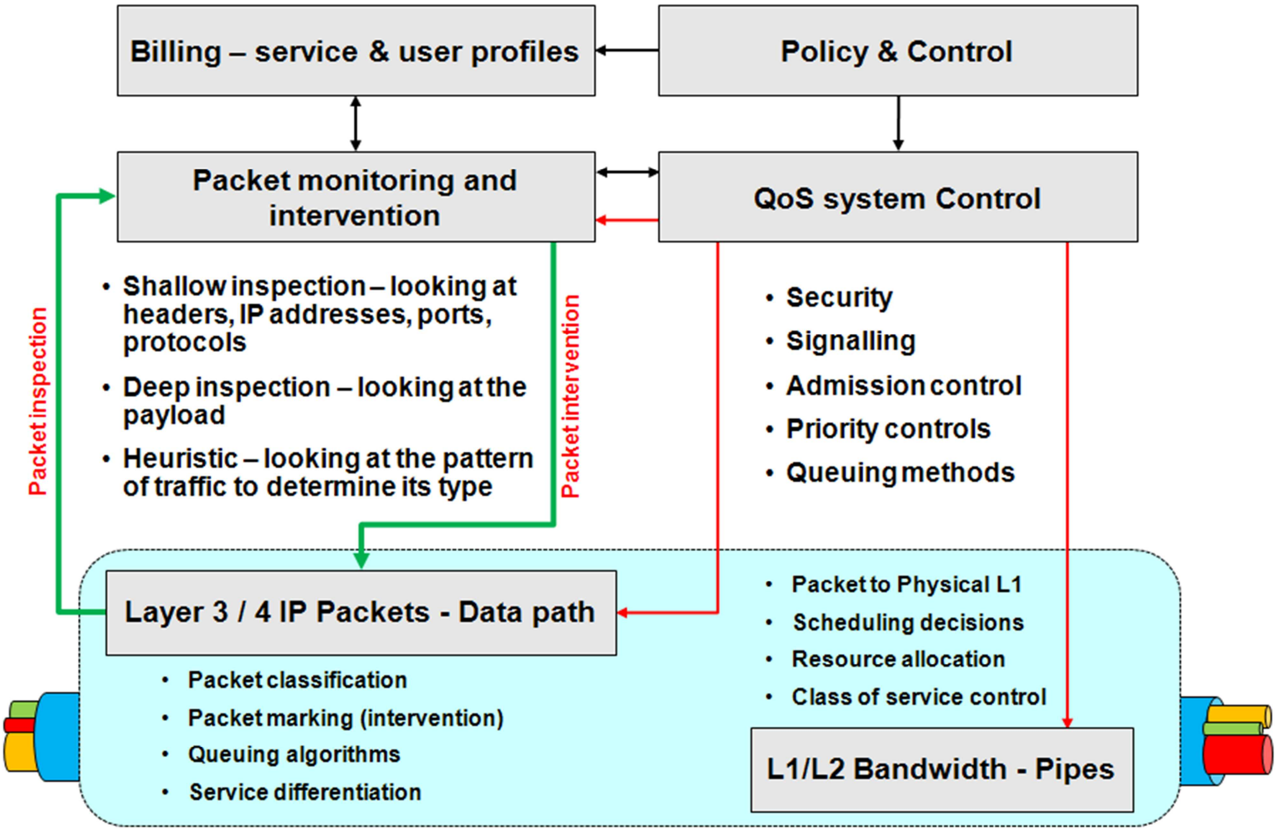

Billing – service & user profiles
Policy & Control
Packet monitoring and intervention
QoS system Control
Shallow inspection – looking at headers, IP addresses, ports, protocols
Deep inspection – looking at the payload
Heuristic – looking at the pattern of traffic to determine its type
Packet inspection
Packet intervention
Security
Signalling
Admission control
Priority controls
Queuing methods
Layer 3 / 4 IP Packets - Data path
Packet to Physical L1
Scheduling decisions
Resource allocation
Class of service control
Packet classification
Packet marking (intervention)
Queuing algorithms
Service differentiation
L1/L2 Bandwidth - Pipes

Figure 4 - Generic traffic management architecture

The majority of IP intervention is dealt with by inspecting packet headers and marking them accordingly for transmission across the network. The inspection equipment will investigate the header information being transmitted across Layer 3 and, based on the criteria set in the policy and control unit, will implement IP header manipulation. The sequence of events is denoted by the green lines on the diagram. The complexity of this is determined by the ISP policy on traffic management, from simple blanket prioritisation such as marking all P2P through to complex bandwidth by user and/or service intervention by changing packet headers and controlling pipe speeds. The control is implemented over the control plane of the network and is denoted by the red lines on the diagram.

To further explain this methodology for traffic management Figure 5 below shows a generic network architecture with the functions associated with particular nodes.

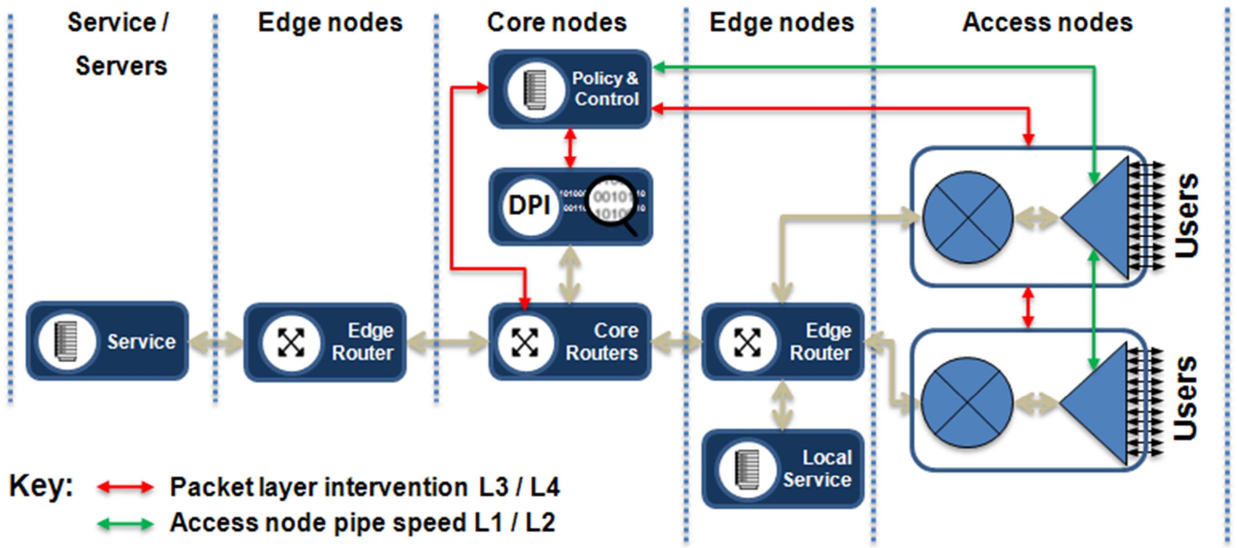


Figure 5 - Network agnostic traffic management example by node

The key aspects of Figure 5 are as follows:

- DPI boxes are deployed in the core network nodes to inspect the packets to determine traffic types. Information is passed to the policy and control node.
- Policy and control units typically contain the traffic management policies and, based on the information received from the DPI box, send control signals to the respective nodes on how to deal with the traffic.
- The red lines indicate the packet based intervention where the core nodes re-label packets based on the priority decided by the traffic management policy, and the access nodes treat them accordingly. As most networks today have the ability to inspect packet headers then all packet header fields in theory could be manipulated.
- The green line indicates control of the access node Layer 1 and Layer 2. For example the DPI box could monitor a monthly usage cap and when the limit is reached could apply a reduction to the pipe speed by allocating less resource to the end user in the access node.

Traffic management architectures differ by access technology but the fundamental principles remain valid for all network types. Appendix E includes some generic network architectures by access types: DSL, Cable and 3G.

3.5 Traffic management versus capacity expansion

From the analysis conducted it appears that traffic management extent and complexity is associated with how much congestion is being experienced or how close the network traffic is to the limit of the network capacity (see Figure 6)

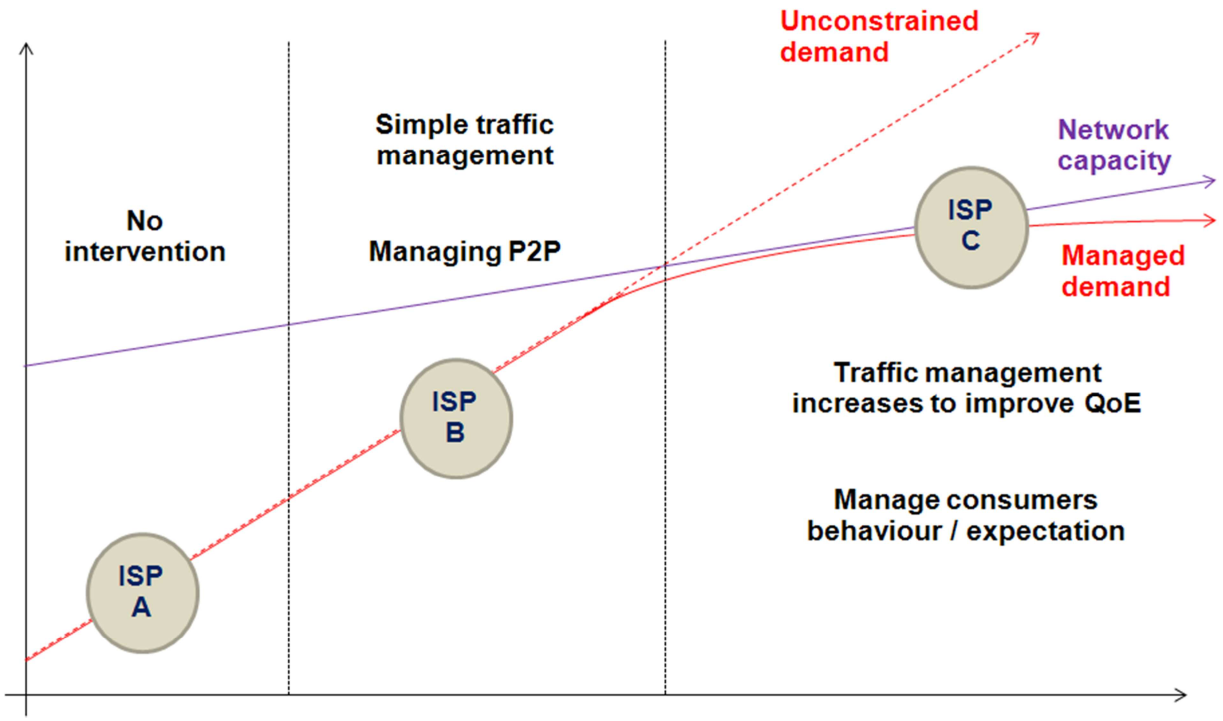

Unconstrained demand
Network capacity
Managed demand
No intervention
Simple traffic management
Managing P2P
Traffic management increases to improve QoE
Manage consumers behaviour / expectation
ISP A
ISP B
ISP C

Figure 6: Traffic management as a response to capacity limitations

Traffic management in general is a response to an emerging need to manage the network resource as it approaches capacity limits. This results in congestion which in turn impacts the Quality of Experience (QoE) for the end users. Figure 6 shows three hypothetical ISPs. The x axis represents the stage of development of the ISP. The y axis represents data demand and available network capacity available to that ISP.

- ISP A has plenty of capacity and can deliver all services at a high QoE as no nodes are congested. Even heavy users do not impact the network.
- ISP B has less headroom and a minority of heavy users are causing congestion. In this case the ISP identifies that P2P users are the root cause and implements a policy where P2P traffic is managed therefore bringing the congestion under control and maintaining a high level of QoE.
- ISP C is reaching its capacity limit, and multiple traffic types are causing congestion. In this case the ISP starts to differentiate between multiple traffic types and/or users. The intervention is based on QoE characteristics of the service. For example, video is prioritised over a non real time application. The complexity of intervention has increased with multiple dimensions and service packages are differentiated to manage the end user expectation and behaviour.

Increasing capacity is often preferred to managing traffic. In practice this may not be possible due to the economic implications or because the fundamental technology of the network is at its limit, in terms of capacity or pipe speeds. Traffic management may be deployed as a ‘stop gap’ before new capacity comes on line. Traffic management can allow ISPs to delay the point at which new capacity is installed. This improves asset utilisation whilst maintaining an acceptable QoS for users until the business case for new capacity is strong enough to justify investment.

Figure 6 can also be treated as a single ISP (hypothetical) evolution over time. The ISP starts at position A, where there is plenty of capacity. It ends at position C, where congestion has to be managed at a finer granularity to retain the QoE and stop potential churn.

3.6 Where is traffic management heading over the next five years

From a purely technical point of view, traffic management will remain as a response to congestion and a mechanism to maintain the highest level of QoE. The most important changes will be that packet inspection capabilities will migrate outwards from the core, enabling a more finely graded and user-specific form of traffic management. Figure 7 below shows a potential scenario based on a hypothetical single network architecture evolution over time.

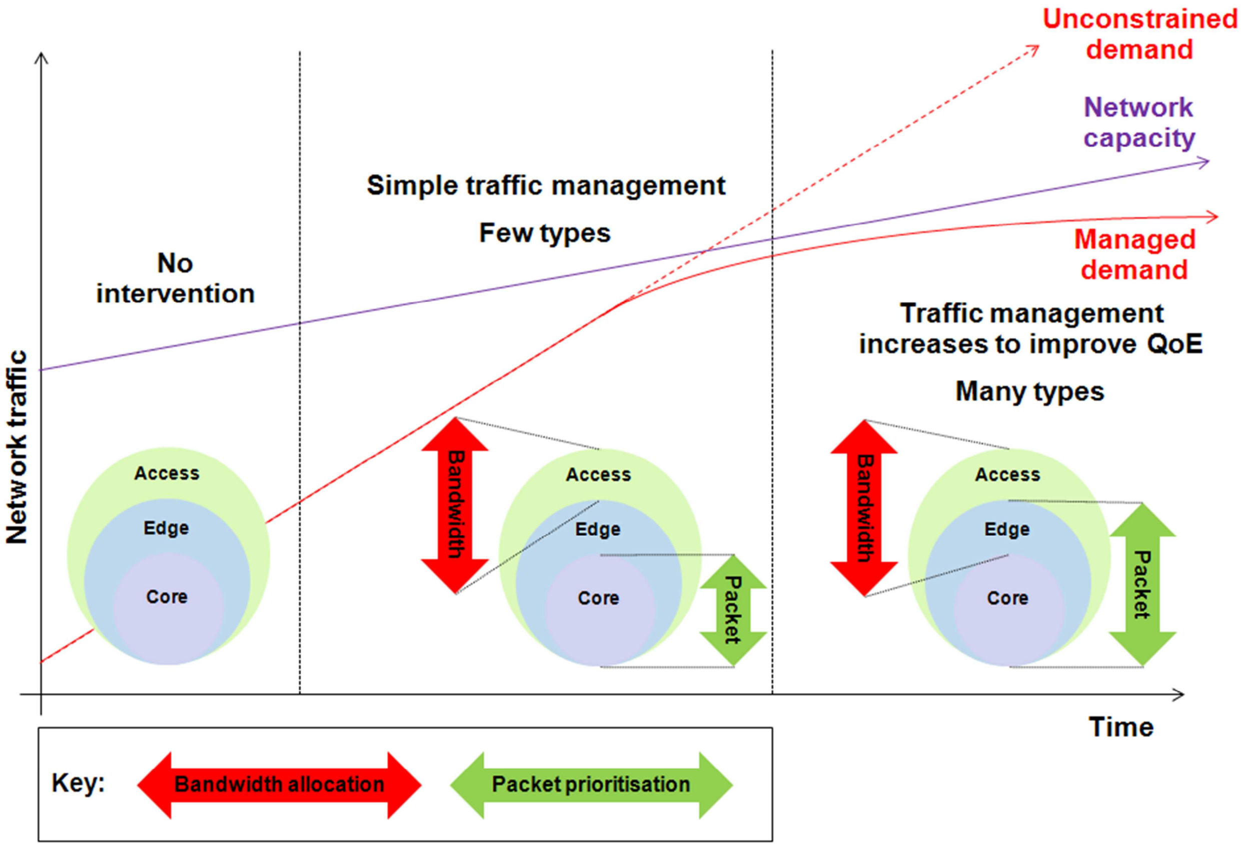


Figure 7 - Traffic Management evolution in relation to network technology

- **No intervention** – Network has plenty of capacity and can deliver all services at a high QoE as no nodes are congested. Even heavy users do not impact the network.
- **Traffic management few types** – Network has less headroom. A small number of users and/or traffic types are the root cause and the operator implements a policy where traffic is managed, therefore bringing the congestion under control and maintaining a high level of QoE. Basic intervention occurs through packet prioritisation in the core and managing bandwidth at the access through traffic shaping and policing intervention.
- **Traffic management many types** – Network is reaching its fundamental capacity limit and multiple traffic types are causing congestion. In this case the network technology starts to differentiate multiple traffic types and/or users. The intervention is based on QoE characteristics of the service. The intervention now has multiple dimensions and service packages are differentiated to manage the end user expectation and behaviour. DPI functionality and packet prioritisation is becoming more complex and migrating from the core to the edge in order to

provide maximum control. Bandwidth control due to fundamental limitations is being applied at the access and edge

3.7 Differences between fixed and mobile networks

We have included in Appendix E a set of generic network architecture diagrams and brief descriptions for each of the three main types of network - DSL, cable and mobile. We have annotated these with information about how traffic management is currently carried out on these networks.

Our analysis indicates that there are some detailed differences in the location, control and enactment of traffic management between the network types. However, these differences primarily reflect the variations in network architecture, rather than any fundamental difference in approach to traffic management.

Each network type features a centralised Policy and Control function, which stores the network management policy and converts this into instructions for specific management actions. Each network type features DPI boxes[7] in the core network which mark traffic types and enable packet prioritisation to be enacted in the core and access networks as determined by the policy. Bandwidth allocation is mostly implemented in the access network, where bandwidth is more scarce and there is greater potential for the actions of individual consumers to impact the QoE of others.

ISPs told us that there is a general trend to move packet inspection and prioritisation towards the edge of networks, where it can provide finer levels of control and thus an improved QoE for consumers. However, the cost of doing this is high, so it will only happen if it can be shown to bring significant benefits. It may be more cost-effective in many cases for ISPs to address congestion by installing additional network capacity rather than by increasing the number of packet inspection points in the network.

The main difference in implementation occurs in the case of mobile networks, which uniquely have a free space radio link as their access medium. Mobile operators also have to manage the challenge of users who move around both within and between cells, and of large changes in local demand, for example during major sporting or entertainment events.

Since radio spectrum is a limited and costly resource, mobile operators always work to make best use of it. This means that the radio access network is likely to remain as the limiting point in terms of traffic capacity on most mobile networks for the foreseeable future. Its limited capacity also acts as a natural traffic management function, which means that mobile operators generally don't have to apply bandwidth allocation for different data services in the access network. They do however apply rules that allocate radio network resources between voice and data traffic, to ensure that each gets a fair share. This isn't an issue for DSL or cable, as in these cases voice traffic is carried on a reserved part of the available spectrum (DSL) or on a separate copper pair (cable). So coexistence with voice is far more of an issue in radio networks, and voice still tends to have priority.

On the basis of our knowledge of the different network types, and the information gained from the interviews, we do not see the need for any fundamentally different regulatory policies for traffic management being required for ISPs using DSL, cable or mobile networks. There will continue to be traffic management policy differences between operators, but these will reflect the demand for their services, the development status of their networks and differences in the capacity of their access networks, rather than any fundamental difference in approach driven by network type.

[7] Though not necessarily using deep packet inspection

3.8 Traffic management today and in the future

ISPs told us that traffic management technology, as applicable to IP networks, is already relatively mature. All ISPs have access to packet inspection techniques that allow them to implement their desired packet prioritisation policies. Collectively they have the ability to measure traffic by user, by traffic type and apply prioritisation by user, by traffic type and by time of day. They can apply usage caps and can control the bandwidth consumed by individual users or types of traffic. They are able to follow and react to trends in the market where some users attempt to disguise a traffic type as another to reduce the impact of traffic management. They can identify and control problem users and prevent them from degrading the service of other customers.

Real-time packet inspection is currently an expensive activity. Like most electronic goods, DPI products will increase in performance and reduce in cost over time. Increased affordability will present an opportunity for operators to move inspection out from the core towards the network edge. The result is likely to be more user-specific traffic management policies in the future, giving operators additional controllability of high-use consumers and traffic types.

Operators will be in a position to offer differentiated packages, for example featuring a guaranteed performance for certain types of traffic (e.g. games or streamed video). Current traffic management technology has the ability to support the delivery of such offerings in principle but the cost of deployment is the issue.

ISPs do not expect any major dislocations or significant changes in the way they manage traffic over the next five years. They see a steadily increasing demand for bandwidth being matched by investment in new capacity, which they may choose to fund partly through differentiated service offerings. This will continue to be supplemented by traffic management to optimise network utilisation, manage peak loadings and ensure that the actions of a few heavy users don't impact the QoE of the majority.

The increasing popularity of streaming video will continue to cause stress for most operators. It demands relatively high bandwidth, tends to need low jitter, and consumers expect it to be delivered at ever-increasing quality and to operate without interruption for several hours (e.g. when viewing a movie). Unlike P2P (which can be de-prioritised or bandwidth throttled) streamed video cannot easily be managed, other than by a co-operative process with content providers to employ more efficient codecs, or to buffer or cache content. ISPs can be expected to develop their traffic management policies to cater for the growing demand for video, whilst maintaining an acceptable QoE for other users.

In mobile, the radio access network (RAN) can additionally be managed through the allocation of codes (effectively, the relative allocation of spectrum).

4 ISPs policies and the current use of traffic management

4.1 ISP policies

The ISPs we interviewed all claim that they use traffic management strictly according to published policies. These policies are either constituted as separate documents on their websites or are effectively incorporated into the details of tariffs.

4.1.1 Fixed ISPs

In Table 5 we have summarised the published policies of a selection of the main[8] fixed ISPs. We have included the top six UK ISPs together with Plusnet which is widely cited for its use of traffic management. The entries in the table show that a policy applies to at least one of an ISP's tariffs but not necessarily all of them.

	Volume limits	**P2P policy**	**Video streaming policy**	**Comments**
BT fixed	Yes	Yes	No	No other traffic management specified on website
TalkTalk fixed	Yes	Yes	No	No other traffic management specified on website
Virgin fixed	Yes	Yes	No	No other traffic management specified on website
Sky fixed	Yes	Yes	No	Heavy users monitored and restricted. Sky "may slow down the speed that all Sky Broadband Connect customers can get on applications such as peer-to-peer networking and newsgroups, which we consider use up a lot of bandwidth and have a negative effect on other customers.[9]
Orange fixed	Yes	Yes	No	"Traffic management is where we sometimes apply restrictions to the amount of network capacity a customer can use, which can affect your throughput speed. We do this to stop a small number of customers who excessively download during peak times (6pm to midnight), as this affects the quality of service we provide to all other customers. It also means that we are able to prioritise certain types of internet traffic on time-sensitive applications, such as our second line phone service or gaming." [10]
O2 fixed	Yes	Yes	Yes	P2P and streaming video allocated bandwidths according to tariff. No other traffic management specified on website
Plusnet fixed	Yes	Yes	Yes	11 traffic types independently treated through both a combination of prioritisation and rate limiting

Table 5: Summary of fixed ISP traffic management policies

[8] According to http://www.ispreview.co.uk/review/top10.php

[9] http://www.sky.com/helpcentre/broadband/set-up/sky-broadband-product-information/

[10] http://shop.orange.co.uk/broadband/broadband-explained#traffic-management

Explicit traffic management

The majority of fixed ISPs employ relatively simple policies which seek to mitigate the impact of 'heavy' or 'problem' users. The policies are based on a volume limit and some restriction of peer-to-peer (P2P). Most ISPs state that only a small number (typically 1% to 5%) of customers are 'caught' by this sort of traffic management. The rationale is always that the excessive use of resources by a minority is unfair on others – hence the policies often include the expression "fair use" in the title. In some cases there are 'application agnostic' usage limits set for each tariff. The lower priced tariffs have limits that would certainly be restrictive to some users, while the higher priced tariffs allow 'unlimited' use, subject to a fair use policy. In the case of Virgin, usage limits are tied to line speed.

There are two notable divergences from the norm among fixed ISPs.

- O_2 has bandwidth limits on both P2P and streaming video. These limits vary according to tariff. These 'limits' can also be read as guarantees.
- Plusnet, which was cited by several ISPs in their responses to Ofcom's consultation document on net neutrality, has a range of tariffs which differ according to the QoS offered on eleven traffic types. Two methods are used - traffic prioritisation and rate limiting. The details given for each tariff are the priority levels (described as platinum, titanium, gold & gold plated, silver, bronze, and best effort) and the rate limit, if applicable, in kbit/s at different times of the day.

The interviews revealed that there was some traffic management being applied which was not apparent from policies. For example, one ISP gives priority for gaming and VoIP traffic in order to improve QoS. Because of the traffic volumes involved, this would probably not have any detrimental effect on other users.

Implicit traffic management

No ISPs told us of any deals with content providers, content delivery networks (CDNs) or disclosed anything that would distort consumers' access to content. However BT wholesale has subsequently announced BT Content Connect which is a form of CDN. CDNs are not seen by ISPs as a form of traffic management. However they are promoted to content providers as a means to improve their connectivity with ISPs, thereby improving QoS for consumers. The fact that CDNs are not generally owned or controlled by ISPs explains their positioning within an ISP's frame of reference.

Partitioning

Managed IPTV services (e.g. BT Vision, Virgin Media Player, Sky Anytime Plus and the forthcoming YouView service) share the same physical access pipe as broadband internet traffic. Whether this affects usable broadband bandwidth depends on the service concerned. When such traffic is 'guaranteed', it limits the bandwidth available for 'best efforts' internet traffic within the finite limits of shared physical resources. Managed IPTV services are typically not included in ISPs' traffic management policies, and this sort of traffic often does not count towards volume quotas. In the case of ADSL we understand that ISPs will check a consumer's connection to ensure that there is sufficient residual capacity for internet traffic before allowing a managed service to be provisioned.

In interviews it was pointed out that IPTV may drive an incremental investment in bandwidth within ISP access networks, creating bandwidth which wouldn't be there if IPTV didn't exist. It was argued that the aggregate effect on other internet traffic would be minimal once this extra capacity is taken into account.

.

4.1.2 Mobile ISPs

In Table 6 we have summarised the published policies of a selection of the main mobile ISPs.

	Volume limits	P2P policy	Video streaming policy	Comments
T-Mobile	Yes	Yes	Yes	"If you exceed our fair use policy, you can still use the internet for the things you love most - like email, Facebook and news sites - and we won't charge you any extra. But we may restrict video streaming, peer-to-peer downloading and other things that affect other people's use of the internet at peak times." [11]
Orange mobile	Yes	See comment	See comment	"Orange may additionally manage customers' data connection at peak times to preserve the best experience for the greatest number of users" [12]
O_2 mobile	Yes	No	No	No other traffic management specified on website
Vodafone mobile	Yes	No	No	VoIP blocked under some tariffs
Virgin mobile	Yes	No	No	No other traffic management specified on website
Three	Yes	No	No	Three does not have a fair use policy but guards against excessive use through volume charging (except in the case of the One plan which has unrestricted access to the Internet)

Table 6: Summary of mobile ISP traffic management policies

Explicit traffic management

All mobile operators have volume charging and/or ‘fair use’ policies, though some have current or legacy tariffs which are headlined as ‘unlimited’. Where low volume limits (around 1GB) are in use, there is little need to differentiate between different traffic types in order to limit heavy users. However some fair use policies do make a distinction between traffic types, and will allow browsing and emails, but not video streaming, once usage limits are exceeded.

Some ISPs block VoIP under some tariffs; this is presumably to account for the potential for VoIP to cannibalise telephony revenue.

Implicit traffic management

No mobile ISPs told us of any deals with content providers, content delivery networks (CDNs) or anything that would distort consumers’ access to content.

[11] http://www.t-mobile.co.uk/shop/mobile-broadband/about-mobile-broadband/

[12] http://www.orange.co.uk/images/editorial/Orange-mbb-Animals-Terms-20101101b.pdf?linkfrom=%3C!--linkfromvariable--%3E&link=box_main_pos_1_1_link_1&article=termsofusemobilebroadbandcurrent

Partitioning

As the radio access layer carries voice in addition to broadband, the bandwidth available to broadband can be affected. None of the published policies give any insight into how bandwidth and priority is managed between voice and data.

4.2 Observations on the current situation

Currently, all fixed ISPs use some form of traffic management. Most use it in a minimalist way. They adopt a simple approach of having relatively high volume quotas and only actively restrict P2P. This is essentially an engineering response to a small minority of very heavy users. In principle there is a trade-off between expanding capacity and more extensive traffic management. We infer from ISPs' behaviour that, in practice, investing in capacity is considered preferable to investing in traffic management. Only where capacity expansion is particularly costly does traffic management feature more strongly in ISP strategies.

The mobile ISPs vary in their approaches. Some attempt to limit usage through volume quotas alone whereas others combine volume quotas with traffic type discrimination. While the original tariffs were often unlimited, reflecting the practice in fixed broadband, the rapid growth in penetration of smartphones mean that most ISPs now offer packages that differ in the volume of data allowed. T-Mobile appears to apply traffic discrimination only once the 'fair use limit' has been reached in order to continue to allow access to certain services even if the limit has been exceeded.

The technology exists to create finely differentiated services by applying different priorities and bandwidths to different types of traffic. In principle this allows QoE to be managed more directly, and tariff packages to be targeted to different user segments. Currently this approach is being used by Plusnet in the UK. In interviews with ISPs we explored whether similar approaches might be adopted more widely. The general view was that any market segmentation advantages would be offset by the increase in network complexity and cost, and the difficulty of communicating policies to consumers.

In interviews we did not find an appetite among ISPs for using substantially more traffic management.

While the net neutrality debate has sparked concerns about the growth of 'covert' traffic management, we found that the use of traffic management in the UK is reasonably overt. This is probably because traffic management is implemented in order to support a process of consumer behaviour change – e.g. reducing use of P2P, or encouraging consumers to trade up to a more expensive package.

None of the ISPs interviewed indicated that they treat traffic differently according to its source. Indeed, as yet, the content-supply side of the two-sided market is hardly developed. No deals between ISPs and content providers were disclosed, and while CDNs have been implemented, they are not being implemented by ISPs, and they are not seen by ISPs as traffic management *per se*[13].

[13] While BT Wholesale has launched Content Connect, a form of CDN, BT Wholesale is not strictly an ISP (ie it does not provide an internet service to end users).

4.3 Current approaches to transparency

All the ISPs we spoke to expressed a commitment to transparency. In practice, the policies can be difficult to locate[14], and some ISPs do not provide great detail in their policies. For example, while a policy may be clear that P2P is moderated or restricted during certain periods, the extent of the impact is not indicated.

This limitation in published information may be a result of the statistical nature of traffic, causing P2P to be restricted only as much as is necessary at any given time. We call such policies 'non-deterministic'. Whereas deterministic policies can be fully described in policies, non-deterministic ones cannot.

During the course of this project there has been some public discussion of T-Mobile's fair usage policies, prompted by the reduction in fair usage quotas. It appears from press comment that T-Mobile is relatively sophisticated in its use of traffic management to distinguish between, say, the content of an email message and an attached file. This level of sophistication is not described in the company's published policies.

[14] Our experience concurs with the mystery shopper exercise reported in the Ofcom discussion document http://stakeholders.ofcom.org.uk/binaries/consultations/net-neutrality/summary/netneutrality.pdf

5 Future scenarios for traffic management in the UK

Our interviews with ISPs suggested that they do not see a need for a radical change in their traffic management policies in the foreseeable future. The perceived problem of a small number of very heavy users consuming disproportionate capacity is largely being controlled by a combination of usage caps and restrictions on peer-to-peer traffic.

This situation may not persist, however. Video streaming is widely expected to grow substantially, and will need to be managed. Network capacity will continue to be added, and will have to be financed. The bandwidth for internet applications may be squeezed from both managed IPTV services and over the top services.

To focus attention on the ways traffic management may evolve we have developed five core scenarios. Apart from the first scenario, the other scenarios could, in fact, be combined.

1. Fair use Minimal, with limitations targeted at excessive users only
2. Traffic management evolves to facilitate congestion-sensitive traffic An engineering response to improve QoS for gaming, VoIP and live video streaming
3. Traffic management evolves in response to growth in video streaming Likely to involve a combination of restriction and monetisation of demand
4. Traffic management evolves as a business/marketing tool Promoting coordination across the two-sided market, using traffic management to define and implement new service packages.
5. Managed services become the norm Users' access connections are routinely partitioned to allow for managed services such as IPTV.

Table 7: Core traffic management scenarios

These five scenarios are discussed further below.

5.1 Scenario 1: Fair use

This scenario assumes that capacity expansion is considered in practice to be preferable to traffic management. This strategy is only likely to be feasible for those ISPs with a low marginal cost of capacity and/or the ability to finance capacity expansion through convincing plans to increase revenue.

5.2 Scenario 2: traffic management evolves to facilitate congestion-sensitive traffic

Current 'best efforts' delivery tends to produce poor QoS under congestion conditions. Some types of traffic – such as VoIP and gaming - are particularly sensitive to latency. Live streaming video is also sensitive to QoS. This scenario assumes that QoS-sensitive traffic (real-time) is identified and prioritised. Provided such traffic is not a large proportion of total traffic, it is assumed that there would be minimal offsetting impact on the non-prioritised traffic.

5.3 Scenario 3: traffic management evolves in response to growth in video streaming

Video streaming is widely regarded to be the next growth area for the internet. There will be wide fluctuations in demand and at peak periods video streaming will need to be limited.

Mobile ISPs may tackle this through volume quotas and/or specific limitations on video. On fixed line ISPs, simple restrictions similar to those applied to P2P are unlikely to be appropriate. This is for two reasons. Firstly, because video streaming is not associated with unlawfulness in the same way, it cannot be throttled back without an impact. Secondly, whereas P2P has been characterised by a small number of very heavy users, streamed video is characterised by a large number of moderate users. We expect ISPs to use a combination of restrictions and premium packages with guarantees. These are more likely to be in the form of bandwidth guarantees (eg an allowance of a certain number of HD streams) than a general de-prioritisation policy which would be unpredictable in its effects.

There is already dialogue between major content providers and leading ISPs regarding optimum coding rates for streamed video traffic. This coupled with further improvements in codecs (leading to better quality at lower bitrates) will allow existing networks to handle some increase in demand.

The architecture of different ISPs' networks affects what can – and cannot – easily be done. However we think that bandwidth guarantees for video will be easier to sell and communicate than differential priorities. The peaks in demand are not completely predictable, so policies may therefore not be fully deterministic.

5.4 Scenario 4: traffic management evolves as a business/marketing tool

This scenario puts traffic management in the hands of the marketeers, not the engineers. We envisage a combination of restrictions and paid-for services, monetising consumers' willingness to pay for content/service bundles.

Some hypothetical examples are:

- a content-led ISP offers a guaranteed QoS for its own content but only offers a 'best efforts' QoS for rival content;
- an ISP offers a tariff package that guarantees QoS within working hours (including 'professional P2P/file transfer applications') which might appeal to home workers;
- an ISP offers a priority service to a small number of very popular websites by caching these sites locally.

5.5 Scenario 5: Managed services become the norm

This scenario assumes that users' broadband connections are routinely partitioned to allow for managed services such as IPTV; this scenario features:

- a managed video conferencing service;
- pushed content – delivery outside peaks.

These five scenarios are used later in the report for assessing the types of consumer information and measurements required. Section 7.4 explores the information requirements by scenario.

6 The effects of traffic management on QoE

There are a number of almost philosophical issues that need to be addressed when looking at the effect of traffic management on quality of service and quality of experience.

Traffic management is not fully observable directly

From the forgoing technical discussion it is apparent that traffic management can act on traffic in different ways. These include:

- guaranteeing delivery of data or reserving bandwidth for that data;
- prioritising certain types of data in the event of queuing;
- de-prioritising certain types of data;
- restricting certain types of data or the bandwidth allocated;
- blocking certain types of data.

At an individual connection or device, a user cannot necessarily observe traffic management directly. He or she can observe the *performance* of an application and decide whether the performance is acceptable or not. If the application works as expected one can infer that the data have arrived in a timely manner. But it is impossible to tell whether the data have arrived only because they have been prioritised, or whether they have arrived because best efforts are perfectly adequate. Conversely, where the performance of an application suggests that data have not arrived in a timely manner it is impossible to observe directly whether this is the result of them having been deprioritised, or of congestion.

Despite the lack of certainty on the above points, a user may however make inferences based on the behaviour of applications compared to previous performance and possibly compared to other applications running at the same time. Some examples are set out below.

- A user may see a disparity between the apparent performance of different applications or websites, or between one user in the household sharing the same connection and another. Some applications are known to be more sensitive to QoS than others[15].

[15] In a previous study for Ofcom we were able to place applications into three categories according to how resilient they are to reductions in QoS:

- **QoS tolerant applications.** Downloading files from iTunes, the casual online multiplayer game, iPlayer Live, iPlayer SD and the on line interactive application continued to work well with packet loss of 1% and latency of 100ms. These represent the worst case conditions for a UK ISP except when there are network problems.
- **QoS sensitive applications.** Skype (VoIP) and YouTube worked well with packet loss of 0.25% and latency of 50ms. These represent average conditions for a UK ISP. Under worst case conditions, these applications continue to work but exhibited problems.
- **QoS critical applications.** The MMOG, iPlayer HD and the VPN started to exhibit reduced QoE with packet loss of 0.25% and latency of 50ms. These represent average conditions for a UK ISP. Under worst case conditions, the

- Users may find applications being blocked or not being usable due to reduced bandwidth. There could be subtle effects such as some types of email attachments not downloading properly whereas others do.
- A user may also see dramatic reductions in the overall performance of their connection compared with previous performance, perhaps because the data rate has been reduced as a result of crossing a usage limit.
- Users may observe variations in performance at different times of day, either because ‘peak hour’ traffic management is being applied to certain services, or because of simple congestion.

It follows that a user can observe performance of applications and infer the role of traffic management in producing that performance, but not know for certain. The status of traffic management being applied to the network or to a user’s connection can only be known for certain by the ISP. The more prevalent traffic management becomes, the more consumers will tend to explain performance problems with reference to the use of traffic management. This implies the need for diagnostic tools to help users understand whether and in what way traffic management is affecting them.

Users will tend to be more alert to ‘negative’ effects of traffic management, such as applications performing poorly, than they will to ‘positive’ effects, which will tend to be taken for granted.

Traffic management on a user’s connection versus traffic management in the network as a whole

There are debates over whether the effects of traffic management are positive or negative. To help bring clarity to this debate we have drawn the distinction between traffic management happening to other peoples’ traffic (the network as a whole), and traffic management on one’s own data (see Figure 8).

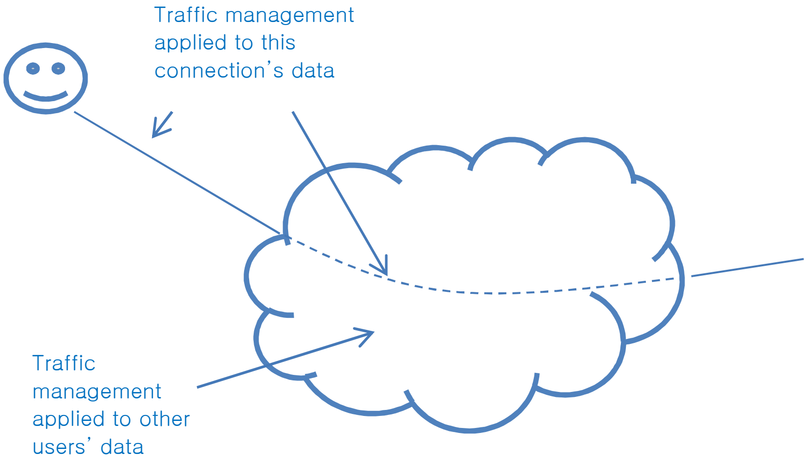

Traffic management applied to this connection’s data
Traffic management applied to other users’ data

Figure 8: Traffic management of own and other users' data

iPlayer HD and the VPN were unusable according to our coding of QoE, and the MMOG exhibited noticeable problems.

Traffic management can have both positive and negative effects, as Table 8 shows. While the lower row in this table summarises the effect of traffic managing other peoples' data, measuring the effects would be virtually impossible[16]. We have therefore restricted our discussion to the effect on QoE of the traffic management applied to a user's own traffic.

	Positive effects on QoE	Negative effects on QoE
Traffic management applied to a user's own traffic	Can guarantee or prioritise data for sensitive applications	Can restrict or block certain applications
Traffic management applied to other people's traffic	Can reduce congestion to manageable levels, allowing fair use for all	Other people's traffic may take priority

Table 8: Positive and negative effects of traffic management

As stated above, the positive effects are unlikely to be as easy to observe as the negative effects, and unfortunately this has the effect of painting traffic management in a more negative light than perhaps it ought. Traffic management is justified by ISPs in terms of fairness. They consider that a relatively small number of users consume a disproportionately high share of resources, and that statistically more users benefit from traffic management than are disadvantaged.

[16] Even with perfect information on the traffic management is being used in the network as a whole, it would be logically impossible to determine the effect of this traffic management on a specific user.

7 User information requirements

The purpose of this chapter is to identify the information that consumers are likely to want.

7.1 When and why would consumers want information on connection performance?

In order to identify specific information requirements we have considered the situations in which consumers might want information on their internet connection performance. There are two basic consumer information situations, "prospective" and "in use", as shown in the flowchart (

Figure 9) and described in the following text:

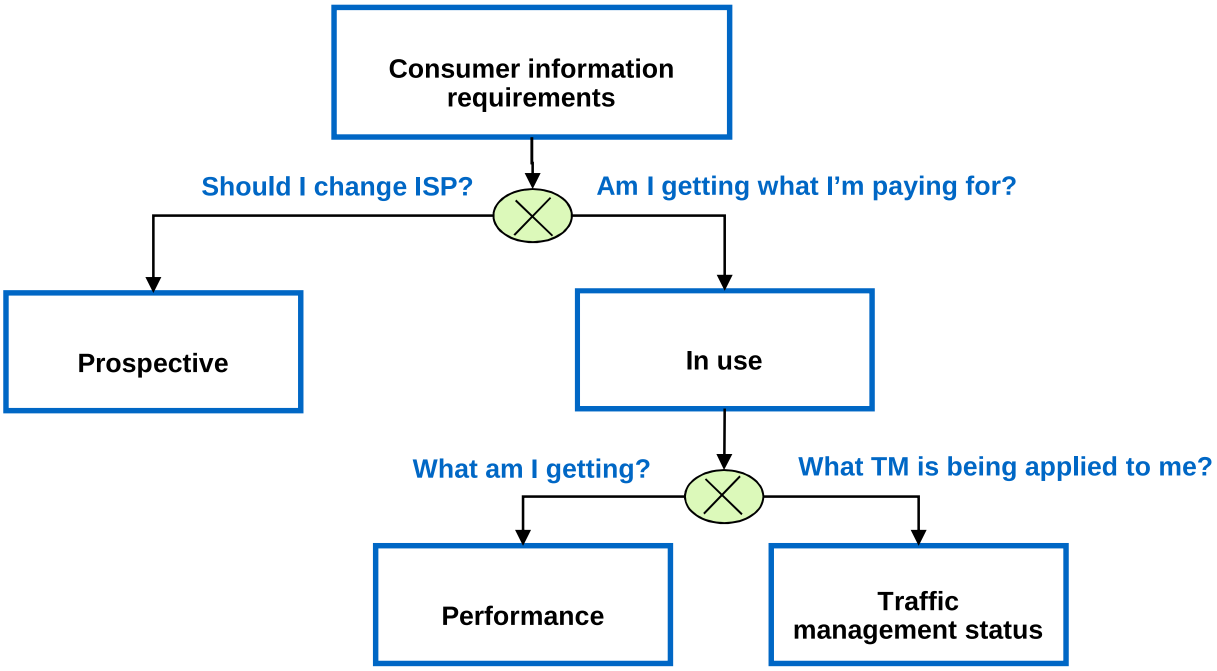

Consumer information requirements
Should I change ISP?
Am I getting what I'm paying for?
Prospective
In use
What am I getting?
What TM is being applied to me?
Performance
Traffic management status

Figure 9: Consumer information contexts

Prospective: In the situation where consumers are deciding between ISPs or between packages, they will be making choices where at least one of the options is a prediction. The information required is whatever is necessary to check the suitability of a package, estimate costs and make comparisons of performance.

In-use: A consumer may wish to have information about the performance and the status of an existing connection. There are two classes of information in this category as explained below.

- **Performance checking:** Consumers may want to check at any time whether an ISP is delivering what was promised, or they may want to diagnose an apparent problem. Performance checking is always based on QoS measurements.

- **Traffic management status checking**: If traffic management becomes more prevalent consumers will increasingly want to determine whether they are being traffic managed. And if usage-based tariffs become more prevalent then consumers will want to find out how they stand relative to caps and thresholds.

7.2 Types of information: pull and push

We can distinguish between two types of information:

"**Pull":** that which is requested by the consumer about an ISP or several ISPs services. "Pull" information requests are initiated by a consumer, with the prime functions of enabling consumers to check: (a) how their connection is performing at the time of the information request; (b) whether any generic (network level) traffic management is in place at the time of the information request; (c) whether any user-specific traffic management is in place at the time of the information request; (d) service levels and costs of different services from different ISPs, to enable an informed choice of provider; and

"Push": that which is pushed by an ISP to customers of a service about their current service, or an alternative service which may better address the customers' needs. "Push" information sessions are initiated by an ISP, and can carry information relevant to: (a) the network as a whole; (b) a customer's connection performance; (c) a customer's service usage; and (d) whether customers are on the most suitable package for their service usage. "Push" communications either enable an ISP to instruct/ control/ persuade a customer to behave in a certain way, or to appear helpful and open to the customer.

In section 7.3, we specify whether each of the items of information consumers might want about their connection performance could be served by Pull, Push or either mechanism.

7.3 What information would consumers want on connection performance?

Below we present a list of what kind of information consumers may need to know about their connection performance, split across prospective, performance and status. For each information item, we also state whether it could be met by Pull, Push, or both types of communications. This list is based on experience and logical analysis as we have not conducted empirical research into consumer concerns. We cannot vouch for the prevalence of each question. The numbering of the questions is for reference purposes and does not imply any priority order.

Prospective Information

1. To know which internet applications and services will be guaranteed (Pull);
2. To know which internet applications and services will work (Pull);
3. To know how well particular internet applications and services will work (even for internet applications and services known to be particularly sensitive to latency or line speed) (Pull);
4. To be able to anticipate which internet applications and services are unlikely to work reliably or at all (Pull);
5. To know which internet applications and services will be limited/restricted, and the details of such restrictions (Pull);
6. To know which internet applications and services will be blocked (Pull);
7. To estimate total costs of obtaining a service level to meet usage requirements (Pull);
8. To judge what is the best package/ service level to meet usage requirements considering cost (Both Push and Pull).
9. To judge the effect of managed services on the broadband connection (Pull)

Performance Information

10. To understand whether ISP is delivering what being is paid for at a discrete time (Pull);

11 To understand which network characteristic (QoS) may be stopping any internet application or service working when they do not work (Both Push and Pull);

12. To know whether the problem can be bought/upgraded around (Both Push and Pull)

Status Information

13. To understand which traffic management function may be enabling or stopping any internet application or service working when they do or do not work (Both Push and Pull);
14. To know whether the problem can be bought/upgraded around (Both Push and Pull)

Figure 10: Detailed consumer information requirements

7.4 Information requirements by scenario

In the table below we list the information types (those listed above) which could be useful or relevant to consumers in each of the five future traffic management scenarios (as presented in Chapter 5).

	Scenario				
	Scenario 1 Fair use	**Scenario 2** Congestion management	**Scenario 3** Video streaming	**Scenario 4** Business tool	**Scenario 5** Managed services
Prospective Information					
1 What's guaranteed?	✓	✓	✓	✓	
2 What will work?	✓	✓	✓	✓	
3 How well will it work?	✓	✓	✓	✓	
4 What will not work	✓	✓	✓	✓	
5 What's limited?	✓	✓	✓	✓	
6 What's blocked?	✓	✓	✓	✓	
7 Cost?	✓	✓	✓	✓	✓
8 Best for me	✓	✓	✓	✓	✓
9 Effects of managed service?		✓	✓	✓	✓
Performance Information					
10 As promised?	✓	✓	✓	✓	✓
11 What QoS is wrong?	✓	✓	✓	✓	✓
12 Can I buy better?	✓	✓	✓	✓	✓
Status Information					
13 What traffic management in action?	✓	✓	✓	✓	
14 Can I pay around?	✓	✓	✓	✓	✓

Table 9: Summary of relevance of information in the five traffic management scenarios

As can be seen from the table, two points are of particular note. First, nearly all the information items may be of interest to some consumers in most of the scenarios. Second, while there are some differences, the scenarios do not differ substantially in terms of the information that may be required. The measurement implications of the above analysis are picked up in Section 9.3.1.

8 Designing a communication approach

8.1 Principles of information transparency

Through our research activities and from synthesising our findings, we have identified three general principles which should underlie any transparent communication about traffic management, so that it is:

- **meaningful** (e.g., have utility),
- **accurate** (e.g., be valid/true), and
- **comparable**.

While these concepts have been derived from the team's understanding of what consumers need, based on experience of consumer decision making in other contexts, similar concepts have been identified and discussed in current academic and applied research literature[17].

Two of the dimensions (meaningfulness and accuracy) are extensively referenced in the literature on information quality. As described below, given sufficient knowledge on what information is needed for and by whom, in the majority of instances it is relatively straightforward to generate guidelines or rules to support the dimensions of meaningfulness and accuracy for transparent communication.

In the context of information about traffic management, comparability as a dimension of transparency is a somewhat more complex principle. Comparability will become increasingly dependent on user context as more complex traffic management regimes are implemented. In essence, to make informed comparisons based on multiple complex rules requires the processing of a lot of rules simultaneously. Solutions to such difficulties in comparability can include the development of representative user contexts or scenarios (e.g., "average family household", "online gamer"), or automated advisers or wizards to recommend a service offering best suited to different usage criteria input by a user.

Here we discuss each of the three dimensions introduced above in more detail.

Accuracy

Any information provided to consumers should be as accurate as possible, whether in a written policy, or real-time status or usage updates. In complex domains with simultaneous variation in multiple dimensions, perfect accuracy can be impossible – there will always be some irresolvable error. In fact, there is always a trade-off between the cost and benefit of seeking accuracy. In the domain of information about the performance of an internet connection and effects on that performance of any traffic management, there are two potential sources of inaccuracy.

- **Predicting performance of access into an ISP**. There is a potential source of inaccuracy in relation to information about the quality of service any particular phone line is able to support. Without performing a test of line quality, it is impossible to know whether a specific

[17] references: http://www.epa.gov/oei/symposium/2010/worthington.pdf; http://asbbs.org/files/2010/ASBBS2010v1/PDF/C/Caratelli.pdf; http://www.google.co.uk/url?sa=t&source=web&cd=13&ved=0CCwQFjACOAo&url=http%3A%2F%2Fpublications.accion.org%2Finsight%2FIS24EN.pdf&rct=j&q=dimensions%20of%20transparency%20of%20consumer%20information&ei=r1BLTcGrAsOxhQeA9vy_Dg&usg=AFQjCNFFwVXPdS8I6V10R_knvBV7hO8jZg

ADSL connection will be able to support, say, an 8Mbps connection, or substantially less bandwidth. There is therefore a limit to the accuracy of prospective information about what services can be supported for a specific customer. Descriptions of the statistical distributions of what services an ISP's customers receive are of course possible, in general terms. But these are different to direct, specific predictions about a prospective customers' line.

- **Forecasting**. It is not possible for an ISP to predict with perfect accuracy how congested its network will be in the future. Given this, where traffic management is triggered by congestion it is not possible to know what levels of traffic management will be in operation at any time of the day, or the exact impact of the traffic management, in advance of its operation and impact. Despite the occurrence of big unexpected events, past performance is as good a guide as there is, and enables intelligent predictions.

Additional accuracy in predictions of performance of access into an ISP, and in forecasting, may be technically feasible, for example through the addition of probes to the network (see section 9.2), but the costs and benefits of obtaining such increased accuracy need careful consideration.

In the context of this discussion, considerations of accuracy as a dimension of the transparency of communications to consumers about traffic management and its effects must centre on whether:

- available measures are communicated honestly;
- the output of future forecasting is communicated honestly;
- complex statistical information is represented meaningfully to users..

Meaningfulness

Any information provided to consumers about traffic management, whether it describes a policy, the effects of implementation of a policy, or a usage allowance, should be meaningful to the target recipient. Important elements to consider in ensuring information is meaningful, are:

(a) that it is **relevant** to the consumer's situation (i.e., that they will be motivated to access the information, a feature of the "what"); and

(b) that it is represented using **easy to understand units, concepts and terminology**, appropriate to the technical literacy of target recipients (the "how").

In relation to relevance, a key consideration relates to what consumers want to use their internet connection for. We assert that consumers generally use the internet to access specific functioning applications and services (for example: websites, video streams, games, social networking sites, email, social interaction, VoIP and so on). This assertion in relation to relevance provides some useful indications or pointers in relation to the second component of meaningfulness, the use of appropriate terminology, units and concepts with which to communicate.

Target recipients of information (about what services are supported, how well and for what price by different ISPs) obviously constitute a diverse audience, with a broad range of technical literacy. Given this broad distribution, a challenge for ensuring communications are meaningful is getting the right balance between excessive simplicity and baffling complexity. A related balance to achieve is between providing too little and too much information. An effective way to address these challenges is to provide different levels of information to different types of user. For example, one view could provide the information that any non-expert user would want/need, with an alternative detailed/experts' view available for those interested. Using this approach, a small minority of people may need additional assistance in understanding the information. This solution is not ideal, but it may ensure optimal meaningfulness for the most people.

With regard to the impact of traffic management on QoE, the most meaningful information possible to communicate to interested users is whether they can use an application or service they wish to; and, if the answer is uncertain, to inform them of their options to ensure access. This would tend to suggest communicating directly about specific applications and services (such as video, photos, video conversations, gaming) in meaningful units (e.g., time – hours, minutes; or units – for example episodes, calls).

It is of course the case that technical literacy develops over time in a population, whether incrementally through the accretion of technical knowledge or experience, or in generational step-changes. As technical literacy develops it is important to ensure that communications remain meaningful to target recipients. It is worth noting here that making metering transparent can be an effective means of supporting the implicit learning of the resource consumed by a device or activity.

Comparability

For information provided to consumers about traffic management to be of use in their decision making requires that different ISPs provide comparable information, i.e., information about the same (potential) variables or features of a service, in comparable units, concepts and terminology. In relation to comparability, a key consideration relates to whether any or all ISPs are able to provide the required information to enable an interested consumer to compare ISPs. Another key consideration relates to the volume of data presented – less information can be easier to digest and compare, and therefore more transparent.

- There are some limits to the comparability of information about traffic management. For example, as described earlier in this report, some ISPs deploy traffic management at specific times of day, whilst others use monthly usage caps. To compare the likely impact of these policies directly is not easy – as there is no automatic translation between the two dimensions.
- To address the fact that different approaches to traffic management are used by different ISPs, the following are potential approaches to support comparability:
    - provide the basic information but do not force any comparability
    - force commonality of practices (data/month or hours/day) to enable direct **price/ quality** comparisons;
    - use generic scenarios to create synthetic comparables (e.g., a typical household) on **price/quality**;
    - personalised scenarios (based on your data/service use over time – this is what it would **cost** you with ISP A, B C...);

Understanding consumer preferences is also important in supporting comparability. Passive consumers may want to be told which suppliers can meet their needs at the best prices. More active/ informed consumers may have a rough wish list and want to evaluate potential trade-offs they can make to get as near to obtaining their wishes as possible whilst saving as much money as possible.

8.2 Prospective information approach

The requirement for prospective information can be met by three elements as listed and detailed further below.

- A QoS Policy Form which gives a complete description of the policies in operation and performance data where policies alone are inadequate'
- A QoE Summary which gives meaningful and comparable information for the popular services'
- A 'wizard' which is capable of helping a consumer choose packages that are appropriate to their needs.

8.2.1 QoS Policy Form

The QoS Policy Form characterises the connection as a whole, the baseline for ordinary traffic, and specific information on any traffic type that is separately managed.

With respect to the policy, the options range from giving a general impression ("We may additionally manage customers' data connection at peak times to preserve the best experience for the greatest number of users") through to being completely specific ("Between 4pm and 11pm your bandwidth for streaming video will be limited to 1Mbps").

For a policy to be 'transparent', it should ideally inform the user precisely what the performance of the connection will be. Some forms of traffic management – such as restricting bandwidth - do allow for predictability. Unfortunately, this is not always possible because performance of a network is generally dependent on the level of traffic, which is not completely predictable.

The default treatment of traffic is 'best efforts'. That is, the ISP attempts to deliver a packet but makes no guarantee to do so, especially if there is congestion. The policy should specify all departures from this default. Thinking ahead to more complex traffic management policies than we currently have, all policies involve applying a traffic management action either unconditionally or in a way that is triggered by some other factor. The policies should be as specific as possible in both describing the condition and describing the traffic management action and its effects.

Fortunately, the technical survey of traffic management suggests that there are relatively few different types of traffic management and that the different options can be represented using a common template as shown below.

For the overall connection, the table below sets out essential information.

Data rate	
Achievable rate, taking into account as much of the user's situation as is reasonable *	

Volume	
• Unlimited	Choose between these and provide the information
• Capped (specify the cap , the alerting procedure, and the consequences of going over the cap)	

For each traffic type, the following must be described. The two techniques of traffic management are covered. Priority level' refers to packet prioritisation and 'data rate' refers to bandwidth allocation.

Traffic type	
Specify the type of traffic	

Priority level	
• Guaranteed	Choose between these and specify the criteria (e.g. times of operation)
• Accelerated (specify criteria) *	
• Normal *	
• Restricted (specify criteria) *	
• Blocked	

Data rate	
• Guaranteed minimum (specify)	Choose between these and specify the data rate in both bps and indicative units of consumption (e.g. type of video supported)
• Maximum, not specially limited	
• Capped or restricted (specify)	

Volume caps	
• Unlimited	Choose between these and specify the data rate in both bps and indicative units of consumption (e.g. hours of video)
• Capped (specify the cap , the alerting procedure, and the consequences)	

Historic data	
If starred above, provide performance graphs based on past measurements. For a new service the performance graphs must be estimated.	

In the above table:

- All classes of traffic not treated on a best efforts basis should be identified.
- There is no requirement to describe how traffic is identified but the description should be sufficiently complete that a user would know whether their traffic is included.
- For each traffic class, the way in which it is handled should be described in as much detail as possible.
- There is no requirement to provide absolute certainty where the impact is affected by demand – which is not fully predictable.
- A reasonable level of certainty should however be offered through historic data or models of expected performance.
- Data on the QoS achieved is more meaningful than data on the level of traffic management employed.
- Where the use of one service (e.g. managed IPTV) affects the use of other services, then the dependency should be explained.
- Particular attention must be paid to services which are, in practical terms, blocked.

We have illustrated an example QoS Policy Form in Figure 11.

ISP Tariff Example

Data rate	5 Mbps
Volume	40 GB/month Email warning at 75% and 90% of limit. Above cap, user can choose to pay a supplement or upgrade. Residual 100kbps is still provided to allow browsing and email

Traffic type	P2P Recognised P2P such as BitTorrent, eMule, Gnutella and newsgroup applications
Priority level	Restricted Between 6pm and 12pm, P2P on our network is deprioritised.
Data rate	Capped All users apart from our 'Mega' tariff are restricted to no more than 2 Mbps upload speed
Volume caps	None
Historic data	See data rates as measured for 3 months Jan 2010 to Mar 2010

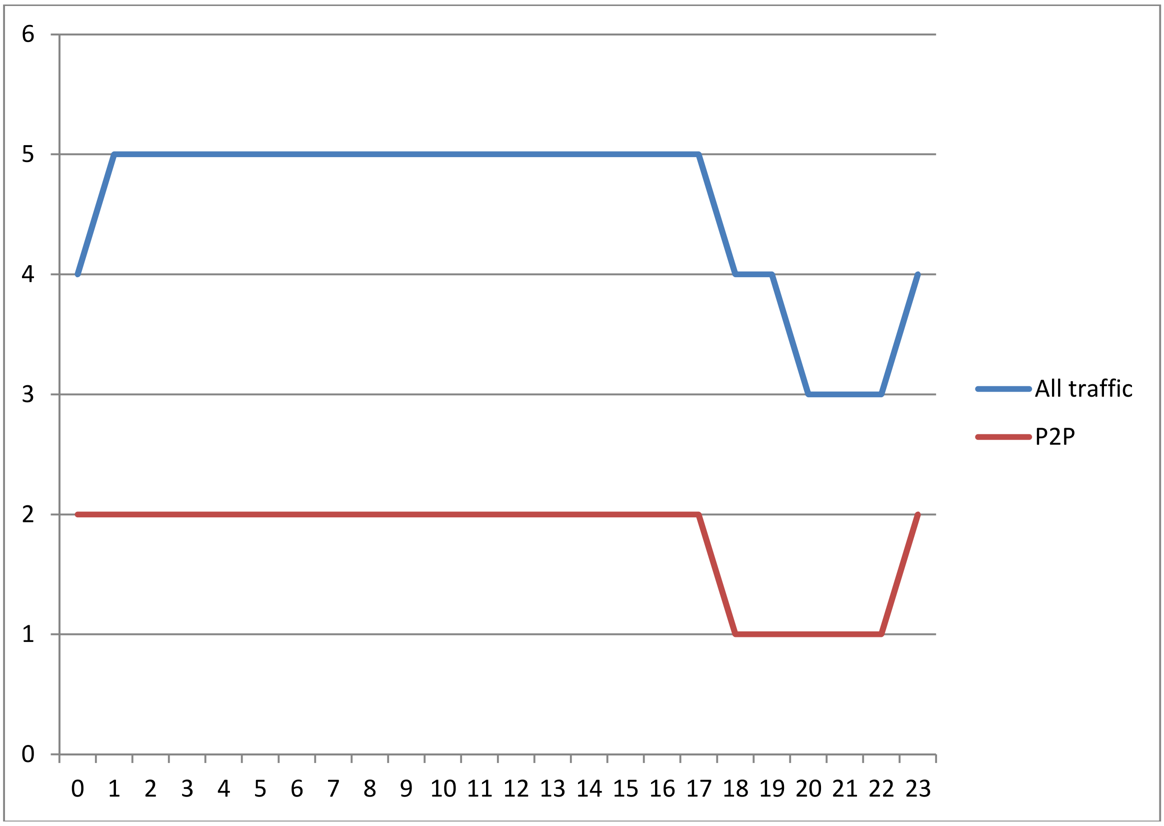


Figure 11: Illustrative example of a QoS Policy Form

8.2.2 QoE Summary

The requirements for each type of information (prospective, performance and status) can be described by different sources, with the two in-use information types most amenable to direct display of current status. In the case of performance, this can be described by real-time quality of service measures (of bandwidth, latency, jitter and packet loss) – which is effectively an aggregate of the inherent network QoS with any QoS effects of traffic management overlaid, at the time of measurement. In the case of status, this can be described simply as a summary of what traffic management, if any, is being applied to a particular connection (or the network as whole) and what proportion of any periodic allowances have been used at the time of measurement.

The most complex question is how to represent prospective information transparently, given the range of relevant variables of which it is constituted.

Transparent representations of Prospective QoE information

Whilst, as discussed further below, we anticipate the development of any consumer-facing representations and graphics describing prospective QoE information to be delivered by the market, according to criteria agreed collectively by ISPs, we have demonstrated in the figure overleaf how it is possible to categorise the service levels of different (hypothetical) ISPs transparently (so that they are meaningful, accurate and comparable).

In the examples here, several service groupings relevant to consumers are represented: Standard Definition video streaming, High Definition video streaming, online gaming, VPN, video conferencing over IP, voice over IP, music streaming, music downloads, video downloads, P2P file sharing, and day to day online activities, labelled e-life: comprising web search/ browsing and transactions.

We have developed, and present here, simple logos – not as suggestions for formal communications, but to test whether to do so is possible. The logos are shown in Figure 12. Logos/graphic representations are presented for downloading, streaming and P2P data methods, and to illustrate where a service's bandwidth is throttled. We also present logos/ graphic representations for a number of service groups: video, audio, gaming, conferencing, P2P. And we suggest how usage limits or caps could be communicated, again in meaningful and relevant units to consumers. The illustrations are not meant to be exhaustive nor complete, but a demonstration of what is feasible.

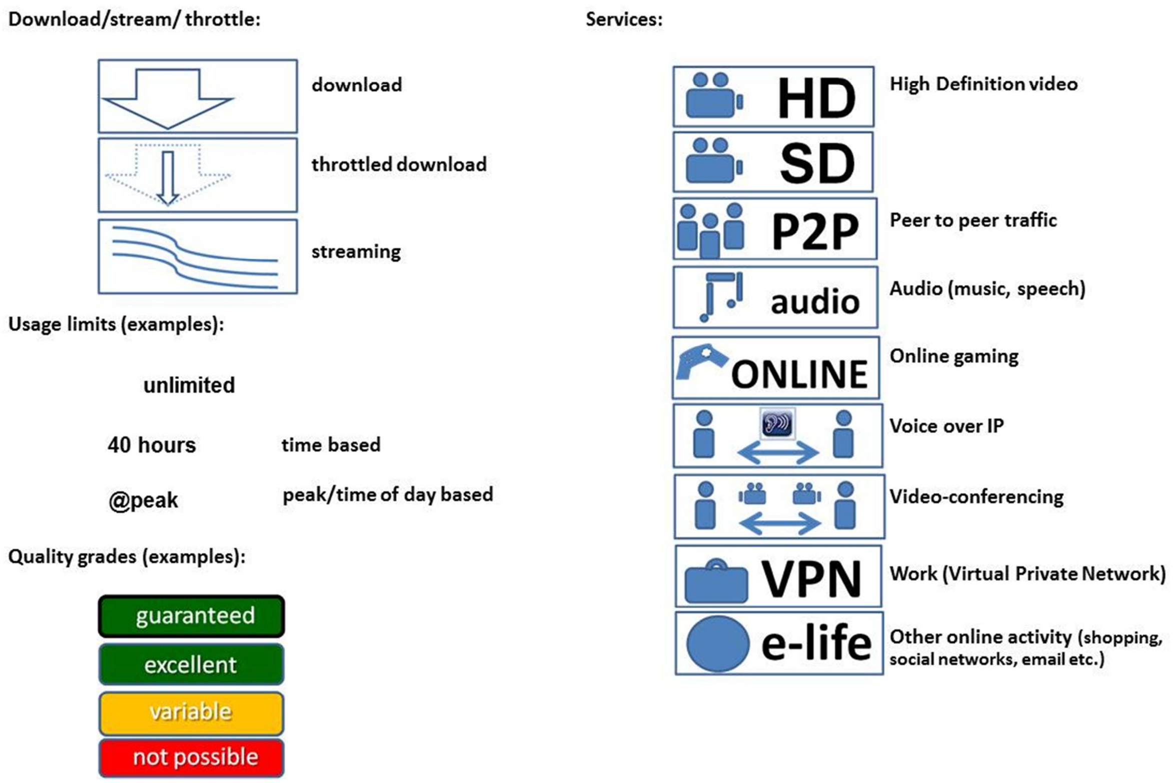


Figure 12: Key to logos

In Figure 13 we have represented the services for three hypothetical ISPs using these logos. We have then grouped the services according to whether the services are: (a) guaranteed (effectively managed services), labelled ‘guaranteed’; (b) strongly expected to work well, labelled ‘excellent’; (c) expected to fail at some points of the day (typically during peak contention and congestion), labelled ‘variable’; or (d) are blocked or firmly expected not to work, labelled ‘not possible’. The three hypothetical ISPs have different package profiles which are used here to illustrate the use of the logos to produce meaningful graphic representations.

The universal implementation of a similarly transparent approach should allow non-expert consumers to identify easily whether a specific ISP’s service/tariff is likely to meet their needs, and to compare at a glance between different ISPs/tariffs. These are most meaningful if the majority of the QoE seen by the consumer is the result of deliberate actions by the ISP, or limitations known to the ISP (e.g. bandwidth restrictions). If the QoE is dominated by other factors, such as radio propagation for a mobile consumer, then the graphic won’t be such a good guide to performance.

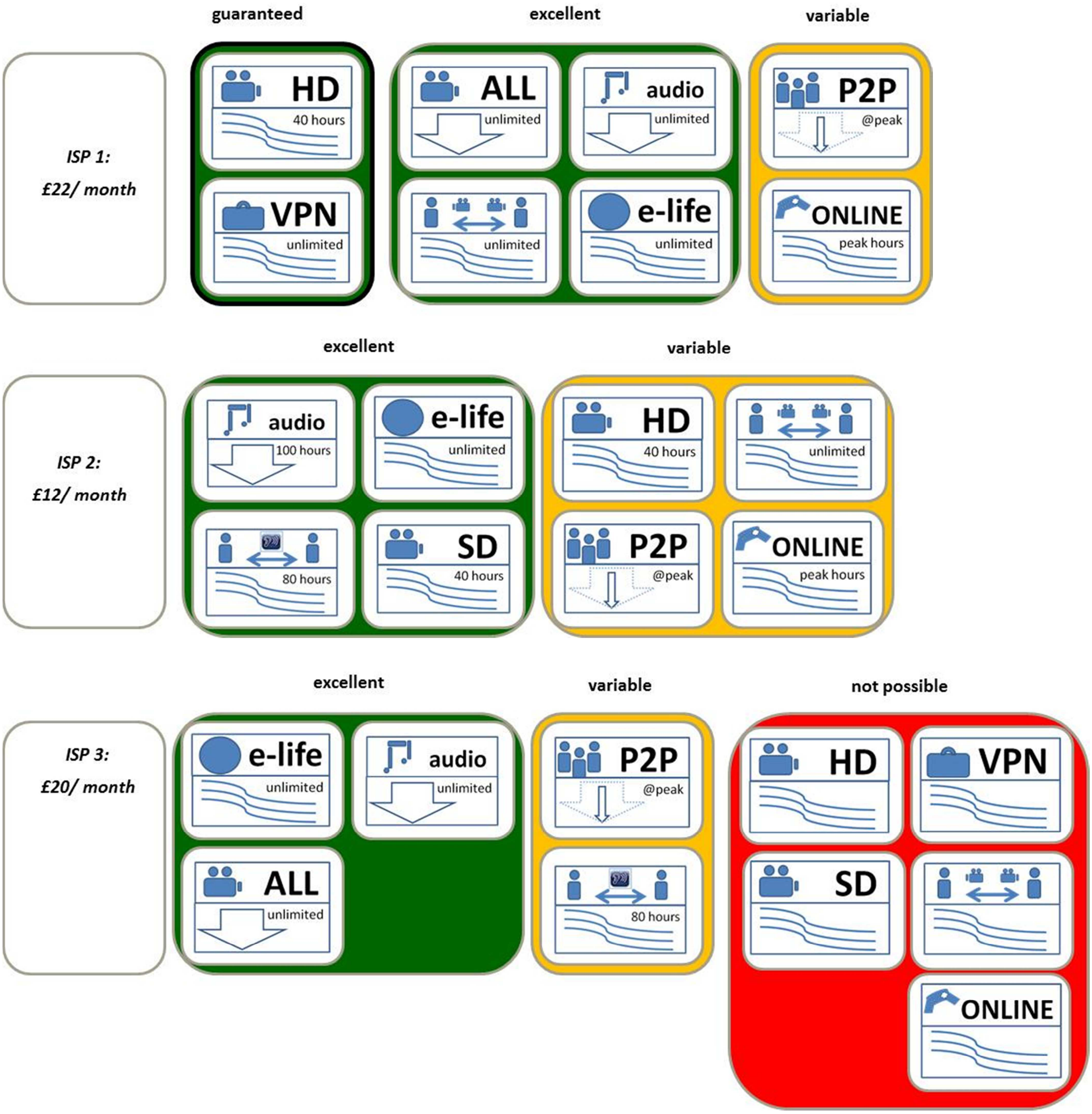


Figure 13: Illustrative QoE Summary for 3 hypothetical ISPs

8.2.3 Thresholds

The implementation of an information provision system such as that outlined above of course requires QoS thresholds to be agreed and set in order that the information provided in such a summary table is as accurate as is possible. Consideration will need to be given in particular to the thresholds between the categories “excellent” and “variable”, and between “variable” and “not possible” (where “not possible” is because of a QoS limitation rather than traffic management policy based blocking). A reasonable approach could be that responsibility for agreeing and setting these thresholds be devolved to ISP industry bodies such as the Broadband Stakeholder Group.

8.2.4 Wizard

For some consumers, a wizard approach may be preferable to the display of a lot of information. Here, consumers enter data describing their usage of services and a wizard recommends an ISP service package appropriate to these needs. A wizard may be most relevant in scenario 4. The design and implementation of any wizard is, again, best left to the market but in order for the wizards to be useful all ISPs must make data available describing the services supported by any of their broadband packages.

8.2.5 Applicability to scenarios

All five of our scenarios can be represented using the QoS Policy Form and the QoE Summary. The more complex traffic management scenario (4) may additionally require the use of a wizard to help people choose. In all cases, we think that ISPs could be asked to agree to provide data to third party wizards if requested.

8.3 Performance and status information approach

8.3.1 Performance

Once threshold QoS parameters are agreed for any service to work, translating a real time network performance test to a working/not working indication for any service is a relatively straightforward task. In addition to the straightforward QoS information, a potential transparent approach to communicating the results of a QoE performance check is shown in Figure 14 below.

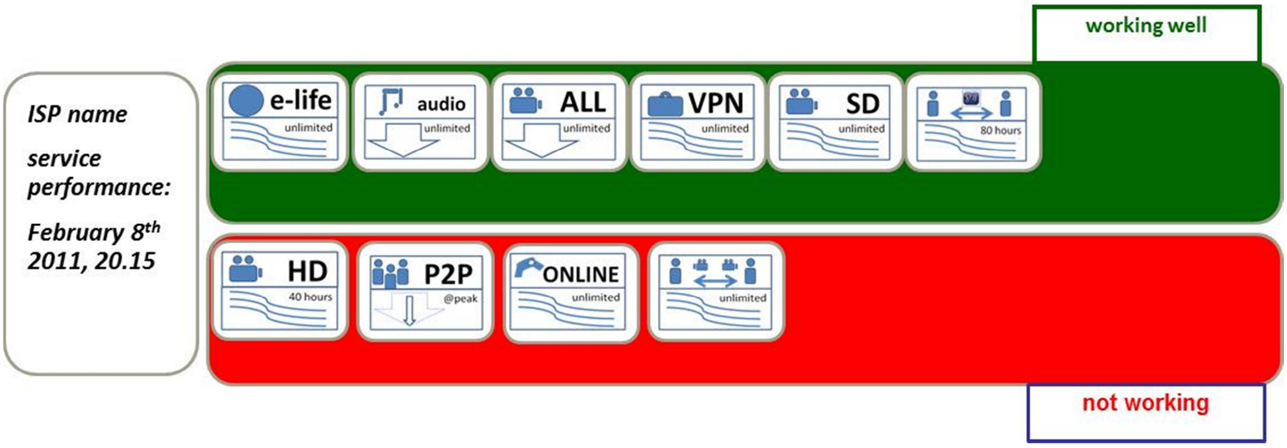


Figure 14: Illustrative transparent QoE performance representations

8.3.2 Traffic management status checking

We defined traffic management status checking above as a process enabling consumers to find out how they stand relative to usage caps and thresholds, and, if possible, to check what traffic management is being applied at the time they make the check.

A potentially transparent approach to communicating the results of a traffic management status check is shown in Figure 15. Whilst simple text descriptions of time remaining are used here, many similar representations could work (e.g., egg timer, pie chart, % used illustrated on a bar) as would simple text statements describing usage relative to caps and thresholds.

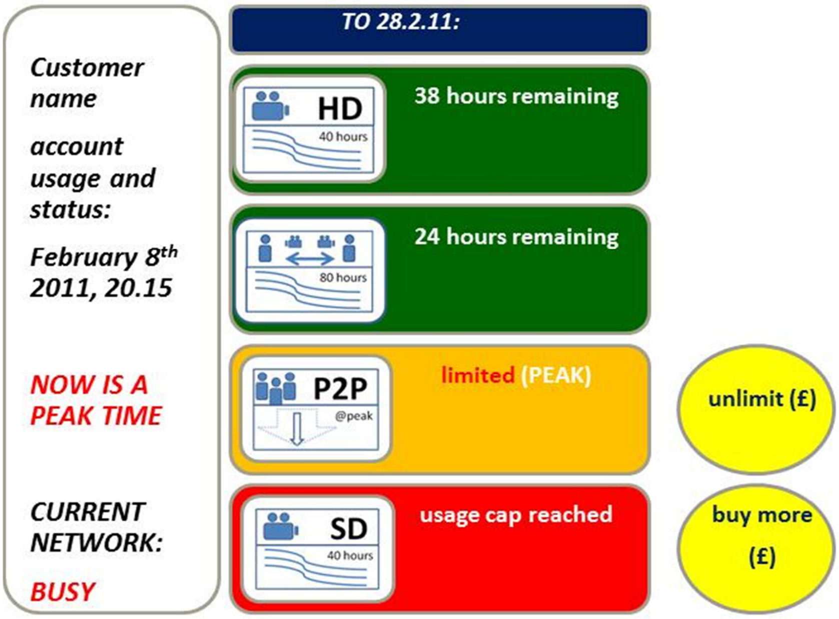


Figure 15: Illustrative transparent traffic management status representations

It is also worth noting that traffic management status updates are those most amenable to "Push" information sessions initiated by an ISP. Examples of traffic management status which could be communicated by an ISP via Push notifications include information relevant to: (a) the network as a whole; (b) a customer's connection performance; (c) a customer's service usage; and (d) whether a customer is on the most suitable package for their service usage.

9 Measurement approaches

9.1 Introduction

This chapter examines the options for providing the data necessary to support the user information requirements introduced in chapter 7 and developed in chapter 8.

Figure 9 in chapter 7 described the reasons why information might need to be provided. In Figure 16 we have identified that the need for prospective information gives rise to the need for time series data capable of predicting future performance, and real time data capable of being used to check performance and the status of a user's usage relative to any caps that may exist.

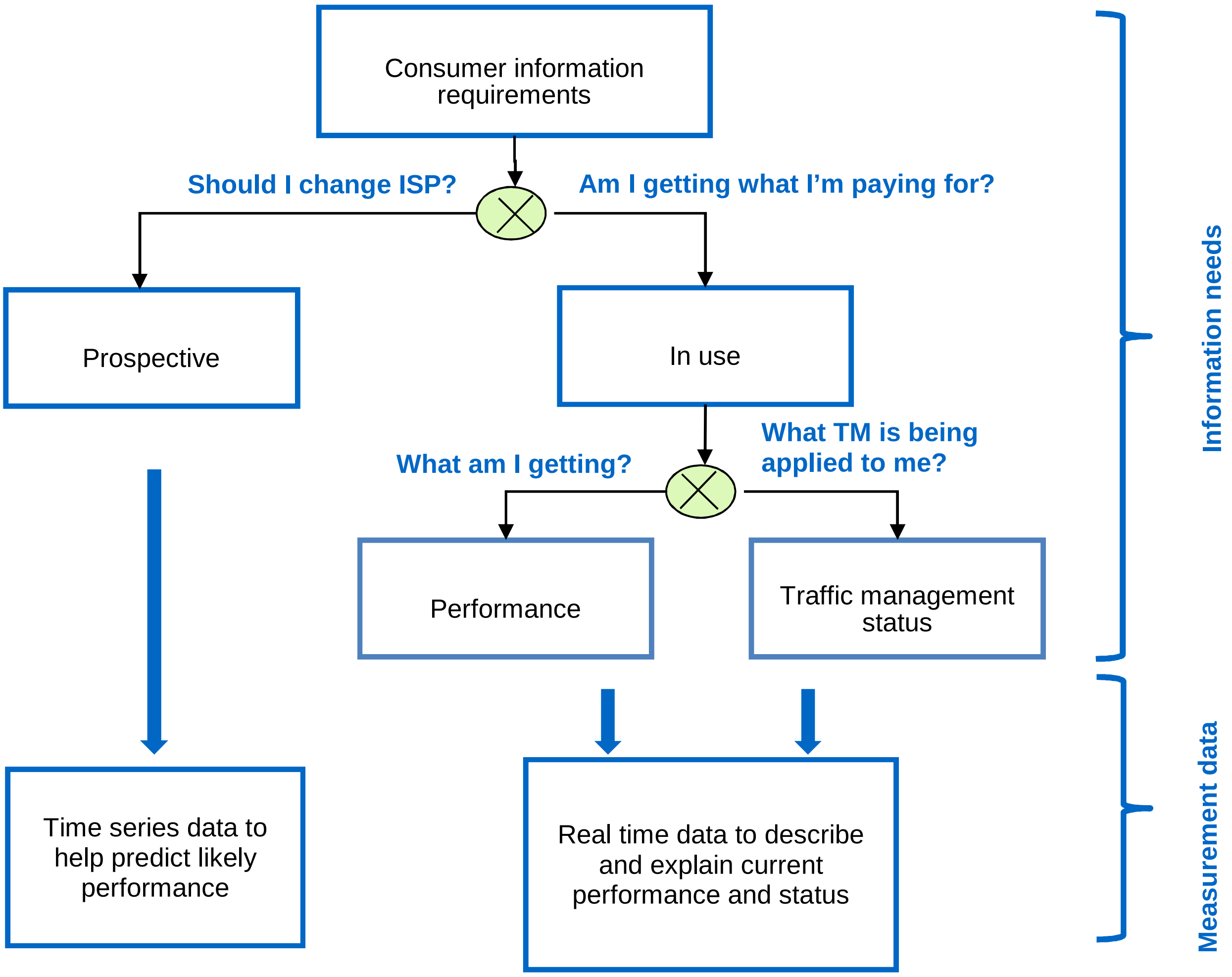

Consumer information requirements
Should I change ISP?
Am I getting what I'm paying for?
Prospective
In use
What am I getting?
What TM is being applied to me?
Performance
Traffic management status
Time series data to help predict likely performance
Real time data to describe and explain current performance and status
Information needs
Measurement data

Figure 16: Measurement data requirements arising out of user information needs

It is helpful to consider the two types of measurement data separately, though in practice it is possible that time series data could be generated by collecting and aggregating the real time data.

9.2 Measurement options

There are a number of different dimensions to measurement and we start by discussing these.

Method

Quality of service can be measured either on existing traffic (typically called ‘passive’ measurement) or on specially injected traffic (typically called ‘active’ measurement). For the purposes of detecting traffic management at a user’s connection or device, the active approach of injecting traffic allows the QoS of specific types of traffic to be detected. However, if traffic volumes are capped, such an approach will eat into allowances.

Location

There are many places within a network and its attached devices where measurements can be undertaken. Principally, these are:

- the user’s device, typically a PC or phone;
- a special purpose monitoring device at a user’s connection;
- within the ISP’s network: at nodes or line cards - network equipment will often incorporate the ability to produce traffic statistics[18] though this ability may not in practice be used;
- at a content provider.

Temporal

The options range from occasional spot tests through to continuous monitoring in the background. Some equipment is able to detect quiet periods at a user’s connection and run tests at that time.

Initiation

Tests can be initiated or controlled by a user, an ISP or a third party.

Having researched the field and spoken to ISPs we consider that there are five main options which are illustrated in Figure 17 and summarised in Table 10.

[18] This DSLAM incorporates stats - http://www.huawei.com/broadband_access/products/dslam/ip_dslam.do?card=2

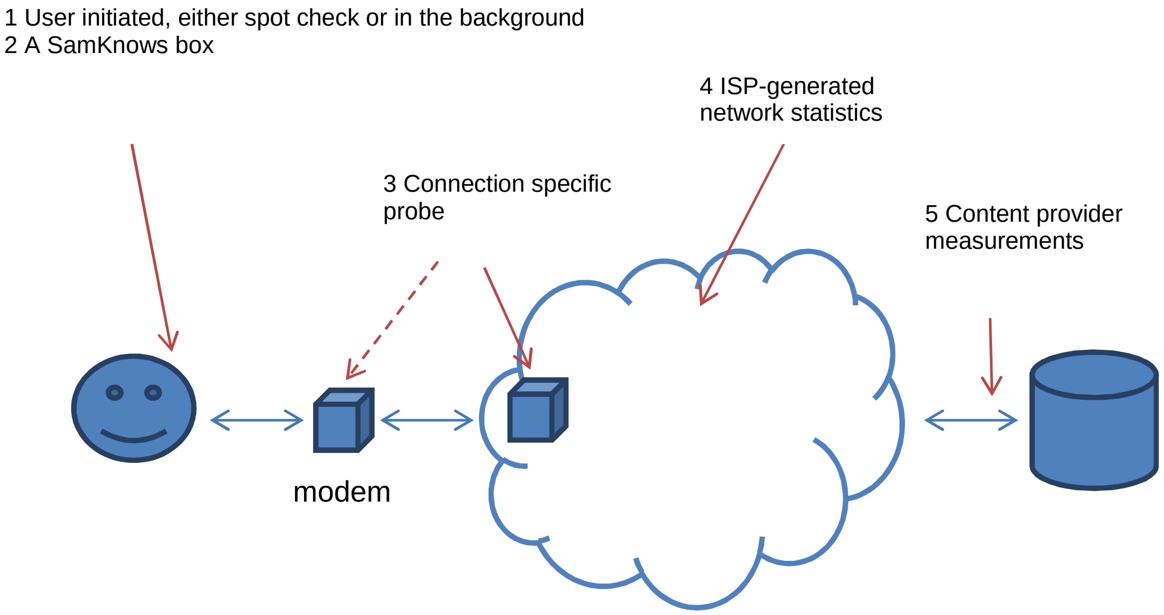

1 User initiated, either spot check or in the background
2 A SamKnows box
4 ISP-generated network statistics
3 Connection specific probe
5 Content provider measurements
modem

Figure 17: Options for data capture

1 User initiated

The principle of running broadband measurement applications on a user's device is established. Sites such as pingtest.net and speedtest.net interact with remote servers to provide basic QoS data. An application called glasnost[19] performs similarly but sends and receives different data types in order to calculate whether there is any traffic management in operation. Such tests are useful for troubleshooting but there is the possibility that ISPs could detect the packets and prioritise them in order to manipulate the result (though we emphasise that we have no indication that such practices are occurring).

It is also possible to have software measuring performance as a background process or as part of an application. These may not easily be possible in some situations, especially on mobile devices where they may impose a significant overhead.

These applications typically send data to a site to be aggregated. Whether this gives a realistic overall picture is questionable because they will tend to be used mostly when problems have been experienced.

2 SamKnows box

Here, software resides either in a special purpose device or in a router or modem, and runs tests initiated and aggregated externally. The company SamKnows has developed this technique and currently monitors six ISPs for Ofcom. The technique is explained in a paper on the SamKnows website[20]. We will refer to the approach as SamKnows because it is the best example of this approach in the UK, though it is no doubt possible for similar techniques to be used by other companies[21].

[19] http://broadband.mpi-sws.org/transparency/bttest.php
[20] http://www.samknows.com/broadband/pm/PM_Summer_08.pdf
[21] References to SamKnows should be seen as generic to the class of technique rather than specific to that company

The SamKnows approach may not be feasible in all circumstances. The software works by generating and sending extra data. This causes an overhead, which may not matter in most fixed broadband contexts, but could eat up allowances in mobile broadband. If a traffic management intervention only applies to a small number of users then the effect may not be noticeable in the sorts of sample size typically used. It may also be inconvenient for the user if, say, volumes of data are injected purely to trigger a volume-related intervention.

In the case of home networks it is necessary to run the tests from a device with priority over other traffic on the network.

3 Connection-specific probe

This option involves measuring aspects of quality of service at a connection from within the network, or from the modem. The ability to measure some parameters is incorporated into some network equipment but may not be used in practice. Such measurements need to be made near the edge of a network where the user context is known. There would be considerable additional cost and complexity in adding a workable measurement capability in some networks.

This option is particularly relevant to the reporting of usage statistics. The architecture of traffic management has a big effect on what is feasible. If traffic management is centralised then data will only be available in an aggregate way. Not all ISPs can offer all details related to an individual connection at present.

4 Network statistics

This technique puts the onus on ISPs to make measurements on their network that characterise QoS and the effect of traffic management. How the measurements are made can depend on the architecture of the network. Network nodes are often already able to report the behaviour of queues, and therefore latency and loss statistics. It may also be possible to aggregate connection specific measurements. Alternatively, in the absence of specific data, it may be possible to model QoS for the expected network conditions.

5 Content provider

In principle, measurements can be performed from a content provider's server. It is likely that from this point it would be possible to expose any differences between ISPs, or how aggregate QoS varies over time. However content providers do not have access to information on users' tariff packages and will not be in a position to determine how much an ISP is actively managing traffic on a particular tariff which, we think, is what consumers would want to know.

9.2.1 Summary

	Method	Location	Temporal	Initiation
1 User initiated	Existing traffic or Injected traffic	User device	Spot test or continuous in background	User
2 SamKnows box	Existing traffic or Injected traffic	User device	Quiet periods	External
3 Connection-specific probe	Existing traffic	Modem or DSLAM/CMTS/Node B	Continuous	ISP
4 ISP-generated network statistics	Existing traffic	Network nodes	Continuous	ISP
5 Content provider	Existing traffic	Server	Continuous	Content provider

Table 10: Summary of characteristics of the main measurement options

9.3 Evaluation

In principle, with enough investment, all the desired measurements can be made, and all the desired data can be generated. In practice we think that such investments should be considered from a cost-effectiveness standpoint. We have considered the applicability of each of the measurement approach to each of the scenarios, and taking into account the scenario and the measurement options we suggest the following.

9.3.1 Prospective

	Scenario				
	Scenario 1 Fair use	**Scenario 2** Congestion management	**Scenario 3** Video streaming	**Scenario 4** Business tool	**Scenario 5** Managed services
Prospective Information					
1 User initiated					
2 SamKnows box	✓ (not to detect TM)	✓ (not to detect TM)	✓	✓ (not with complex tariffs)	✓
3 Connection-specific probe					
4 ISP-generated network statistics	✓	✓	✓	✓	✓
5 Content provider					

Table 11: Options for prospective measurements

The two main approaches relevant to a consumer looking to evaluate ISPs other than his/her own are the SamKnows approach and ISP-generated network statistics. Because scenarios 1 and 2 would tend not to affect enough users to make a SamKnows approach cost effective, we consider that data will be better provided from within the network. For the other scenarios both of these techniques can be used. However the SamKnows approach may not be feasible with complex tariffs (scenario 4) because of the need for a large enough sample of each tariff. In the case of mobile networks, the SamKnows approach may impose an unacceptable overhead, suggesting that network statistics are the only solution. Not all QoS parameters are easily measured from within a network, however.

9.3.2 Performance checking

	Scenario				
	Scenario 1 Fair use	**Scenario 2** Congestion management	**Scenario 3** Video streaming	**Scenario 4** Business tool	**Scenario 5** Managed services
Performance Information					
1 User initiated	(✓)	(✓)	(✓)	(✓)	(✓)
2 SamKnows box					
3 Connection-specific probe	(✓)	(✓)	✓	✓	✓
4 ISP-generated network statistics	✓	✓	(✓)	(✓)	(✓)
5 Content provider					

Key: (✓) indicates that the option may be appropriate but with some qualification

Table 12: Options for performance measurements

There are more options in this case. ISPs may not recognise user-initiated measurements as valid, so the best option is ISP measurements. We consider that the lower variability in scenarios 1 and 2 will tend not to require connection specific probes. SamKnows boxes are not appropriate to this application because only a small proportion of users have these boxes.

9.3.3 Traffic management status checking

	Scenario				
	Scenario 1 Fair use	**Scenario 2** Congestion management	**Scenario 3** Video streaming	**Scenario 4** Business tool	**Scenario 5** Managed services
Status Information					
1 User initiated					
2 SamKnows box					
3 Connection-specific probe	✓	✓	✓	✓	✓
4 ISP-generated network statistics	✓	✓	✓	✓	✓
5 Content provider					

Table 13: Options for status measurements

The information is essentially only held by the ISP so it needs to be provided from within the network, either connection-specific or for the network as a whole.

9.4 Conclusion

There is no one solution to the provision of measurement data.

In the case of prospective data capable of giving greater certainty over the details of how a traffic management policy is working in practice there are two main options – the SamKnows approach and ISP generated network statistics. Both of these approaches are capable of providing consistency and repeatability. The SamKnows approach is based on sampling, so where the tariff or circumstance has low prevalence, the SamKnows sample will possibly be too small to provide statistical reliability. Alternatively, where there are numerous different tariffs and situations, the number of distinct SamKnows samples will tend to proliferate. Using ISP generated data can potentially get around these problems but such data has its own limitations and there might still be a need to audit data produced by ISPs.

For real time performance information, there is always the possibility of users initiating tests using third party software and servers. Alternatively, ISP data can be used. Once taken in conjunction with the need for status information, it becomes clear that ISP data (either connection-specific probes or network statistics) would be a strong option. We envisage that the ISPs would provide a performance and status interface in real time. This should be connection specific if possible, but some aspects could be representative if connection specific data are not available because of the architecture.

Mobile broadband is more difficult in this context than fixed. The susceptibility of mobile broadband to all sorts of factors that affect the radio access network means that past data may not have much predictive power. Data can be collected from within the network but will have to be aggregated and will not be very relevant to what is essentially a location-sensitive service.

10 Conclusions

The current approach adopted by most UK ISPs to traffic management can be characterised as 'minimalist'. This form of traffic management is mainly designed to promote 'fair use' so that heavy users are not able to consume such a disproportionate level of network resources that they degrade the service available to moderate or light users. The information that ISPs currently provide to consumers on their traffic management describes their policies in broad terms but often not in sufficient detail that a user can predict the precise level of traffic management they will experience. This is partly because the level of traffic management at any one time can depend on the level of congestion, which in turn is subject to statistical variations.

This report has described a series of future scenarios that would involve greater use of traffic management. The scenarios have been developed by considering commercial and technical trends. These scenarios show traffic management either affecting more users or being more complex, or a combination of the two.

Assuming that consumers should have 'transparent' information about the traffic management envisaged in these scenarios, the question then arises as to what this information should be, and how it should be obtained. 'Transparency' can be considered to include meaningfulness, accuracy and comparability.

Meeting the requirement for transparency in these scenarios would involve more detailed information being given to consumers than at present. This information would be necessary to describe the packages on offer and to allow consumers to check that the delivered services accord with the package descriptions.

The representation of the effects of traffic management on performance is reasonably straightforward. This study has shown that common templates are feasible, given that traffic management interventions are broadly only of a few basic sorts. Applying the transparency criteria of meaningfulness, accuracy and comparability leads to three 'representations':

- a detailed policy, supplemented by time series data, which describes QoS
- a summary which presents the QoE of popular applications
- data for wizards (only necessary in the more complex scenarios).

There remains a need to translate between QoS and QoE in a consistent way. For example, the same threshold for the data rate necessary to support an HD video stream should be applied by all providers. The production of these criteria and thresholds could be undertaken by Ofcom or may possibly be devolved to an industry body such as the Broadband Stakeholder Group.

Even now, some tariffs make real-time status information desirable. Applying concepts of transparency to the traffic management scenarios could make such information essential in the future. Where usage limits are easily reached, then consumers will arguably need real time status information in order to regulate their usage patterns. If status information were provided it would fit comfortably alongside the provision of comprehensive real time information on QoS and any applied traffic management interventions.

By its nature, some traffic management is non-deterministic. For example, the effect of packet prioritisation is relative to other traffic and therefore depends on network conditions. In such cases, a policy alone cannot fully describe the QoS on offer: the level of traffic management and the effect on an individual connection may be impossible to predict with certainty. This facet of traffic management sets a limit to the achievable transparency.

Greater certainty for consumers when trying to compare packages could be provided through the use of time series data to show how much traffic management had been applied in the past. For example, it might be possible to state that on a typical weekday evening, P2P traffic was reduced to an average data rate of 5 Mbit/s on a connection capable of supporting 20 Mbit/s.

Providing such data is not straightforward. Though there is no single ideal way in which performance can be measured, on balance ISP-generated data will probably be the best long term solution. However, the architectures of ISP networks differ, and some ISPs are in a better position to gather such data from within their networks than others. For ISPs with centralised traffic management, imposing a requirement to measure and publish performance would be particularly costly. Accordingly we conclude that transparency will be enhanced by ISPs providing time series data from within their networks, and that the provision of such data should be encouraged. However, the case for mandating such data will need to be assessed carefully. The imposed costs would need to be evaluated relative to consumer benefits, and both sides of this trade-off will vary according to the traffic management scenario in operation. At present, the ‘fair use’ approach to traffic management is probably sufficiently minimal in its impact on most consumers that time series data would only be of benefit to a small proportion of consumers.

The Broadband Stakeholder Group has published its voluntary industry code of practice on traffic management transparency for broadband services. The code of practice includes a key facts indicator (KFI) which is similar in intent to the QoS Policy Form described in this report.

Appendix A Terms of Reference (from Ofcom ITQ)

Background

The internet is increasingly central to the lives of citizens, consumers and industry. It is a platform for the free and open exchange of information, views and opinions; it is a major and transformative medium for business and e-commerce, and increasingly a mechanism to deliver public services efficiently. As such it provides access to a growing range of content, applications and services which are available over fixed and wireless networks.

Many of these services, particularly those which contain video content, require high capacity networks to deliver them. Some networks are already experiencing congestion problems as consumers use ‘bandwidth hungry’ services. Even in the longer term, as next generation networks are deployed, there may continue to be congestion problems, particularly in wireless networks.

In response, network operators and internet service providers (ISPs) are making greater use of traffic management techniques. These can allow them to handle traffic more efficiently, to prioritise traffic by type, to charge for guaranteed bandwidth or to block or degrade the quality of certain content. It is important for ISPs to be able to clearly communicate the impact of their traffic management techniques to consumers, so that consumers can make informed choices about their broadband services.

Objective

This technical study will review the (current and likely future) traffic management techniques used by ISPs and the options for their characterisation, such as the Quality of Experience (QoE) provided to different types of online services. We have drawn up a list of questions to be answered by this study, divided into two categories depending on whether they relate to traffic management techniques or QoE:

Traffic Management	**Assessment of QoE**
• Which traffic management technologies and approaches are used that are under the direct control of ISPs and how are these likely to develop in coming years? • Which traffic management technologies and approaches are used that are not under the direct control of ISPs and how are these likely to develop in the coming years? • What role with Content Delivery Networks play? • Will these technologies be (or continue to be) network specific, or deployable across multiple network types? Will there be any changes to where traffic management technologies are deployed in the network? • What impact (positive and/or negative) does traffic management employed by a particular ISP have on: 1. Access to the internet by consumers; 2. Access to the internet by content providers; and 3. Other ISPs that may be part of an end-to-end exchange of data? • Are there any barriers to achieving end-to-end traffic management? Is co-ordination required between different traffic management technologies operated by interconnected ISPs? • Is there the potential for an *arms race*? For example, will it be possible for content owners to categorise their content in such a way to circumvent traffic	• What are the different options for measuring and characterising the impact of traffic management on consumer internet connections? • Which of these approaches (if any) would enable a meaningful repeatable (quantitative) comparison of performance across different ISP providers? • What are the relative merits of using a QoS and QoE approach for characterising connection performance and establishing a potential minimum level of required internet connection performance? • What effect would connection sharing, e.g. using a WiFi access point, have on these measures? • To what extent could these measurements of connection performance be communicated in a meaningful way to consumers? • How will measures of QoS and QoE stay relevant over time, i.e. can they be technology and application neutral? Will it be possible to update them as networks evolve?

management approaches? • Can traffic management cause further congestion in core networks? For example, if packets are dropped by ISP routers (rather than delayed), can this result in excessive retransmission by servers? Are traffic management techniques “polite” to the rest of the Internet?	

This study is intended to provide some of the technical input into Ofcom’s policy activity on Net Neutrality and Traffic Management. We therefore require this study to be completed within 3 months.

Deliverables

We suggest the following deliverable schedule:

1. A draft final report, at month 2
2. A final report, at month 3
3. A presentation to project team members, at month 2 or 3 as appropriate;

Appendix B Ofcom's technical questions

The technologies and approaches are covered in detail in Chapters 3 through to 5. Our key conclusions are bullet pointed below:

B.1 Which traffic management technologies and approaches are used that are under the direct control of ISPs and how are these likely to develop in coming years?

- The majority are under the direct control of ISPs
- The technologies and approaches are reasonably well established. ISPs use what they call "DPI boxes" to identify and categorise packets. Despite the name, Deep Packet Inspection is not necessarily used
- Subsequently packets may be treated differently according to their priority. At nodes packets may queue for longer or may be dropped according to their priority.
- Traffic may also be restricted in data rate
- No account is taken of packet priorities as signalled in the headers of incoming packets
- Packet priorities are generally stripped out on leaving the ISP's network
- In radio access networks, there are ways to prioritise within the radio layer. However it is technically difficult to integrate radio and IP management
- The technologies are not predicted to change. However there are general moves towards more distributed forms of traffic management (moving from the core towards the edge of networks) which would allow more finessed approaches.

B.2 Which traffic management technologies and approaches are used that are not under the direct control of ISPs and how are these likely to develop in the coming years?

- The ISPs we spoke to did not consider the technologies and approaches that are not under their control. Fundamentally ISPs restrict their interpretation of traffic management as helping them manage their network.
- Managed services and CDNs are not conventionally thought of as traffic management but they do have an effect on QoE. Both are set to become more prevalent. They may not even be visible from the ISP's perspective.

B.3 What role with Content Delivery Networks play?

- CDNs are in existence but are not perceived as being part of traffic management. ISPs did not highlight the role of CDNs and regarded them exclusively as a matter for content providers. The recent launch of BT Content Connect might be seen to counter this, but BT Wholesale is not an ISP.
- CDNs are marketed as offering end users a better experience. They can always be justified on network efficiency grounds.
- The launch of BT Content Connect as, in effect, a CDN within an operator's network is a new development. This should benefit consumers in helping to provide a better QoE for content-hungry applications, such as video streaming. Ofcom may wish to look into any impact this may have in the market for independent CDNs.

B.4 Will these technologies be (or continue to be) network specific, or deployable across multiple network types?

- In principle packet priorities could be retained across network types. In practice there is no interest in end to end traffic management within the internet. Managed services can cut across network types, but may or may not be regarded as services on the internet.
- We found that ISPs operating the three main types of network (DSL, cable and mobile) used broadly the same traffic management principles.

- There were differences in emphasis between network types – for example we found that DSL operators used all five of the intervention types described in section 3.3. Mobile operators used all except bandwidth allocation by traffic type (our reference B3), and cable operators used all except packet prioritisation by user identity (our reference A1)
- There are good reasons for these differences – see section 3.3.3 for details
- In view of the above findings we do not see any need for regulation of traffic management to apply differently to operators delivering service over different types of network

B.5 Will there be any changes to where traffic management technologies are deployed in the network?

- Most ISPs observe a trend to put traffic management closer to the edge and closer to users.
- This allows more finessed and individualised packages.
- But the extent of this trend does depend on cost. There needs to be a business benefit to justify the additional investment that would be needed.

B.6 What impact (positive and/or negative) does traffic management employed by a particular ISP have on: 1. Access to the internet by consumers; 2. Access to the internet by content providers; and 3. Other ISPs that may be part of an end-to-end exchange of data?

- In summary, most ISPs regard traffic management as a way of delivering an acceptable QoS for all, and penalising users who use more than their fair share.
- Very heavy users may be throttled back and/or removed
- There is no evidence in our interviews of ISPs using traffic management to affect access from content providers or other ISPs.
- Specifically:

1. Consumers that are particularly heavy users of certain types of traffic (e.g. P2P) can expect restrictions to be applied to their service to prevent undue impact on the service of other consumers. At peak times (typically evenings) users may see a reduction in overall internet speed. Some users may also experience a usage ‘cap’ which, if exceeded, will result in their service being restricted in speed for a period of time.

2. We have been told of an instance in the past where an ISP limited the bandwidth available for a particular type of traffic from a specific content provider. This was intended to provide consumers overall with a better service by limiting the coding rate of all data streams from this provider. This restriction has now been removed and we were not told of any similar restrictions currently in place in the UK. However, current regulation would not prevent such a restriction being applied again by an ISP.

3. ISPs told us that they did not take account of any packet prioritisation information in the headers of traffic entering their network. They apply their own traffic management policies to the data as it travels across their networks. They didn’t expect that ISPs receiving data from their networks took account of their prioritisation either. We may therefore conclude that the overall QoE enjoyed by a consumer will reflect the summation of the traffic management policies being applied by all the operators in the chain between the source and destination of a link. For each traffic type the result will reflect the most restrictive management applied to that traffic type by any operator in the chain.

B.7 Are there any barriers to achieving end-to-end traffic management? Is co-ordination required between different traffic management technologies operated by interconnected ISPs?

- Currently ISPs are not looking to achieve end to end traffic management.

- An end to end approach would effectively allow the ISP closest to the customer to detect and set the priority of each packet, and for this priority to be recognised by subsequent networks. In practice this is not done.
- There is also a bandwidth allocation mechanism built in to the standards that would allow end-to-end PVCs to be created. Such techniques are not used in the open internet.
- We are not sure if end-to-end traffic management is an achievable goal, in the short term at least. One ISP might apply a particular traffic management policy which is appropriate to the architecture and capacity of its network. The next ISP in the end-to-end chain might have a different architecture or capacity restriction (e.g. having a mobile access network) and could therefore be unable to honour the packet prioritisation or bandwidth allocation applied by the first ISP.

B.8 Is there the potential for an arms race? For example, will it be possible for content owners to categorise their content in such a way to circumvent traffic management approaches?

- In practice a variety of techniques are used by ISPs to identify content types. The vendors of packet inspection equipment send out updates to ensure their equipment retains the ability to detect content, even if attempts have been made to ‘hide’ it
- The ISPs did not tell us of any difficulties in identifying content
- There is no reason to believe that content identification is a losing battle – ISPs and vendors are confident that they can keep up with the development and use of new traffic types.

B.9 Can traffic management cause further congestion in core networks? For example, if packets are dropped by ISP routers (rather than delayed), can this result in excessive retransmission by servers? Are traffic management techniques “polite” to the rest of the Internet?

- Not all traffic management techniques, and not all protocols, will result in retransmission.
- However, reducing the priority of TCP/IP data can cause data loss, which in turn requires retransmission.
- The sender will reduce the data rate accordingly, but retransmission is still likely
- If retransmission emerges as a problem, then bandwidth restriction is an alternative which does not cause the same problems.

Appendix C Ofcom's QoE questions

The technologies and approaches are covered in detail in Chapters 6 through to 9. Our key conclusions are bullet pointed below:

C.1 What are the different options for measuring and characterising the impact of traffic management on consumer internet connections?

The options all involve a combination of a written policy, a set of measurements, and a way of representing the information. The main ones are:

- A QoS Policy Form giving full details
- A QoE Summary giving highlights and indicating applicability
- A wizard to help in complex choices
- A real time connection status dashboard.

C.2 Which of these approaches (if any) would enable a meaningful repeatable (quantitative) comparison of performance across different ISP providers?

All have their place, and it depends how far traffic management moves from its current 'fair use' paradigm. We identified five types of measurement approach:

1 User initiated

2 SamKnows box

3 Connection-specific probe

4 ISP-generated network statistics

5 Content provider.

Of these, 2, 3 and 4 are most relevant. The SamKnows approach (2) can be used for validation but may not be practical to give complete characterisation of all tariffs. In-network measurement (3&4) is constrained by architecture and by the sorts of information that can be provided from within a network.

C.3 What are the relative merits of using a QoS and QoE approach for characterising connection performance and establishing a potential minimum level of required internet connection performance?

Transparency involves three factors which cannot always be simultaneously satisfied. These are:

- accuracy
- meaningfulness
- comparability.

The QoS and QoE approaches are both necessary. QoS is more accurate and QoE is more meaningful. A method to convert QoS to QoE is needed, and this involves establishing minimum performance levels for applications. The levels must be consistent between ISPs.

C.4 What effect would connection sharing, e.g. using a WiFi access point, have on these measures?

ISPs deal with connections, not individual devices. Thus in general connection sharing is not well accommodated within the frameworks suggested. A particular issue is usage limits where one user of the connection may cross thresholds to the detriment of other users of that connection.

C.5 How will measures of QoS and QoE stay relevant over time, i.e. can they be technology and application neutral? Will it be possible to update them as networks evolve?

Both applications and traffic management will evolve. QoS is more likely to stay relevant than QoE. There is no problem in principle in updating as required. The QoS Policy Form should remain relevant over time as it is technology-neutral and application-neutral. The QoE Summary is necessarily application-specific (in order to be more meaningful) so it will need to be updated to reflect application developments.

Appendix D An ISO Model View of Traffic Management

D.1 Introduction

The ISO model is the reference model for IP networks so we have included a simple explanation here. In order to describe traffic management we will look at:

- The way in which IP networks work, based on the ISO model
- The way in which congestion and QoS are managed
- The way in which applications and protocols work
-

D.2 IP Traffic and the ISO Model

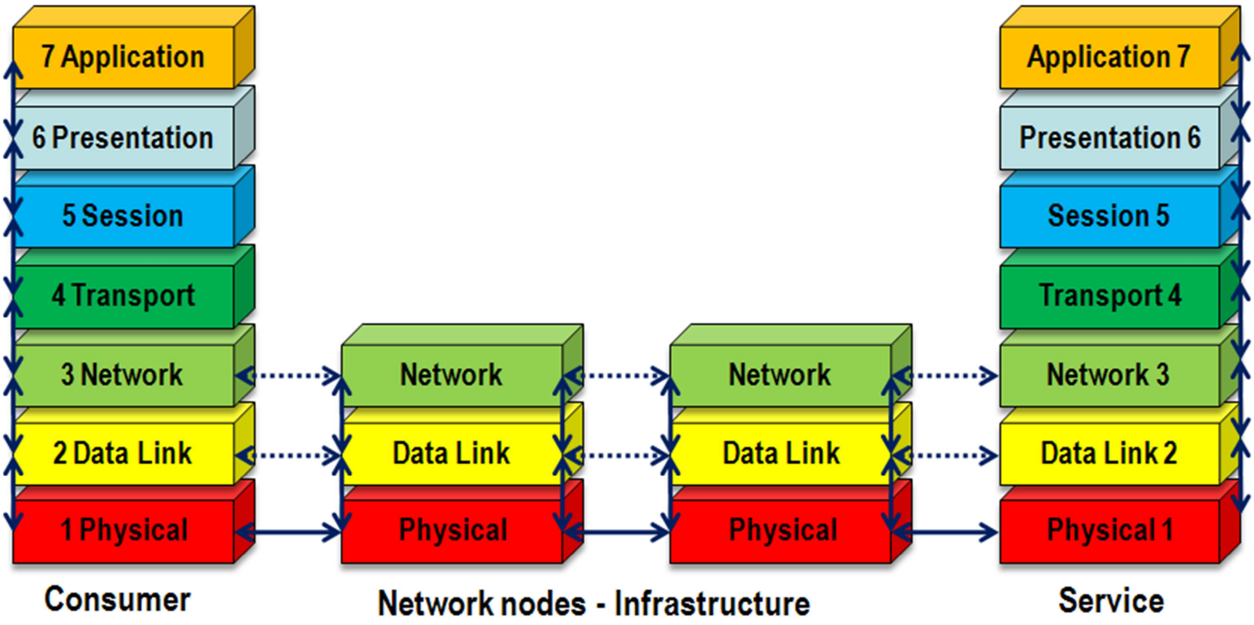


Figure 18: The ISO 7 Layer Communications Network Model

Figure 18 above summarises the arrangement of the well-known ISO 7 Layer model for communications systems.

With respect to the internet model, 5 main layers are used:

- Layer 1: Physical layer – defines the electrical / interface technology.
- Layer 2: Data link (Media Access Control – MAC) task is to take the Layer 1 transmission and convert this into a stream of error free scheduled data over the specific technology being deployed.
- Layer 3: Network layer – Forwards and routes packets based on a priority implemented in the 'Internet Protocol – IP'.
- Layer 4: Transport – Uses protocols such as TCP/IP and UDP to determine the format which the data is transmitted over the Network Layer.

- Layer 5-7: Application – Uses high level protocol such as HTTP, DNS etc. The application layer is not an application as such, it provides the ability to have a ‘network transparent’ common system for resource allocation and partitioning.

In the IP world, traffic is typically managed and controlled at Layer 3 and above. In contrast, in the telecoms world, resource allocation controls apply to Layers 1 and 2. This means that these networks operate differently under overload conditions.

In an IP network, congestion occurs at Layers 1, 2 and 3. Services are managed at Layer 3 and above, whilst congestion in Layers 1 and 2 is managed by schedulers running in Layer 2. These schedule data and prioritise access to the physical resources available in Layer 1.

D.3 Data Network Protocols and QoS

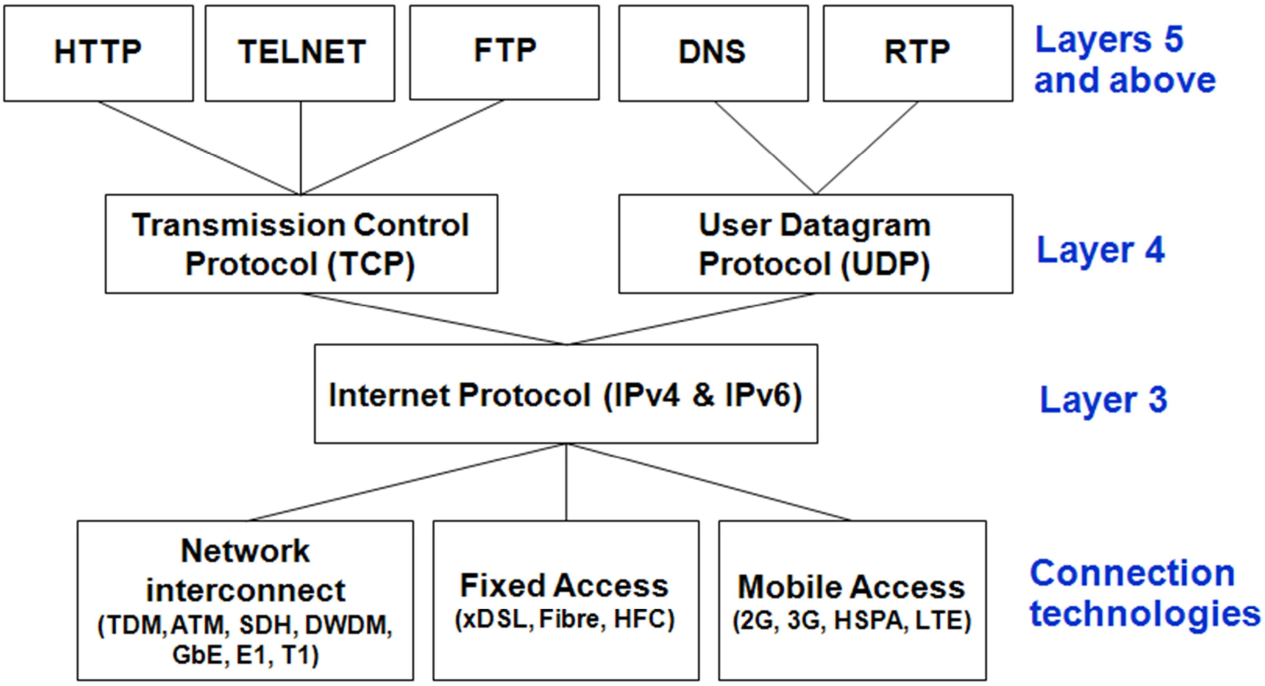

HTTP
TELNET
FTP
DNS
RTP
Layers 5 and above
Transmission Control Protocol (TCP)
User Datagram Protocol (UDP)
Layer 4
Internet Protocol (IPv4 & IPv6)
Layer 3
Network interconnect (TDM,ATM, SDH, DWDM, GbE, E1, T1)
Fixed Access (xDSL, Fibre, HFC)
Mobile Access (2G, 3G, HSPA, LTE)
Connection technologies

Figure 19: Data Network Protocols

Figure 19 above shows the main protocols used in data networks, and their relationships to each other.

In summary, HTTP is the protocol used for web browsing, Telnet is a bi-directional interactive text-oriented protocol and FTP is used for file transfer. These protocols all feature flow control, with the ability to re-transmit lost or damaged packets. TCP supports these protocols by including flow control to provide a reliable, ordered delivery of IP traffic over the network.

On the right hand side of the figure, DNS provides the internet ‘phone book’ and RTP is the standard protocol for sending audio and video content over IP networks. UDP delivers these protocols via a service that emphasises low latency over reliability of transmission.

TCP and UDP traffic have different transmission requirements and constraints. Ideally, they would be managed in different ways.

UDP provides a transport layer with the source and destination ports (the IP addresses are actually in the IP layer (Layer 3). UDP does not support flow control and retransmission and is typically used for streaming applications.

TCP is used to transmit data reliably. It is less suited for real time streaming due to the overhead required to implement the retransmission of lost packets.

IP is the basic protocol of the Internet which operates at Layer 3 and in isolation is unreliable for guaranteed delivery of data. The transport layer 4 deploys the techniques used to create the data packets which are delivered over IP layer 3, therefore layer 4 processing determines the QoS for layer 3 to process.

D.4 Application of QoS in the ISO Model

In many IP networks there is no connection between QoS management in the upper layers and physical provisioning in Layers 1 and 2. Operators rely on supplying adequate physical network capacity, and constraining traffic at Layer 3 or above, in order to avoid overload. Connection speed can be controlled via Layers 1 and 2, but systems at these levels have no knowledge of traffic type. This means that all forms of traffic running over the connection are impacted by any changes to capacity.

In the past, networks were usually designed for a single application, for example, voice or data. The evolution of networks to carry multiple different traffic types has led to the management of traffic, and hence the control of QoS, being distributed more widely across the ISO stack, depending on the traffic management objective, for example:

- IP QoS control – managed at Layer 3 and above
- Physical connection speed – managed at Layers 1 and 2, and driven by prioritisation from layer 3 and above.

Mobile operators, whose networks are often heavily loaded, face special challenges when more capacity is needed in the access network. They may not have access to additional radio spectrum, and moving to smaller cells to improve frequency re-use may require new base sites to be build and backhaul to be installed. They therefore make extensive use of resource scheduling in Layers 1 and 2 in order to make the best use of available capacity. This provides a more graceful degradation of performance under overload conditions than the alternative of allowing IP rules to battle inefficiently for capacity in a constrained radio access network.

However, the mobile community is also following the same evolution as fixed and currently implement a hybrid model across 2G – 3G – HSPA, where services such as voice take priority over data. Data services are generally treated equally with little QoS differentiation. As mobile networks evolve to handle more complex data services it is highly likely that the evolution experienced in the fixed world will be deployed in the mobile world by implementing more complex IP level differentiation.

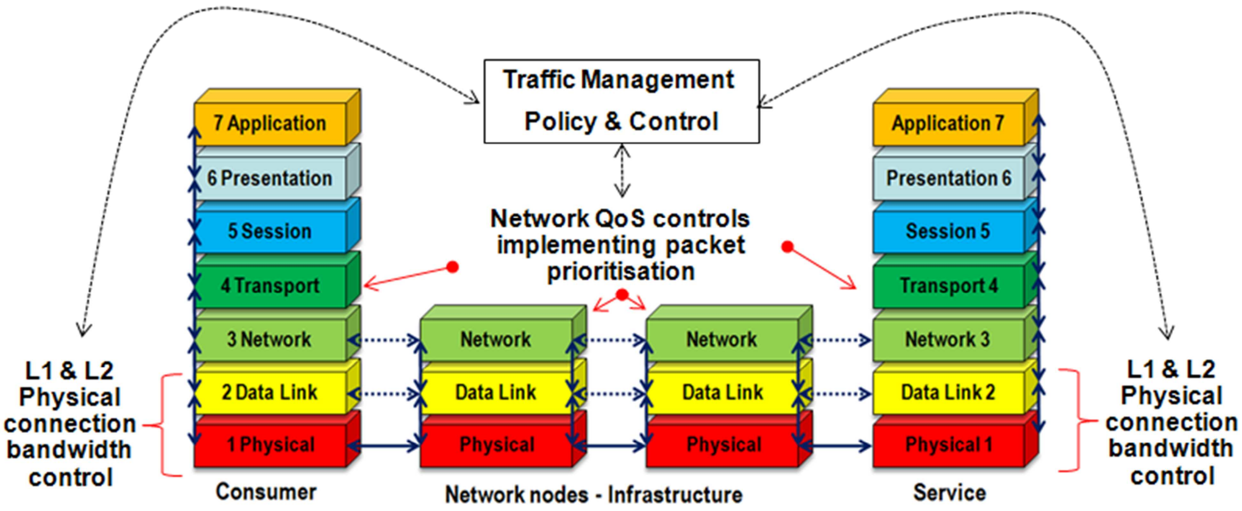


Figure 20: The ISO 7 model and traffic management controls

Figure 20 summarises the main elements of managed system where policy and control implements the ISP profile through managing bandwidth at layers 1 and 2, and QoS mechanisms (packet prioritisation) at layers 3 and 4.

Packet prioritisation can be conducted by identifying packets through techniques such as DPI and marking them accordingly and/or traffic shaping/policing for known traffic types within control of the network.

Operators believe they have the information and technology to understand where congestion is beginning to be a problem and to manage this. In the short term they do this by limiting the capacity available for certain traffic types, and in the longer term this is achieved by installing additional capacity.

Appendix E A Network Type View of Traffic Management

This appendix shows schematically how the principal components of traffic management can be implemented in different network types. The diagrams have been compiled from public sources describing networks internationally and are not intended to depict the networks in use in the UK.

E.1 DSL

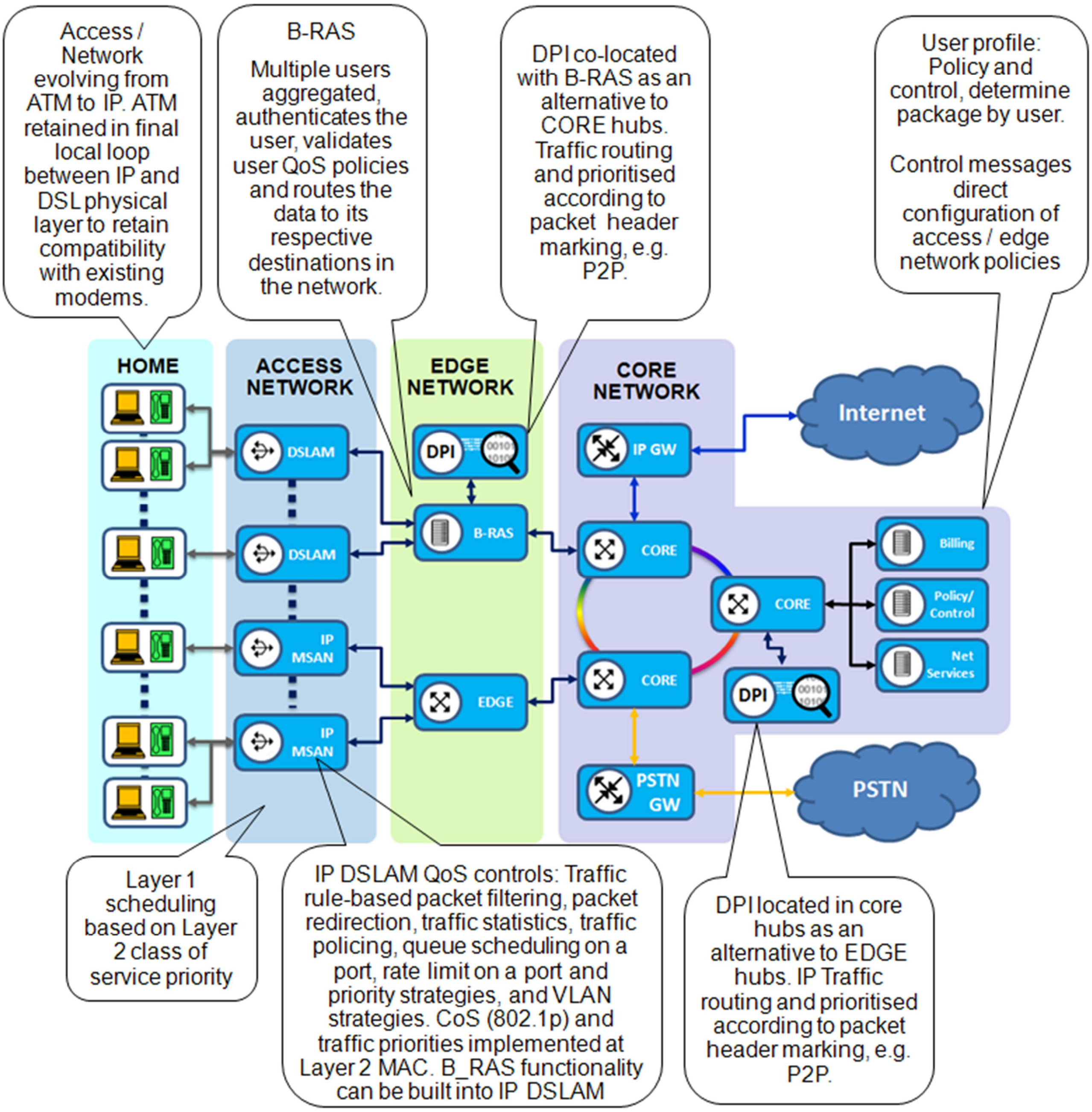
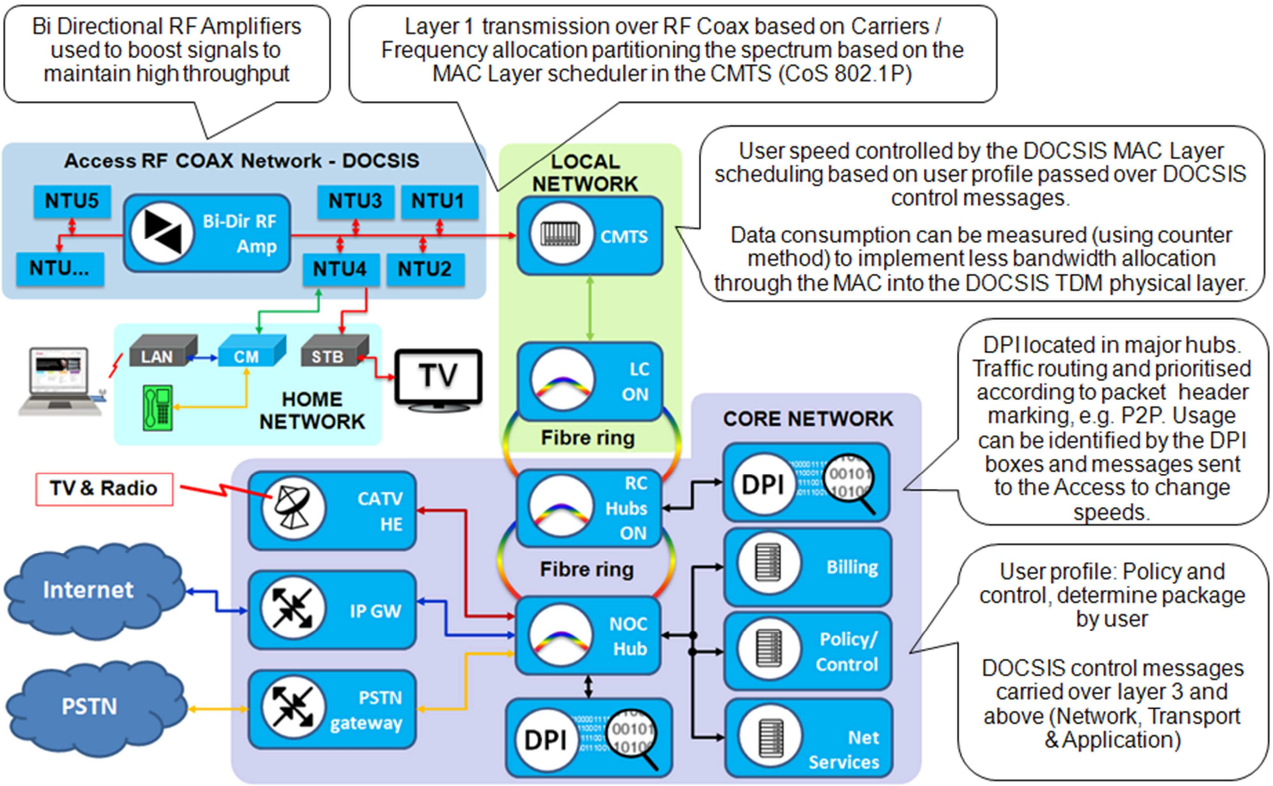

Access / Network evolving from ATM to IP. ATM retained in final local loop between IP and DSL physical layer to retain compatibility with existing modems.
B-RAS
Multiple users aggregated, authenticates the user, validates user QoS policies and routes the data to its respective destinations in the network.
DPI co-located with B-RAS as an alternative to CORE hubs. Traffic routing and prioritised according to packet header marking, e.g. P2P.
User profile: Policy and control, determine package by user.
Control messages direct configuration of access / edge network policies
HOME
ACCESS NETWORK
EDGE NETWORK
CORE NETWORK
Internet
DSLAM
DSLAM
IP MSAN
IP MSAN
DPI
B-RAS
EDGE
IP GW
CORE
CORE
CORE
DPI
PSTN GW
PSTN
Billing
Policy/ Control
Net Services
Layer 1 scheduling based on Layer 2 class of service priority
IP DSLAM QoS controls: Traffic rule-based packet filtering, packet redirection, traffic statistics, traffic policing, queue scheduling on a port, rate limit on a port and priority strategies, and VLAN strategies. CoS (802.1p) and traffic priorities implemented at Layer 2 MAC. B_RAS functionality can be built into IP DSLAM
DPI located in core hubs as an alternative to EDGE hubs. IP Traffic routing and prioritised according to packet header marking, e.g. P2P.

Figure 21: Typical Structure of ADSL ISP Connectivity

E.2 Cable


Bi Directional RF Amplifiers used to boost signals to maintain high throughput

Layer 1 transmission over RF Coax based on Carriers / Frequency allocation partitioning the spectrum based on the MAC Layer scheduler in the CMTS (CoS 802.1P)

User speed controlled by the DOCSIS MAC Layer scheduling based on user profile passed over DOCSIS control messages.

Data consumption can be measured (using counter method) to implement less bandwidth allocation through the MAC into the DOCSIS TDM physical layer.

DPI located in major hubs. Traffic routing and prioritised according to packet header marking, e.g. P2P. Usage can be identified by the DPI boxes and messages sent to the Access to change speeds.

User profile: Policy and control, determine package by user

DOCSIS control messages carried over layer 3 and above (Network, Transport & Application)

Figure 22: Typical Structure of Cable Access Network

Figure 22 shows an architecture implementation based on cable networks using Hybrid Fibre Coax (HFC) running the Data Over Cable Service Interface Specification (DOCSIS). This modulates data onto carriers which fit within the 8MHz PAL TV channel allocations of European CATV systems. DOCSIS supports a maximum downstream (to the consumer) data throughput of 55.62Mbit/s per channel, using up to 256-level QAM modulation. The upstream throughput is a maximum of 10.24 or 30.72Mbit/s, depending on DOCSIS version.

The Cable Modem Termination System (CMTS) is connected on the network side to redundant fibre rings, which in turn connect to the operator's fibre backbone network. Consumers have Cable Modems (CMs) which provide access to one or more DOCSIS channels. Contention is specified by the number of consumers connected to each channel.

Cable operators have good control over contention and congestion, and are less limited by available bandwidth in the access network than ADSL operators. They can change allocations of consumers to channels, or allocate additional channels to meet changing demand. The DOCSIS 3.0 specification provides greater upstream throughput per channel and allows channels to be configured together to provide downstream throughputs of 222.48Mb/s (4 channels) or 444.96Mbit/s (8 channels).

Cable operators can apply traffic management on a per-user basis in the access network and de-prioritisation of selected traffic types in the core network, typically at one of a small number of major backbone hubs.

As with other network technologies and topologies, DPI is used to identify traffic and packets are marked according to a traffic management strategy controlled centrally (Policy & Control). Bandwidth allocation will be implemented by using the information provided by the packet marking and the user profiles instructions also derived from the traffic management strategy implemented through the Policy and Control centre.

E.3 Mobile

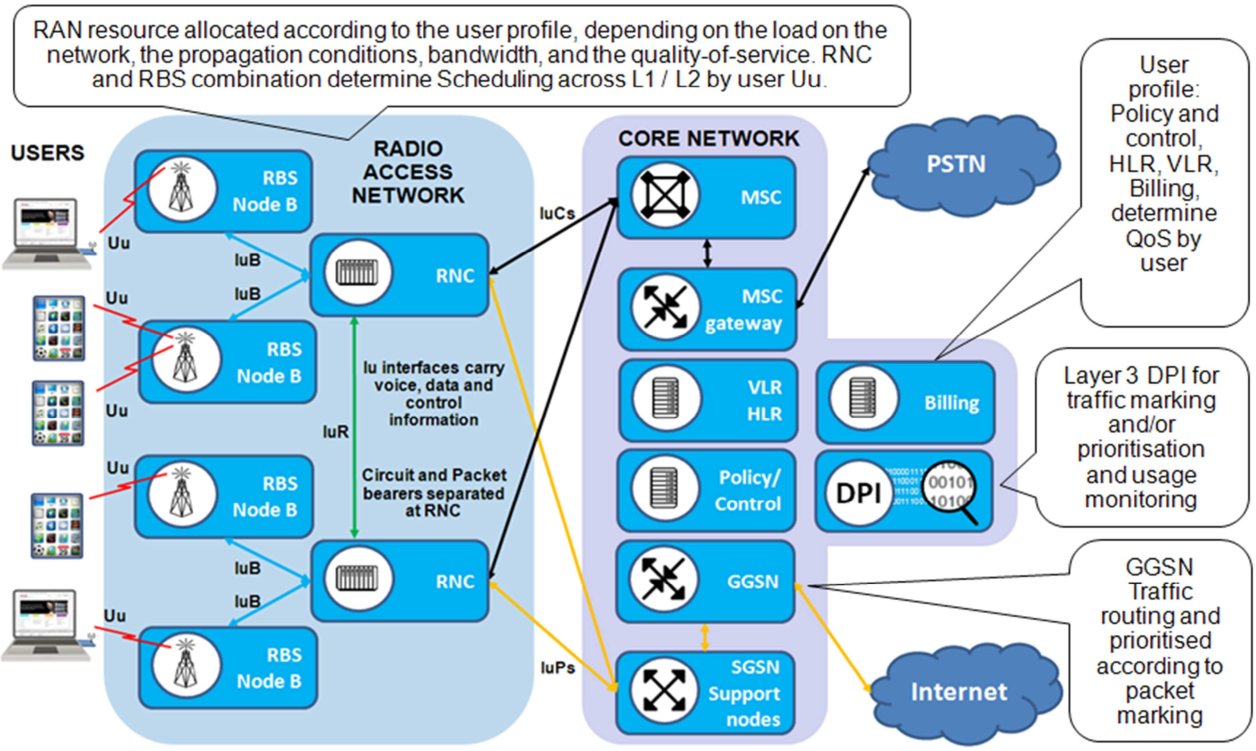

RAN resource allocated according to the user profile, depending on the load on the network, the propagation conditions, bandwidth, and the quality-of-service. RNC and RBS combination determine Scheduling across L1 / L2 by user Uu.
USERS
RADIO ACCESS NETWORK
RBS Node B
RNC
Uu
IuB
IuR
IuCs
IuPs
Iu interfaces carry voice, data and control information
Circuit and Packet bearers separated at RNC
CORE NETWORK
MSC
MSC gateway
VLR HLR
Policy/ Control
GGSN
SGSN Support nodes
Billing
DPI
PSTN
Internet
User profile: Policy and control, HLR, VLR, Billing, determine QoS by user
Layer 3 DPI for traffic marking and/or prioritisation and usage monitoring
GGSN Traffic routing and prioritised according to packet marking

Figure 23: Typical WCDMA Mobile Network Architecture

Figure 23 above shows the architecture of a typical WCDMA (3G) mobile network. The main determinant of QoS in mobile networks is the available radio bandwidth. This will be constrained by:

- The amount of radio spectrum available to the mobile operator
- The intensity of installed infrastructure (i.e. how large the cells are)
- The transmission protocol used (number of bits/Hz carried)

In 2.5G (GPRS) networks, operators will typically pre-allocate a small proportion of radio network capacity to GPRS data. This will be shared amongst data users on each cell. In 3G networks, voice calls are often given priority over data, which means that data capacity will vary with voice loading on each cell. On some networks voice traffic may restrict data capacity to the point where data users are offloaded onto a GPRS cell in the same area.

Typical methods of traffic management employed by mobile operators include use of ‘traffic consumed’ counters, which measure the amount of data generated and consumed by users. These reside in the core network (e.g. on the GGSN) and communicate with RNCs which instruct the RAN to drop data rates when consumers approach or exceed pre-defined limits.

Mobile operators also have the ability to limit usage of traffic types (e.g. P2P) within their core network and to limit data entering their networks.

DPI located in the core is used to identify traffic and packets are marked according to a traffic management strategy controlled centrally (Policy & Control). Bandwidth allocation will be implemented in the RAN by using the information provided by the packet marking and the user profiles instructions also derived from the traffic management strategy implemented through the Policy and Control centre.