Final
2016-06-09

NetWorld 2020 ETP

Service Level Awareness and open multi-service internetworking
- Principles and potentials of an evolved Internet ecosystem

White Paper
(Including Appendix)

Executive Summary

The value creation of the current Internet is substantial and growing, both directly and indirectly. It constitutes a fundamental building block for many businesses. The basic Internet access service constitutes the dominant business for an increasing number of telecommunication companies (Telcos) and Internet Service Providers (ISPs). The current Internet is very efficient in IP packet transport and very suitable for elastic applications, which can adapt to the varying network conditions and are not highly demanding in terms of network performance. However, the visions and expectations for the future and the 5G networks go far beyond what the Internet can support today. The shortcomings of the IP backbone inter-domain networking (internetworking) solution as of today should get more attention.

The all-IP property of the 4G network will also be fundamental to the 5G solutions. This will demand the consideration of principles and properties of the basic Internet access service, together with the overall principles, properties and potentials of the 5G portfolio of multi-service networking solutions. This is fundamental, as the all-IP traffic and the various 5G services, whether an access service, an end-to-end value added connectivity service or a higher-level application service of a certain vertical, will in general share the same physical infrastructure resources.

While recognizing and protecting the strong properties of the current Internet and its basic Internet access service, it is also important to be aware of other inherent properties and even shortcomings of the current Internet and of the single-traffic-mode operation (aka. best-effort). These shortcomings, such as the inherent dilemma of buffering (larger buffers increase statistical multiplexing gain, link filling and resource utilization, while at the same time increasing the delay) and the lack of protection of critical traffic, must be considered, both technically and in terms of the basic Internet business models. Any solution that proposes to mitigate the shortcomings must ensure that the well-functioning and the strong properties of the current Internet are not compromised.

This whitepaper proposes key principles and properties of future open multi-service internetworking that balances the properties of the basic Internet access service and those of future Value-Added Connectivity Services (VACS) in a constructive way. The fundamental motivations are three-fold: i) creating higher value for the customers and ensuring customer choice, ii) achieving higher efficiency of the network resources used, and iii) unleashing the innovation potentials of value added and end-to-end assured connectivity services from any end-point to any end-point on the Internet.

These innovation potentials, both those foreseen and those not yet discovered, are anticipated to be substantial. Network Service Providers (NSPs) have great innovation potentials, but there are even greater potentials for the larger volume of emerging and evolving Online (digital) Service Providers (OSP, or often called OTT) and SMEs that can now innovate on-top-of such value-added connectivity services. From these connectivity services they can develop their businesses and create assured quality and differentiated digital and online services for the benefit of their customers, considering a variety of consumer, business and public sector markets.

However, there are substantial challenges in establishing a cost-effective and flexible solution for the future that can accommodate the foreseen variety of applications, and evolve into an effective approach at a global scale. It is important to recognize that this is not just a challenge for individual

Telcos or ISPs. This is an industry wide coordination challenge, or even a society-wide coordination challenge. In order to succeed, the society at large, including the regulatory authorities, must welcome the emerging solutions as legitimate and as a good balance to the dilemmas we are facing as an industry and as a society.

The goal of this whitepaper is to contribute to a vision for the future that is needed by the industry and the society at large. Such a vision is needed in order to meet the variety and range of requirements and demands from different applications and how they can best be aligned with the most suitable type of connectivity and service level in order to fulfil the on-demand requirements of consumers, business and the public sector. A well-balanced traffic management scheme and a managed approach to inter-domain traffic exchange are fundamental solution elements. This will enable the most resource and cost efficient solutions, which again will increase innovation, industrialization and social welfare. In this way, the full exploitation and innovation potentials of Horizon2020 and 5G research and innovation activities can be achieved.

In summary, the whitepaper reflects upon the need for evolving the Internet networking and ecosystem, and suggests desired properties and goals of an open multi-service internetworking approach. It provides an outline of suggested solution elements, including potential baseline business models and approaches to service level awareness that are based on four main traffic modes: Basic Quality (BQ, aka. best-effort), Background (BG, aka. least-effort), Improved Quality (IQ) and Assured Quality (AQ). The whitepaper suggests and provides arguments for the overall properties and policies along these main modes, and concludes by recognizing short term opportunities and overall challenges ahead, while focusing on some selected research challenges. Details on these proposals, further analysis and accompanying elaborations are provided in the Appendix.

List of Contributors

Name	Company / Institute	eMail Address
Editorial Team Members		
Håkon Lønsethagen (Team Leader)	Telenor, Research	hakon.lonsethagen@telenor.com
Angeles Vazquez-Castro	Universitat Autònoma de Barcelona (UAB)	angeles.vazquez@uab.es
Felipe Gil-Castiñeira	Gradiant / University of Vigo	fgil@gradiant.org
Kostas Tsagkaris	Wings / Intelligent	ktsagk@wings-ict-solutions.eu
Maciej Muehleisen	ComNets, Hamburg University of Technology	maciej.muehleisen@tuhh.de
Murat Tekalp	Koc University, İstanbul	mtekalp@ku.edu.tr
Nikos Chavaranis	Wings	nxavar@wings-ict-solutions.eu
Olav Østerbø	Telenor, Research	olav-norvald.osterbo@telenor.com
Contributors and Reviewer		
Agnes Salvatori	Airbus Defence and Space	agnes.salvatori@airbus.com
Alexander Geurtz	SES	Alexander.Geurtz@ses.com
Andra Lutu	Simula Research Lab	andra@simula.no
Andreas Timm-Giel	ComNets, Hamburg University of Technology	timm-giel@tuhh.de
Anna D'Annunzio	Telenor, Research	Anna.DAnnunzio@telenor.com
Anne-Grethe Kåråsen	Telenor, Group	anne-grethe.karasen@telenor.com

Arnor Solberg	SINTEF	Arnor.Solberg@sintef.no
Artur Krukowski	Intracom S. A. Telecom Solutions	krukowa@intracom-telecom.com
Carlos Oliveira	Ubiwhere	coliveira@ubiwhere.com
Caner Aksoy	NETAS Telecommunication	caner.aksoy@netas.com.tr
Christine Leurquin	SES	Christine.Leurquin@ses.com
David Fernandez	INDRA	dfernandezp@indra.es
Diego Lopez Garcia	Telefonica	diego.r.lopez@telefonica.com
Dirk Trossen	InterDigital UK, Inc	Dirk.Trossen@InterDigital.com
Eliot Minn	E2E Services Ltd	eliotminn@e2eservices.co.uk
Fabrice Guillemin	Orange Labs	fabrice.guillemin@orange.com
Franck Fleurey	SINTEF	Franck.Fleurey@sintef.no
Franco Davoli	DITEN, University of Genoa	franco.davoli@cnit.it
Gino Carrozzo	Nextworks	g.carrozzo@nextworks.it
Guidotti Giovanni	Finmeccanica	giovanni.guidotti@finmeccanica.com
Isabelle Tardy	SINTEF	Isabelle.Tardy@sintef.no
Jean-Sébastien Bedo	Orange Labs	jeansebastien.bedo@orange.com
Joan Manuel Cebrian	INDRA	jcebrian@indra.es
Jordi Domingo	Universitat Politècnica de Catalunya (UPC)	jordi.domingo@ac.upc.edu
Klaus Moessner	Univ. of Surrey	k.moessner@surrey.ac.uk
Mats Eriksson	Arctos Labs Scandinavia AB	mats.eriksson@arctoslabs.com
Maurizio Caruso	E2E Services Ltd	mauriziocaruso@e2eservices.co.uk
Michele Luglio	University of Rome "Tor Vergata"	luglio@uniroma2.it

Necip Gozuacik	NETAS Telecommunication	gozuacik@netas.com.tr
Niclos Chuberre	Thales Alenia Space	nicolas.chuberre@thalesaleniaspace.com
Olav Nedrelid	Domos Labs	olav@domos.no
Patrick Waldemar	Telenor, Research	Patrick.Waldemar@telenor.com
Philippe BOUTRY	Airbus Defence and Space	philippe.boutry@airbus.com
Rui A. Costa	Ubiwhere	rcosta@ubiwhere.com
Tiago Batista	Ubiwhere	tbatista@ubiwhere.com
Valerio Frascolla	Intel Mobile Communications GmbH	valerio.frascolla@intel.com
Özgü Alay	Simula Research Lab	ozgu@simula.no

Table of Contents

Executive Summary 2
List of Contributors 4
Table of Contents 7
I. The need for evolution of Internet networking 9
II. Desired properties and goals of open multi-service internetworking 12
  a. Customer centricity and customer choice 13
  b. Unleashing the innovation potentials for all 13
  c. Improving the network efficiency 14
III. Outlining key solution elements 15
  a. Four main traffic modes 15
  b. Service level awareness 17
  c. Open Multi-Service Internet 18
  d. Business models and charging principles 20
IV. Challenges going forward 21
V. Concluding remarks 23
Appendix 25
  1) Properties of the current Internet networking 25
  2) Regulation and Net neutrality 27
  3) Product areas benefitting from Value Added Connectivity Services (VACS) 28
    a) Rich and Unified Communication Services 28
    b) Live events and content delivery 29
    c) Business Cloud Services 29
    d) "Tactile Internet" and Advanced Digital Consumer Services 29
    e) Internet of Things and M2M 30
  4) Dynamic traffic management and efficiency gains 31
    a) Example case: Conference hotel 31
    b) Admission control and network performance monitoring 33
    c) Conference Hotel – Case description 34
  5) Service Level Awareness and User Interaction 38
    a) Service Level Indicators in real-time 38
    b) User Interaction with the Network 39
  6) Managed and Assured Service Quality Traffic Exchange 39

7) Research and Innovation Challenges 41
a) Service level awareness and user experience in softwarized network infrastructures 41
b) Control and energy-efficient engineering of multi-class network information flows........... 42
c) Need of standardization 43

I. The need for evolution of Internet networking

The Internet has become a fundamental infrastructure of today's economy, with businesses of great variety creating trillions of dollars in revenue, and with value creation continuing to grow, both directly and indirectly. The basic Internet access service constitutes the dominant business for an increasing number of Telecommunication Companies (Telcos) and Internet Service Providers (ISPs).

The current Internet is very efficient in IP packet transport and very suitable to elastic applications that can adapt to the varying network conditions and that are not highly demanding in terms of network performance. However, the visions and expectations for the future and the 5G networks go far beyond what the Internet can support today.

The all-IP property of the 4G network, that will also be fundamental to 5G solutions, demands the consideration of principles and properties of the basic Internet access service, together with, and as an integral part of, the overall principles, properties and potentials of the 5G portfolio of multi-service networking solutions. This is fundamental, as the all-IP traffic and the various 5G services, whether an access service, an end-to-end value added connectivity service or an higher-level application service of a certain vertical, will in general all share the same physical infrastructure resources.

While recognizing and protecting the strong properties of the current Internet and its basic access service, it is also important to be aware of other inherent properties and even shortcomings of the current Internet, the associated inter-domain networking (internetworking) solutions and the single traffic-mode operation (aka. best-effort). These shortcomings must be considered, such as the inherent dilemma of buffering, which allows better link utilization via the increase of statistical multiplexing gain, link filling and resource utilization, at the expense of larger delay, the lack of protecting critical traffic and less than optimal routing properties. The analysis of the shortcomings must take into account both the technical and the basic business models of Internet networking, along with the resulting economic properties. Any solution that proposes to mitigate the shortcomings must ensure that the well-functioning and the strong properties of the current Internet are not compromised.

The best-effort and single-traffic-mode nature of the current Internet delivers a varying performance end-to-end and the Quality of Experience (QoE) for many applications is unpredictable. It is not well suited to cope with the high demands and potentially conflicting requirements of the increasingly diversified applications and services of the future. Problems, as experienced by the end-users[1], can for instance be too low screen resolution and undesirable end-to-end buffering delays, distortion of the visual images, online gaming delays or unpredictable file upload duration. SMEs that consider new cloud services may find the Internet based service offerings too unreliable for their critical business functions, while the traditional premium provisioned VPN based solutions are inflexible and too expensive. The network operators are faced with the dilemmas around overprovisioning and the

[1] The notion of end-customer (the customer paying for the service offered to the end-user or consumer of the service) may be used. The notion of end-user might be different from the end-customer. For simplicity, this distinction between the two is not elaborated upon, and in general considered out of scope of the whitepaper.

best-effort nature of the Internet[2], which prevents solutions for efficient provision of capacity to time critical traffic. For more information on shortcomings, see the Appendix, Section 1) below.

Nonetheless, the Internet end-users are demanding higher and higher speed and better performance along with the continued technology performance improvements, higher broadband access speeds, and performance developments within computer, smartphone, software and device technologies. The rapid emergence of smart devices, video streaming and the Internet of Everything (IoE) technologies create new demands for broadband and Internet services. In addition to the QoE uncertainty of the current Internet, there are advanced services that today cannot be provided in a feasible way, such as high-definition real-time video communications, assured low delay service, demanding business critical cloud Internet services, as well as advanced live event content delivery services.

For many years, the above shortcomings have been recognized and researched, but differentiated and assured quality connectivity services have so far not been developed, enabled and offered into the market as an open and general Internet scale offering. We can easily agree there is a chicken-and-egg challenge here, as well as a regulatory dilemma. When there are no such products offered to the market, the customers cannot express their demands and hence, the rationale for developing industrialized and user-friendly solutions cannot be easily provided. Moreover, the regulatory conditions for evolving the Internet solutions and introducing a greater variety in access and end-to-end network services have been unclear in the light of net neutrality issues. The optimal solution requires complex multi-actor alignment, both technically and in terms of business models.

These dilemmas trigger and motivate this whitepaper. The whitepaper proposes key principles and properties of future open multi-service internetworking that balance, in a constructive way, the properties of the basic Internet access service and those of future value added connectivity services. The fundamental motivations are three-fold: i) creating higher value for the customers and ensuring customer choice, ii) achieving higher efficiency of the network resources used, and iii) unleashing the innovation potentials of value added and end-to-end assured connectivity services from any end-point to any end-point on the Internet. An illustration of the innovation potential is provided in Figure 1 below.

---

[2] Note that while *best-effort* in practice implies a single "best-effort-class" for handling the Internet traffic, this is from the internetworking and end-to-end perspective of the Internet. The Internet infrastructure itself is provided by individual ISPs whose network infrastructures (intra-domain) do support multiple network services and traffic classes as well as more advance means of traffic engineering (for instance based on DiffServ and IP/MPLS) where for example business VPN traffic can be protected and given priority over the Internet traffic. For private managed IP networks there are also multi-service inter-domain solutions available. However, these network operating modes are not made available as part of the current Internet.

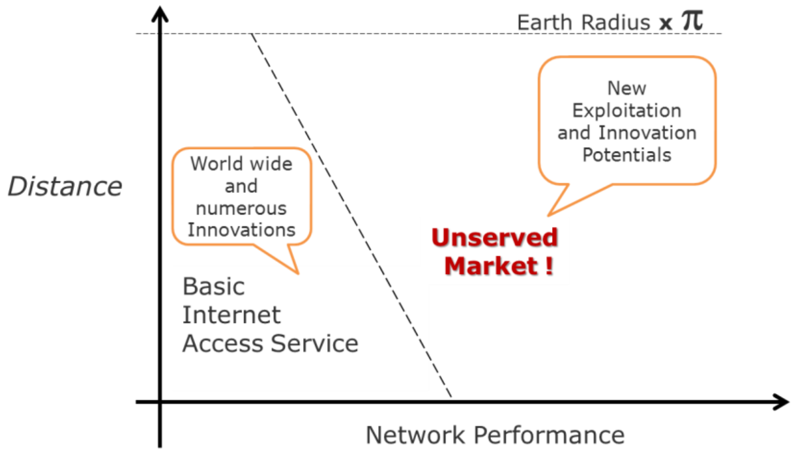

Figure 1 Unserved Market – Innovation Potentials

As shown in Figure 1, the Internet connectivity that can be accommodated by the basic Internet access service has some natural limitations. The typical experience of the end-customer is that given any situation (whether fixed, fixed wireless, nomadic wireless or highly mobile) the network performance has limitations. This can be in terms of throughput, delay (latency), delay variation (jitter) as well as service availability. By not having a fixed scale on the network performance axis, the illustration indicates that these limitations vary over time and according to the settings. Moreover, the figure illustrates that the typical user experience is such that, as the remote end-point is further away (increase in distance), the typical throughput will decrease and the jitter will typically increase. The decreased throughput is partly due to an increase in the likelihood of packet loss.

These innovation potentials are anticipated to be substantial. That is, the NSPs have great innovation potentials, but there are even greater potentials for the larger volume of emerging and evolving Online (digital) Service Providers (OSP, or often called OTT) and SMEs. These can now innovate on-top-of such value added connectivity services, develop their businesses and create assured quality and differentiated digital and online services for the benefit of their customers, considering a variety of consumer, business and public sector markets. The offering of assured quality connectivity is in particular important to the SMEs and OSPs that are not able to build large or global scale private and managed quality IP backbone networks that can mitigate the current Internet shortcomings. The general, standardized and open offering of assured service quality connectivity will enable the SMEs and the OSPs to differentiate and monetize their online content and application offerings. They will open up new innovation potentials for many businesses and SMEs that we currently have not even seen.

However, there are substantial challenges in establishing a cost-effective and flexible solution for the future, which can accommodate the foreseen variety of applications and evolve into an effective approach at a global scale. It is important to recognize that this is not just a challenge for individual Telcos or ISPs. This is an industry wide coordination challenge, or even a society-wide coordination challenge. In order to succeed, the society at large, including the regulatory authorities, must welcome the emerging solutions as legitimate and as a good balance to the dilemmas we are facing as an industry and as a society.

The goal of this whitepaper is to contribute to a vision for the future that is needed by the industry and the society at large. Such a vision is needed in order to meet the variety and range of requirements and demands from different applications. It is necessary to determine how each application can best be aligned with the most suitable type of connectivity and service level in order to fulfil the on-demand requirements of consumers, businesses and the public sector. A well-balanced traffic management scheme and a managed approach to inter-domain traffic exchange are fundamental solution elements. This will enable the most resource and cost efficient solutions, which again will increase innovation, industrialization and greater social welfare. In this way, the full exploitation and innovation potential of Horizon2020 and 5G research and innovation activities can be achieved.

In summary, the whitepaper reflects upon the need for evolving the Internet networking and ecosystem. It suggests desired properties and goals of an open multi-service internetworking approach and provides an outline of suggested solution elements, including potential baseline business models and approaches to service level awareness that are based on four main traffic modes: Basic Quality (BQ, a.k.a. best-effort), Background (BG, a.k.a. least-effort), Improved quality (IQ) and Assured Quality (AQ). The whitepaper suggests and provides arguments for overall properties and policies along these main modes. It concludes by recognizing overall challenges ahead, while focusing on some research challenges. Details on these proposals, further analysis and accompanying elaborations are provided in the Appendix.

Whether or not the notions of “***Evolved Open Internet***” and “***Open Multi-Service Internet***” should be used is also a public relations and regulatory issue that can potentially become a stumbling block. Bear in mind that regulatory authorities may choose to use other terms for future regulation framework policies. Hopefully, these terms and the below concepts and proposals can still be semantically valid, while other terms and wordings may finally be globally recognized and used.

II. Desired properties and goals of open multi-service internetworking

This section and the rest of the whitepaper is assuming that the terms “***Open Multi-Service Internet***” or “***Evolved Open Internet***” are both valid and legitimate terms, denoting the networks of networks that carry the mix of traffic from both the basic Internet access service and the multiple traffic types, as needed by the different advanced services of the future evolved Internet. The traffic modes and the principles suggested below are intended to accommodate the future needs for a variety of connectivity types and services, where the internetworking solutions can be established in a way that optimizes network resource utilization and that can enable overall cost efficiency.

The desired properties and goals of the future and Evolved Open Internet should recognize three fundamental perspectives in a well-balanced and effective way, as suggested below. These identified key properties or goals lead to several solution implications, as well as research and innovation challenges. Many of the solution implications are addressed further in the paper’s main sections, while a selected set of relevant research challenges are addressed in the Appendix.

a. Customer centricity and customer choice

The variety of service levels, the main traffic modes introduced above (BQ, BG, IQ and AQ), and the overall dynamic traffic management approach should be transparent to the end-users through service level awareness mechanisms and indicators. These mechanisms and indicators and the more or less managed quality level for all services, including that of the basic Internet access service, should be provided in ways that are feasible for each of the specific application types and end-user settings. The anticipated service level indicators (further details are provided below) enable a richer set of options and application variants to be provided to the end-users, accommodating the different settings and urgencies of the user and the user context.

The choice of preferred service level and traffic mode can then be flexibly and dynamically made according to settings, urgency and preferences of the end-user. The selection is then made either directly, according to preferred connectivity mode, or indirectly, through the choice of application, application variant or application service level. The service levels should be measured from the specific end-user perspective by appropriate and standardized service KPIs (Key Performance Indicators), e.g. assessing quality and delay of video flows.

In addition, reasonable measures for monitoring and publishing the general service levels of the basic Internet access services (i.e. the BQ mode) of the different types of broadband service providers should be established in a way that covers a representative set of locations and end-user settings. By effectively monitoring the general quality level of the BQ mode and enabling the suggested service level indicators for end-user devices, a sufficient transparency can be achieved. In this way, it will be possible to ensure that the NSP policies for provisioning the other traffic modes and value added connectivity services will not be to the detriment of the basic Internet access services.

(Note: This future anticipated and suggested situation is very similar to today's situation, where radio access network resources for the provisioning of telco voice calls have priority over the broadband data service. However, the mobile operators typically ensure that new voice calls can be rejected if the limit of calls has already been reached. In this way the broadband data service is maintained within reasonable bounds.)

b. Unleashing the innovation potentials for all

SMEs and enterprises require reliable connections with predictable performance for running their business and for offering digital and online services to their customers, anywhere and at any time. The networking solutions for the future should enable and support value added connectivity services and features for innovation in new advanced and managed quality online application services for a variety of markets. This will enable the OSPs of different verticals to create and offer their differentiated and assured quality service goods profiled to their targeted customer segments. It is important to mitigate the limitations of the current networking solutions addressing the SME and enterprise markets by offering and unleashing the innovation potentials of end-to-end assured connectivity services from any end-point to any end-point on the Internet.

This implies that any SME or OSP should be able to get access to Value Added Connectivity Services (VACS) through standardized APIs. They should be able to interact with their preferred partner NSP, which can be a one-stop-shop that is able to interact with subsequent NSPs involved in the VACS offering needed to support the services managed or offered by the SMEs and OSPs.

Customer friendly charging principles should be established, enabling pricing of VACS and advanced application services, where what is paid for is reasonably in line with the resources consumed end-to-end. This should also include the option of having the initiating party paying for two-way traffic between the communicating end-points. That is, business models and technical support for end-to-end money flows across the actors of the Internet value chain should be supported. This is discussed in more detail in Section III - d below.

A major challenge for performing such a big evolutionary step is to ensure that domestic or local (media) information formats are transformed into common IP-format (including the broadcasted information) in order to enable consistent end-to-end behaviour.

c. Improving the network efficiency

Value creation and resource usage efficiency from differentiated traffic, enabled by Quality of Service (QoS) mechanisms in the customer premises equipment and their local customer networks, should be taken good care of and supported by dynamic and reasonable traffic management at the network edge of the NSP. Basically, these value creation and traffic handling mechanisms should be recognized and carried end-to-end across the backbone networks. Likewise, the energy efficiency mechanisms and support should be an integral part of these end-to-end evolved Internet capabilities. The energy consumption of the Internet should be considered as a management goal, to be suitably traded off with user QoE and preferences.

It is necessary to design networks that are able to adapt the consumption of resources, not only to the number of users, but also to the required quality for the services they are demanding. Basically, the best cost efficiency is achieved by having a mix of all traffic types being facilitated, under reasonable traffic management, by shared physical infrastructure elements, including switches, routers and links. An example case (the so-called "Hotel Case"), including numerical figures, is provided in the Appendix, Section 4) below.

The evolution of the Internet towards a multi-service era necessitates the adoption of dynamic and reasonable traffic management at the network edge. This must be applied across the customer gateways, the access links, the aggregation segments and the service provider edge nodes, and should be carried end-to-end through correspondingly feasible and matching Assured Service Quality (ASQ) interconnection services, solutions and business models. This will enable novel solutions and business models, expanding the Internet ecosystem.

In conclusion, the challenge is to optimize the dynamically varying demand from a variety of users, their applications and corresponding service types, and accommodating this demand in the most resource efficient way. A network should be well-dimensioned in order to accommodate the general anticipated demand, with a good or excellent quality perceived by the user. In the worst-case scenario, where the network is not able to provide the expected performance (for example, in case of link failure or because a user is moving to an area with poor wireless coverage and capacity), it should alert the user and offer a chance to handover to another application mode or another application which requires less resources, or even to finish the session or connection in a gracefully and controlled manner.

III. Outlining key solution elements

The above sections have provided motivation for evolving the networking solutions and pointed out desired overall properties and goals. This section introduces what are believed to be key solution elements of such an evolved Internet networking approach. It presents the four main traffic modes, elaborates on service level awareness, introduces the concept of "Open Multi-Service Internet" and suggests a baseline business model and some associated charging principles.

a. Four main traffic modes

The four main traffic (connectivity[3]) modes have already been mentioned. Arguments for the importance of being able to match the properties of the connectivity with the different specific needs of the applications are given above. Being able to carry a mix of different traffic types with different properties over the same physical capacity is a very efficient way of carrying traffic. This can enable a substantially higher network resource utilization and efficiency gain as compared with carrying just a single traffic type over the same physical capacity.

Depending on the specific settings, the optimal number of traffic types (a.k.a. Traffic classes, or QoS classes) can vary. For instance, the 4G or LTE standardization has identified thirteen classes or QoS Class Identifiers[4] (QCIs). However, for the purpose of this whitepaper (with a wide and general audience) we have identified key traffic differentiation properties and effects and what we call the "four main traffic (or connectivity) modes". These four modes give us the vocabulary needed to carry across several key messages on the main properties and positive effects of traffic differentiation. We also recognize that these positive effects will come as a result of also ensuring a dynamic and reasonable traffic management.

First, we identify the ***Basic Quality mode (BQ)***, which we associate with the current Internet access service and the current Internet traffic (aka. best-effort). This is the dominant or only traffic mode today, used in particular for interactive web-browsing, interactive social media, and basic streaming services. We expect this mode to be the dominant and most used for the foreseeable future, due to having the highest efficiency and the lowest cost of end-to-end IP packet transport. The applications

---

[3] The notions of "traffic mode" and "connectivity mode" are used somewhat interchangeably. However, the notion of "traffic mode" is used when addressing aggregates of traffic flows, such as at the traffic exchange points, while the notion of "connectivity mode" is often used while speaking of the end-customer connectivity (flow).

[4] 3GPP TS 23.203 Policy and Charging Control Architecture

that are suitable for this mode are elastic applications, which can adapt to the varying network conditions and are not highly demanding in terms of network performance.

The next mode in our line of arguments is the ***Background mode (BG)***. This mode is important to avoid having non-critical data interfere negatively with the BQ traffic. The BG mode (aka. least-effort) is useful for non-critical traffic or applications usage, such as email synchronization, background backup or file up-/download. The BG mode should be an extension of the Internet access service.

On the other end of the scale we identify the ***Assured Quality mode (AQ)***. This traffic mode is used when absolute (specific) service level guarantees are requested, provided or delivered. This traffic mode is used for applications that are in need of absolute predictable network performance, in terms of throughput, delay (latency), delay variation (jitter) or packet loss. In addition, the AQ mode can be associated with various service availability levels. Example applications that can find the AQ mode suitable are Telepresence, HD Voice, Assured Quality Live Video Streaming, or Critical business applications. Typically, the AQ mode is supported by explicit Resource and Admission Control (RAC) functionality in order to control the network performance and signal back to the user in real-time whether or not the service is available, e.g. signal back a "busy tone" when resources are not available for the specific AQ mode service requested.

In addition to these three modes, the ***Improved Quality mode (IQ)*** should also be supported. This is motivated by the need for higher efficiency and lower costs. While the AQ mode is rather costly to implement, the IQ mode can be implemented in less costly ways, while still having the benefits of allowing somewhat improved network performance, in relative terms, as compared with the BQ mode. The increased performance level compared with BQ will depend on the operator policy, product profile and the general network usage conditions.

Note that while IQ, BQ and BG are all given relative priority among each other, the AQ mode is given absolute priority. In this whitepaper, we do not go into the details and complexity of how the modes can be mapped to traffic classes and how these traffic classes are handled by the specific network QoS mechanisms. However, we do recognize that the IQ mode may come in several flavours such as for lower delay or for higher throughput. The four modes are summarized and ordered according to network performance in Figure 2 below.

The four traffic modes will give the NSP the tools for new innovations into connectivity services that in turn can accommodate new innovations for the OSP. The OSPs can now differentiate and offer new products, even with assured service quality. This will benefit OSPs belonging to any kind of vertical product area, even new OSPs offering advanced and demanding services based on high-performance connectivity services or based on low-end innovations, where the BG mode can allow additional savings.

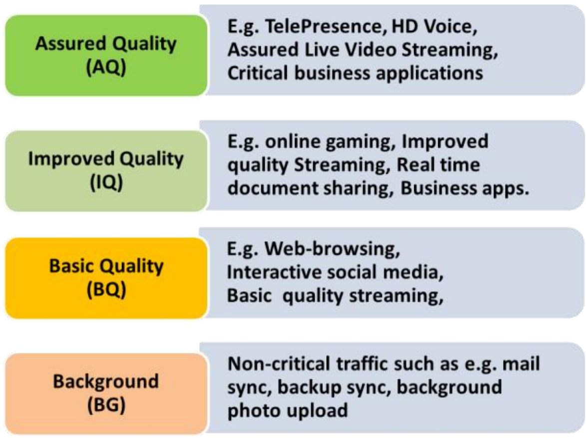

Figure 2 Four main traffic (connectivity) modes, with example applications

Appendix Section 3) below follows up on how these modes can provide the basis for new innovation, while Appendix Section 4) below discusses how these traffic modes can enable higher network utilization and create additional value for the customers while lowering the overall cost. Figure 3 below illustrates the positioning of the four modes in terms of network performance. The BQ mode corresponds with the Internet access service, complemented by the BG mode, while IQ and AQ can be associated with value-added services. Furthermore, it is anticipated that the value-added services have a pricing structure such that what is paid for is somewhat in accordance with the resources used by the service.

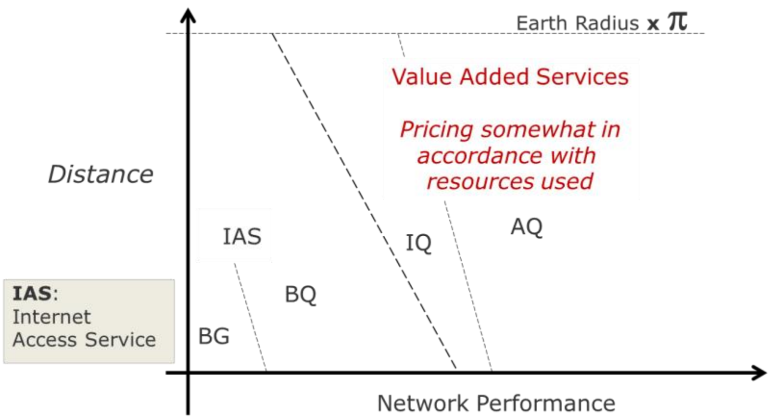

Figure 3 Four main traffic (connectivity) modes. Accommodating different network performance needs

b. Service level awareness

In this section we take a closer look at what the four connectivity modes can mean for the end-user. That is, Service Level Awareness is considered from a customer-centric point of view. It is valuable or even important for the end-user to know the application performance to be expected from a particular network at a particular point in time. This concern is quite pertinent already with today's Internet (and for what we here call the BQ mode) and the numerous applications already provided.

The typical questions or concerns a user has are for instance: What is the performance of the network or the application I would like to use right now? For instance while riding the train or bus: Should I even bother to click on this video news clip on my smartphone, or will I just get annoyed by the poor experience or perhaps it will not work at all? If I pick up my tablet, will I then get a good experience with the larger screen size? Or, when being at home watching an interesting video on my PC: Can I get a good or excellent experience if moving onto my 55" HD TV, or is it not even worth the try?

Future needs for more clever interaction with the end-user, as well as giving the user a possibility to communicate intents or a flexible and efficient way of "signalling" service requests, constitute a whole new area that has been given little attention up until now, as these questions are more or less incompatible with the current Internet way of working.

In Appendix Section 5) below we provide some initial ideas and food for thought on this broad topic. Like for a car, one needs some common understanding of "things" that one interacts with, like the steering wheel, the speedometer, and the gear-shift. So, what are the common Service Level Awareness and Indicator concepts and human-machine interfaces needed for the future? This is a key topic raised and discussed in this whitepaper. However, while some common and general concepts are needed, it should also be possible to add solution-specific features and design properties to allow for both competition and creative designs in user interaction.

c. Open Multi-Service Internet

From the above discussion we propose the following concepts and illustration to pull together the four traffic modes (or customer-focused connectivity modes) and the various wholesale focused internetworking solutions, anticipating that the notion of ***Internet*** can also be used in complementary ways beyond the current usage of the term. More specifically, by the illustration in Figure 4 below we suggest a way to understand the notion of ***Open Multi-Service Internet***.

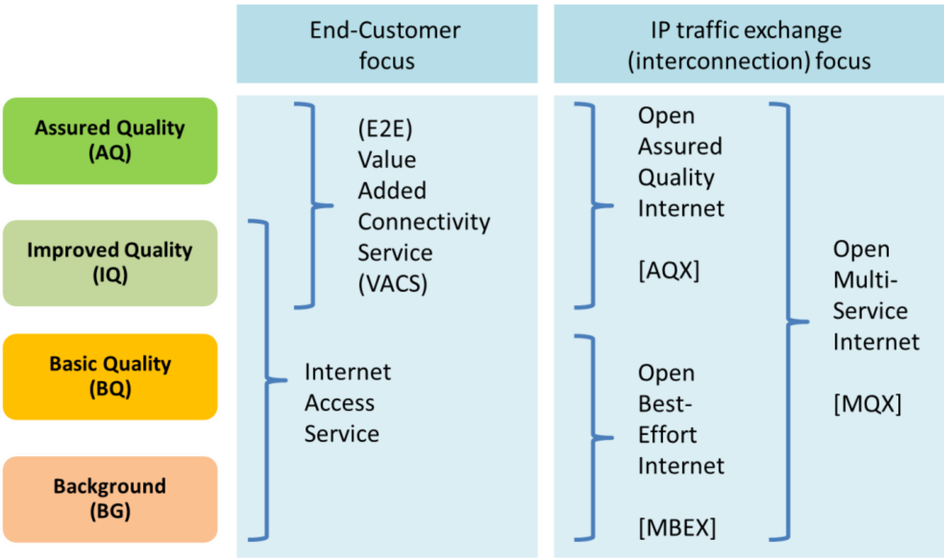

Figure 4 Elements of an Open Multi-Service Internet approach

This illustration should be considered with the following understanding and accompanying notes. Consider first the "end-customer focus". We recognize the connectivity modes that are offered as part of the basic Internet Access Service on one hand, and the connectivity modes that are associated with end-to-end "Value Added Connectivity Services" (VACS) on the other hand. It is suggested that the IQ mode can come in two variants. That is, a) it can be offered as an IQ access connectivity, to be considered as a potential future extension of the general Internet Access Service, and b) it can be offered as an end-to-end IQ connectivity that is realized by a traffic class having an end-to-end implementation. Compared with the AQ (which is solely an end-to-end connectivity mode and concept), end-to-end IQ connectivity is realized with a more relaxed resource management scheme, in order to avoid some of the more detailed signalling and enforcement cost elements that the AQ mode typically requires.

Next, we take a closer look at the interconnection or IP[5] traffic exchange part and the wholesale solutions to accompany these end-customer connectivity offerings. We first make some observations regarding the current situation. Connectivity that requires support of AQ mode solutions, such as the Telco voice service, is facilitated by a managed quality private IP/MPLS backbone separate from, and parallel to, the current (best-effort) Internet, and typically sharing the common infrastructure capacities with the Internet traffic. Moreover, many of the Internet services are supported by "back-office" server-to-server (datacentre-to-datacentre) connectivity. This traffic is carried by other managed quality private IP/MPLS backbone networks for instance to facilitate Content Delivery Network (CDN) traffic or other "back-office" traffic of the global OTTs. Also in these cases, the quality managed traffic typically shares the common core infrastructure capacities with the Internet traffic, and the NSP must also today carefully manage a mix of best-effort Internet traffic and the managed quality IP traffic.

---

[5] The IP traffic exchange solution may also be complemented by an interconnection solution for sub-IP layer technologies.

The solution approach addressed in figure 4 is the complementary and anticipated (yet to be realized) Open Assured Quality Internet that is able to accommodate the AQ and the IQ traffic modes, supporting managed quality connectivity (i.e. VACS) offerings across network domains, from any end-point to any end-point of the Internet. Such an approach is considered as very efficient and suitable for the future evolved Internet ecosystem. This future open assured Internet must be facilitated by a new Assured Quality traffic eXchange (AQX) business model and solution, as suggested in the next subsection. Moreover, the combination of this AQX traffic exchange solution and the best-effort (BQ) traffic exchange can be considered in combination as the overall traffic exchange solution for the Open Multi-service Internet, as illustrated.

Moreover, we anticipate that also the best-effort traffic exchange will evolve and potentially become a more managed traffic exchange solution (Managed Best-Effort traffic eXchange - MBEX), where acceptable but still more flexible levels of congestion and delays will be suggested and agreed for different network segments, so that also the BQ and BG traffic modes are monitored more closely and handled in satisfactory ways. See also the discussion in Appendix Section 4) below. Hence, it is anticipated that the overall traffic exchange solution of the open multi-service Internet becomes a Managed and assured Quality traffic eXchange solution (MQX).

The notion of “Open” suggests an open ecosystem where all actors, various NSPs and various OSPs can participate through standardized interconnection and interworking interfaces and APIs. These are then the wholesale interfaces between partner actors, although often these actors are also competitors. Moreover, standardized interfaces are also needed for end-point (host) applications to enable signalling and interaction with peer applications as well as with the network. So far, the demand and regulatory setting for such an open multi-service Internet has been unclear. However, as regulatory concerns are now getting clarified and new regulation is emerging, the increased demand for advanced applications and services that require consistent and predictable QoE can trigger changes to the industry. See also Appendix Section 2) below on “Regulation and net neutrality”. Applications such as telepresence and live-event content streaming services are two use cases that may trigger the “bootstrapping” of such a more holistic Internet approach.

d. Business models and charging principles

An overall objective is to develop business models and charging principles that can align the incentives across the diverse set of actors in the ecosystem, and alleviate the current tensions and disputes that are rather common in today’s Internet industry. The charging principles should be simple but sufficient and show a future path for the creation of new added value and advanced multi-provider services. They should better reflect that what is payed for is somewhat aligned with the resources consumed by the service, across the chain of service elements.

Appendix Section 6) below on “Managed and Assured Service Quality Traffic Exchange” and the section above on “Open Multi-Service Internet” suggest that the Assured Quality traffic eXchange (AQX) solution and business model should be simple but sufficient for its purpose. This will typically mean that there is no need for end-user flow awareness at the Point of Interconnection (PoI), even for the end-user connectivity offerings with managed end-to-end quality (i.e. for VACS). Accordingly, the Sending-Party-Network-Pays (SPNP) approach is suggested as the baseline approach among the NSPs for the AQX approach.

It should be noted that the business model for the wholesale offerings by the NSP to the OSP is a different and separate issue and is out of scope of this whitepaper. In general, we note that it will typically depend on the specific service type and the product area ecosystem (i.e. the specific vertical). However, some indications are indirectly provided by the anticipated end-customer and retail oriented charging principles for VACS in the following.

The anticipated general solution should prepare for a setting where on-demand and real-time end-to-end quality management of the end-user connectivity (i.e. VACS) can be satisfactorily handled by coordinating the policy control and enforcement at the service nodes of the edge NSPs serving the end-points that take part in the VACS. By these policies, the VACS traffic is steered onto the Assured Service Quality (ASQ) paths that carry the traffic aggregates across network domains. The suggested default charging principle for peer-to-peer VACS is one where the end-customer pays for the access part as well as for the transit traffic from his end. That is, from his NSP across any transit NSPs to the remote edge NSP. The end-customer of the opposite end pays for his access part, as well as for the transit traffic in the opposite direction (i.e. the traffic that is sent from the remote peer).

In addition, charging principles are also needed in order to support various scenarios where one end-customer (the “initiating end-customer”) can pay for traffic in both directions. In this case, the initiating party pays his NSP, and this NSP must then pay the remote NSP for the traffic in the opposite direction. This “Initiating Party Network Pays” (IPNP) principle enables support of end-to-end money flows. For instance, money may flow from end-points of customers willing to pay into network regions that are less well-developed, with end-customers that lack the financial strength to pay for VACS. Hence, also low-end customers and the NSPs in these regions may now benefit from VACS and the advanced applications enabled by such VACS.

Moreover, the charging principles and mechanisms should also accommodate the case where the end-customer is paying the OSP for advanced application usage, including the VACS part that enables this advanced application. This case will typically include some way of value sharing with the partner NSP that offers the wholesale service enabling the VACS. Standardized interfaces for such services and charging mechanisms[6] are needed to support this and the above cases, with their many variants, in an open way.

IV. Challenges going forward

There is one question that should be addressed. Why have the changes envisaged above not happened so far? The answer that the current best-effort Internet is the best to serve all foreseeable needs is most likely not the best one. On the contrary, to make such changes happen there are numerous inter-related and challenging elements that must be solved in a coordinated fashion. First,

[6] These interfaces and suggested charging principles are currently addressed by the EU project 5G Exchange. See http://www.5gex.eu/

there must be a positive business case for involved actors. However, it is rather difficult to predict the potential of a new product offering when there are so many uncertainties and several of the key solution elements (both at protocol and integrated architecture level) are currently unavailable. Moreover, to introduce a new product into the market with a higher potential than just that of a single-Telco offering requires new ways of assured quality traffic exchange (as discussed in this whitepaper). Hence, there are many multi-actor coordination challenges ahead, both technically and business wise, which furthermore introduce uncertainties to the business case. There are also enterprise internal issues regarding responsibilities across related products, product migration and re-organization issues that add to the complexity and uncertainties.

An additional hurdle is the uncertainties around net neutrality. A single-Telco initiative and approach would very likely result in severe negative publicity, as it is impossible not to run into net neutrality discussions. The same would be the case if a few Telcos collaborated and partnered to develop Internet style solutions for a general approach to broadband-based differentiated and assured quality connectivity services. These seemingly high-risk dilemmas do call for clarity on regulation and public acceptance up front. As already mentioned, this whitepaper aims to discuss and help mitigate these dilemmas and to propose goals and principles pointing to a balanced and hopefully acceptable way forward, not just for the Telcos but triggering innovation potentials for all actors in the ecosystem. It should also be noted that the regulatory uncertainties are much lower now that the final regulation by the European Parliament and the Council on the Open Internet has been published (on November 25, 2015). See Appendix Section 2 for further discussion.

Luckily, there is no need to solve all challenges and coordination issues at once. The ultimate and very best solutions do not need to be available from day one. There are several product opportunities and business cases that can be developed into positive cases without much difficulty, i.e. sufficiently positive to spark innovations, investments, deployments and new offerings, if the surrounding uncertainties and risks are lowered. Seemingly good cases are business cloud connectivity and assured Internet cloud services, assured quality telepresence and assured quality live-event content delivery, to suggest a few. In the short term, these product offerings are in particular important for value creation in the SME segment. (For a more elaborated discussion on bootstrapping issues and approaches, see the ETICS deliverable D2.3[7].)

Then, beyond bootstrapping, looking into the more advanced solutions foreseen by this whitepaper, we do see several research challenges. They need to be tackled as part of the bigger picture of upcoming techno-economic opportunities and challenges driven by new related networking concepts such as cloudification, softwarization or virtualization. Specifically, management of network resources and user experience over softwarized and reconfigurable networks call for powerful methods to both measure and adaptively track QoS and QoE. Mappings between QoS and QoE may moreover vary depending on the location and communication context. Optimization of resource usage and energy consumption may require sophisticated ways to gather statistical information of network dynamics at various scales, both in time and space.

[7] See ETICS web page: https://www.ict-etics.eu/publications/deliverables.html

Control and engineering of multi-class flows (connectivity) is also an important technical challenge, in particular in heterogeneous network environments. The efficient allocation of physical and/or virtual resources, while providing differentiated in-network treatment, will require the identification of ontologies of traffic as well as logical mappings among connections. Such mappings must be able to meet traffic demands and differentiation requests, while having control over the reliability of the aggregated and individual flows. Indeed, innovative concepts related to media coding and network coding seem indispensable to cope with the upcoming strong reliability, low cost and maintenance and high availability requirements. A fundamental issue also arises when it comes to preserving confidentiality in such a transmission flow processing context.

It is evidently of paramount interest for all parties to bring research and innovation into standardized and open frameworks. Standardization of new concepts such as service level awareness or multi-class traffic flows can allow the identification of design constraints agreed by different players. Moreover, standardized architectures of softwarized and virtualized networks at different scales across networks will be of fundamental relevance to incorporate multi-class resource and traffic management. The lessons learnt should be considered while addressing the well-known gaps between academic research and standardized solutions. See Appendix Section 7 for more details on research challenges.

V. Concluding remarks

This whitepaper looks into the future of 5G networks, anticipating the innovation potentials of many verticals, evolving existing ones and triggering new ones. The variety of requirements associated with the diversity of future applications and services cannot be well and efficiently supported and delivered over just a single-traffic-mode internetworking approach, as is an inherent property of the current Internet. While recognizing and protecting the strong properties of the current Internet and its basic Internet access service, it is also important to be aware of other inherent properties and shortcomings, both technical and with respect to issues regarding lack of incentive-compatible business models.

In particular, the whitepaper points to future approaches that can unleash the innovation potentials for services and network operations requiring higher network performance and more flexibility than what can be offered and enabled by the current best-effort Internet. These innovation potentials, both those foreseen and those not yet discovered, are anticipated to be substantial. The NSPs have great innovation potentials, but there are even greater potentials for the larger volume of emerging and evolving OSPs and SMEs. They can now innovate on-top-of such Value Added Connectivity Services (VACS), develop their businesses and create assured quality and differentiated digital and online services for the benefit of their customers, considering a variety of consumer, business and public sector markets and verticals.

This whitepaper proposes key principles and properties of a future open multi-service internetworking approach that balances, in a constructive way, the properties of the basic Internet access service and those of future Value Added Connectivity Services (VACS). The fundamental motivations are three-fold: i) creating higher value for the customers and ensuring customer choice,

ii) achieving higher efficiency of network resources usage, as well as iii) unleashing the innovation potentials of value added and end-to-end assured connectivity services from any end-point to any end-point on the Internet.

A key message is that Service Level Awareness should become an integral part of future access and connectivity services, application services and their offerings. Service Level Indicators should eventually be directly or indirectly available, even in real-time, to properly inform the customers regarding service offerings, availability and what choices are available to them to better serve their needs. We recognized that the importance, urgency and expectations of using a specific service changes from one instant to another. That is, the customer preferences are highly context specific. The future internetworking solutions and service offerings should be able to take this into account. We predict that the end results will be improved customer experience, network and energy efficiency, as well as boosting innovations across verticals and stakeholders.

Appendix

According to the structure above, the sections below elaborate on some of the key topics addressed in the main body of the whitepaper.

I. The need for evolution of Internet networking
   - Section 1 elaborates on the properties of the current Internet networking, inherently resulting from the one-traffic-mode-only approach, including also some business considerations.
   - Section 2 reports briefly on Regulation and net neutrality.

II. Desired properties and goals of open multi-service networking
   - Section 3 provides some examples of, and reflections around, product areas that can benefit from value added connectivity services.
   - Section 4 elaborates on the topic of reasonable and dynamic traffic management and the network resource efficiency gains that can be achieved. One simplified numerical example is provided (the so-called “Hotel Case”).

III. Outlining key solution elements
   - Section 5 provides examples and illustrations around Service Level Awareness and addresses briefly the topic of user interaction.
   - Section 6 addresses the topic of managed and assured service quality traffic exchange.

IV. Challenges going forward
   - Section 7 provides an indication on some advanced features of open multi-service networking that call for novel and challenging research directions.

1) Properties of the current Internet networking

The current Internet is very efficient in IP packet delivery, suitable to elastic applications that are not highly demanding in terms of network performance or network performance stability. The value creation of the Internet is substantial and growing. The overall capacity and throughput of the Internet will steadily increase and as such offer improved value and customer experiences. Protecting the value creation potentials of the open Internet on a fair basis is in the interest of all players.

However, it is also important to be aware of the inherent properties as well as the shortcoming of the current Internet. The best-effort and single-traffic-mode nature of the open Internet delivers unpredictable performance and QoE for many applications and usages. It is not suited to cope with the increasing and potentially conflicting requirements of the increasing diversity of applications and services of the future. Hence, in addition to the QoE uncertainty property of the Internet, there are advanced services that generally cannot be provided for in a feasible way, such as high-definition video communications, assured low delay service, demanding machine-type communication, business critical cloud Internet services, as well as advanced live event content delivery services.

One of the inherent shortcomings or dilemmas is related to buffering and applications in need of low delay. Larger buffers are good for increasing statistical multiplexing gain and high link utilization, while small buffers are needed to support low-delay services.

The efficiency of traffic routing varies as IP packets on the Internet are routed according to their destinations only, and the hot-potato principle is often used (selecting the nearest exit point). No other routing policies across network domains are generally supported, such as according to the delay or resilience requirements of the flows. Moreover, ISPs are treating traffic volume in peering and transit relationships as a way to measure market power. This may also lead to inefficient or conflicting routing policies. Note also that resilience protection according to the criticality of the traffic is not possible.

What customers pay is often disconnected from, or has little correspondence with, the quantity of resources used across the Internet. Light users are often subsidizing the heavy users. These issues get worse with higher access speeds, as the ratio of access link capacity over core link capacity is decreasing (an indication of burstiness potential). As a result, content providers or end-users have limited incentives to use bandwidth efficiently.

Many applications need a certain level of throughput and delay performance in order to function according to users or application demand. Such inelastic traffic should be supported by some kind of admission control and resource management at critical locations, to assure satisfactory performance and predictable QoE for these applications, while also ensuring sufficient resources for the elastic traffic. The current Internet does not support such a mechanism.

Moreover, there are some fundamental challenges in overall traffic control. The foundation of the current Internet is based on the IP / Transport Control Protocol (TCP) of end-points (hosts / servers / devices), which uses packet loss as a real-time signal of congestion and adapts the traffic load pushed onto the Internet based on this. In this sense, a reasonable level of packet loss is inherently part of a well-behaving open Internet. On the other hand, the User Datagram Protocol (UDP) based traffic does not have such a back-off mechanism. In principle, one can argue that this traffic type is given higher priority. The assumption is that this will not create a problem as long as the UDP volume is relatively moderate. However, these assumptions lead to several concerns that are debated and challenged in several ways. For instance, there are observations that competing TCP variants are developed that do not back-off in the way that standard TCP should do. Hence, these applications try to grab more resources off the Internet at the sacrifice of other applications (and their users).

Energy efficiency in all network segments is a further concern. The current Internet is far from behaving according to a certain degree of proportionality between energy consumption and traffic load. There have been concerns that the energy per carried bit might become a true bottleneck in the growth of the whole Internet ecosystem.

Taking a closer look at the business aspects, some concerns are raised as well. The current Internet traffic exchange business model (IP peering, IP transit) has weaknesses or several issues in alignment of incentives among the players. There are often peering / transit disputes and arguments about who will pick up the bill of investing into new capacities, for instance in reducing congested exchange points or (long-haul) links. We see examples of inter-carrier long-haul and peering point excessive congestion. However, as noted above, some congestion and packet-loss is acceptable and is actually considered as good and an inherent property of a well-functioning best-effort Internet.

Moreover, it is not possible to have the initiating party paying for two-way traffic between the communicating end-points. That is, there is no business model and technical support for end-to-end money flows across the actors of the Internet value chain. As one example, it is not possible to have an SME in a developed country paying for an on-demand two-way AQ connectivity to enable a Telepresence session with a partner SME in a less-developed country.

2) Regulation and Net neutrality

As already mentioned above (see Section IV) the uncertainties around net neutrality is perhaps among the greatest hurdles, and a reason why the changes envisaged in this whitepaper have not happened so far. As part of this uncertainty are even issues on publicly acceptable terminology, which can become a stumbling block in itself. Is it even appropriate to speak of the (future, to be established) “Open Multi-Service Internet”, or will a substantial public opinion object and say that the “Internet” is what it is today and the meaning of this term should not change?

That is, the suggested and anticipated evolved Internet or internetworking approach will imply a multi-service Internet carrying multiple traffic types, both best-effort (Basic Quality, BQ) Internet access service traffic and multiple traffic types that are needed by various other services and their service modes. In any case, all the traffic of the future network and digital services is anticipated to be carried end-to-end over shared physical resources to ensure efficient network resource utilization and reduced costs for meeting all customer’s needs. Even today, these shared resources carry the IP traffic of diverse needs in a balanced and managed way: Traffic coming from a multitude of services, such as the Internet, the legacy voice service, as well as from business specific VPNs or various private managed networks. However, the traffic of these other specific services originates as part of dedicated specialized and rather costly silo’ed and often more inflexible solutions. There is no general approach to broadband-based differentiated and assured quality connectivity from any end-point to any end-point on the Internet.

However, the regulatory environments are changing. On November 25, 2015, the final regulation by the European Parliament and the Council on the Open Internet was published. This document gives a final answer after the long uncertainty on the topic.

The document states that:

“Providers of electronic communications to the public, including providers of Internet access services, and providers of content, applications and services shall be free to offer services other than Internet access services which are optimised for specific content, applications or services, or a combination thereof, where the optimisation is necessary in order to meet requirements of the content, applications or services for a specific level of quality.

Providers of electronic communications to the public, including providers of Internet access services, may offer or facilitate such services only if the network capacity is sufficient to provide them in addition to any Internet access services provided. Such services shall not be usable or offered as a replacement for Internet access services, and shall not be to the detriment of the availability or general quality of Internet access services for end-users."

Furthermore, the Treaty on the Functioning of the European Union states: "a regulation shall have general application. It shall be binding in its entirety and directly applicable in all Member States."

This evolution of regulation (which will be followed-up by BEREC[8] guidelines), while protecting the positive properties of the current best-effort Internet access services, will give the possibility for market players and consumers to negotiate on some specific characteristics of the services when a particular content, applications or services need reasonable traffic management, efficient and flexible use of network resources. We believe that this regulation is a positive step and that such possibilities will open up for new innovations and services, as discussed in this whitepaper. We hope that the envisaged approaches presented in the whitepaper towards key solution elements can help the community to make progress in this direction.

3) Product areas benefitting from Value Added Connectivity Services (VACS)

The following subsections provide an indication and some suggestions on how product specific ecosystems or verticals might benefit from VACS.

Note that we are not intending to be complete, but we present several examples that illustrate how new product areas can be enabled in this new innovation space.

a) Rich and Unified Communication Services

Nowadays, different OTT applications are providing several communications services demanded by users. Nevertheless, the users only use such services with a *good* Internet connection and typically rely on the voice calls provided by their phone operator for critical situations or whenever they need a reliable service. The IP Multimedia Subsystem (IMS) based services programs promoted by the GSM Association, such as the Rich Communication Services (RCS), were designed to provide interoperable services between carriers, guaranteeing (among other features) the QoS. In order to guarantee these features, operators need to control the access network.

Thanks to Service Level Agreements (SLAs), it would be possible to use any access network. For example, phones could use the local Wi-Fi access to connect to the operator-provided voice or video service. In such cases, voice and video are real-time critical services and have to be specially treated.

It is essential to remember that traditional voice communication services still account for around 60% of annual operator's service (around $600 billion in 2014).

[8] Body of European Regulators for Electronic Communications (BEREC)

But this is not only valid for operators. New or existing companies can provide innovative high quality multimedia services, avoiding the existing problems associated with OTT applications (e.g. unpredictability).

b) Live events and content delivery

Live events are followed by hundreds of thousands or even millions of users (for example, sports such as soccer games, tennis competitions or races). Broadcast technologies are usually adopted to make the video signal available at the same time to all the users. Globalization and the demand to have access to any kind of event (even small and medium events) from anywhere are pushing the creation of platforms that provide access to live events over the Internet using unicast connections. Such platforms are stressing the existing network to its limits.

Users need a network that satisfies this growing demand to access live contents, so the solution has been to include special elements such as caching systems at edge nodes (for delayed viewing) and multicast support. It also requires mechanisms supported by SLAs in order to guarantee that high quality streams reach those network elements and the final user in a proper and predictable way.

According to Cisco, this kind of traffic will increase by 33% between 2014 and 2019, with the potential to accommodate the needs of all the viewers of live events.

This opens a new opportunity to develop advanced services around live content delivery. For example, interactive events or second screen enhanced applications.

c) Business Cloud Services

Companies are increasingly moving their computing needs to the cloud, leading to an explosion in traffic.

Users want to experience cloud services, at anytime, anywhere and on any device. Nevertheless, not all the applications have the same requirements. For example, the upload of an attachment for an email can accept a delay, but a collaborative text editor requires the text to be updated almost immediately, or the experience for the users will be quite poor.

According to Cisco, the amount of global traffic crossing the Internet and IP WAN networks is projected to reach 2.0 ZB per year by 2019 (it was less than 700 EB in 2014). Cloud providers should be able to differentiate themselves by introducing mechanisms that ensure the QoS according to application type and customer choice. Thanks to those mechanisms companies will be able to move also advanced applications to the cloud, services that will not work today (e.g. advanced editing tools for multimedia, CAD tools, etc.).

d) "Tactile Internet" and Advanced Digital Consumer Services

The "tactile Internet" is a concept that describes how communication networks should be improved to be adapted to the human reaction time and perception, making it possible to create realistic telepresence systems or control remote robots that perform critical operations. The tactile Internet will require an extremely high reliability and low latency.

Online gaming is not as advanced as tactile Internet applications, but is a popular service with similar, although less strict, requirements. One of the QoE metrics for online games is game lag. A high game lag has a negative effect on consumer experience, and main causes are the computational and network constrains of the execution environment of the game client and game server. The constant increase - and falling cost - of computing power and network bandwidth have allowed Massive Multiplayer Online (MMO) games to acquire a wide consumer base. However, the situation is far from perfect, because recurring network latency spikes create frustrating moments for gamers. Latency varies depending on a number of factors, such as the physical distance between the end-systems. A longer distance means additional transmission length and possibly more routers involved (each incurring some packet processing delay), which results in higher latency. For an Internet connection, transmission length, medium and routing complexity can be very pronounced. Aiming to reduce latency and to provide a high QoE, the cloud gaming service OnLive (as one example) has developed a solution by establishing peering relationships with multiple Tier 1 network ISPs and choosing an optimal route between server and user.

MMO generated $24.4 billion of revenue in 2014, with an expected average annual growth of 7.9% until 2017, according to a market report by ad2games and newzoo.

A specialized services Internet can reduce the cost of such approaches and make them available to a wider range of MMO service providers and their users.

e) Internet of Things and M2M

The Internet of Things (IoT) paradigm describes how almost any physical device will be connected and will exchange data.

It is hard to imagine such amount of devices connected, but it is obvious that the nature and requirements of the traffic exchanged will depend on the particular application or use case. IoT use cases are extremely varied and so are their requirements. Major categories include smart home, smart car, health monitoring, emergency response, and production management. Smart home and smart car use consumer level bandwidth, disproportionate to their traffic needs. The same applies to health monitoring applications outside of hospitals. The latency issue that can arise from such application is almost completely offset by the extreme bandwidth overprovisioning that occurs in this situation. In the case of health applications in hospitals, the payloads can be much larger and latency requirements much more stringent as well. M2M communication systems usually exchange information and control messages, which may be critical to maintain the correct operation of the controlled devices. For that reason, nowadays many systems have to use dedicated links, but this could change with a network that ensures quality at an agreed level across the network domains.

In emergency response applications, the IoT devices must collect data about the emergency situation and transmit it to a crisis management centre, where it can be assessed by specialists (e.g. doctors in case of a serious car accident). The specialist directs the emergency operatives to the best handling of the situation. Such application can only rely on wireless communication. Worst case scenarios about the wireless communication must be considered. This leads implementations, such as 6-inaction by the University of Ljubljana, to include a gateway in the information system design. This is deployed to the crisis situation and transmits data collected by the IoT devices to the crisis management centre through a satellite connection. Offering AQ service level in an Internet connectivity approach can enable the deployment of emergency crew solutions without a gateway, significantly reducing the cost of creating and deploying this kind of solution. The use of admission control policies and dynamic traffic management technologies is needed to make such a service offering possible and cost effective. The resulting situation would be similar to the high priority that ambulances and police cars enjoy on public roads when they are assigned to an emergency.

The possibilities for new business opportunities are next to limitless, but clearly a technology that guarantees the right service levels for each type of IoT application will be the key ingredient for success. For example, critical communications will require a provider that ensures the quality from end to end. Such critical communications can be exemplified in hazardous “Industry 4.0” work environments where usage of IoT solutions is very important in order to prevent unpredictable operation failures and work accidents. Different kinds of IoT sensors, such as wearable devices, that capture data from work environment and vital parameters of work safety can be monitored to extract maximum amount of information from the hazardous production sites. The managed and assured information flows from such sensors and devices can enable numerous new applications and solutions in the IoT and “Industry 4.0” fields.

Cisco expects 50 billion “things” to be connected to the Internet by 2020, generating 500 ZB (zettabytes) of traffic. According to Gartner, IoT will generate $235 billion in sales in 2016.

4) Dynamic traffic management and efficiency gains

In this section, we take a closer look at the network efficiency potentials of traffic differentiation as compared with just having only one “best-effort” traffic mode. We do not go into details on the proposed four traffic modes and the QoS mechanisms and the specific traffic classes that may be used in order to realize traffic and service differentiation. We only notice that the QoS mechanisms will vary from one network segment to another and that the motivation should be to facilitate simple but sufficient mechanisms, whether the QoS solution and scope is designed for the network access segment or end-to-end. This section will also discuss the important roles of admission control and network performance monitoring, respectively.

a) Example case: Conference hotel

In Section 4 c) below a numerical example can be found of what we have called the “Hotel Case”. This case is chosen as one that is quite recognizable and represents potentials that are relevant in many settings beyond this particular example, ranging from home and SME local networks to mobile cellular networks and even to satellite network segments. The challenge is to consider the dynamically varying demand from a variety of users, their applications and corresponding service types and how the demand can be accommodated in a more resource efficient way. The core issue boils down to the following: How can the hotel network owner create the highest utility for the hotel’s guests from the resources she has invested in, noticing that her guests have quite varying needs, intentions, expectations and willingness to pay, with all these demands dynamically changing from minute to minute?

The goal and the challenge is to facilitate connectivity, not only to the elastic flows that can all be aggregated into a traffic bundle and handled by the BQ mode, but to provide efficient support to the inelastic flows as well, those that need to be supported by the IQ or even the AQ mode. When considering utility for the users and how this can be quantified, the notion of QoE is important. QoE can be quantified in terms of Mean Opinion Score (MOS), where a score of 5 indicates ***excellent*** experience while a MOS of 1 is ***bad*** experience or useless service. A score of 2 indicates ***poor***, a score of 3 indicates a ***fair*** experience, while a score of 4 maps to ***good*** experience.

In Figure 5 below, one can find illustrative MOS curves for the different services considered for our example case. In general, the MOS will depend on many factors where the measurable network performance metric (throughput, packet loss, delay, delay variation) are all important, although to a varying degree depending on service type. In general, the inelastic services are characterized by the rather steep shape of the MOS curves, whereas the elastic services are characterized by a more gradual slope and are more flexible in terms of relating network performance to QoE. By recognizing these fundamental properties of the different service types, the dynamic traffic management mechanisms can take advantage of these properties to achieve a higher network utility. See Appendix Section 4 c) for details.

The baseline case is a well-dimensioned network to accommodate the general anticipated demand with a good or excellent MOS outcome (Scenario 1). This allows connectivity for premium video and premium streaming, which represents inelastic traffic flows suitable to be handled by the AQ and IQ connectivity modes and to be delivered with predictable QoE. In addition, the goal is to accommodate the elastic basic browsing and basic streaming services with a similar good or excellent MOS. In the second scenario, we foresee a more or less unexpected growth in demand. First, we ask how many more premium video communication flows can be accommodated while still maintaining an acceptable MOS for all services (see scenario S2-A). The calculations show an increase from 20 to 52 premium video flows and a total increase of flows from 150 to 202 flows.

In scenario S2-B, on the other hand, we rather ask how many more browsing or basic streaming flows we can accommodate while still supporting an acceptable level for the MOS. The calculations show an increase of the “browsing and basic streaming” flows from 100 to 240. Note that this way of adding flow bandwidths (throughput) is in general not fully accurate. The MOS curves are in practice more complex and not only dependent on throughput. For instance, the slope of the MOS curve for elastic flows will also depend on the user context and might also be steeper. Moreover, the induced packet rate variation, buffering and multiplexing is a complex process not taken into account in these simplified calculations. However, the principles and basic approach to dynamic traffic management are still valid and the calculation results do provide an indication of the potentials, even if the particular calculations are on the optimistic side.

A more complex modelling and calculation has also been performed to consider much more realistic short term traffic variations of the flows, which implies that it is not possible to load a link to 100%, since this will lead to unacceptable delays and packet loss. These variations are handled by the buffers in routers, which do induce delay and delay variations (jitter). These more advanced calculations are based on strict requirements on delay and jitter for the premium video and premium streaming, while requirements can be much more relaxed for the elastic flows. The calculations show a total capacity need of 233.6 Mbit/s if no differentiation is supported, while the total capacity need can be reduced to 207 Mbit/s if differentiation is supported. In conclusion, the dynamic traffic management scheme must consider both buffer queuing and scheduling of the different traffic classes and flow (service) admission control in order to control the load and manage the overall utility and QoE in an optimal way, both from the user and the network operator perspectives.

b) Admission control and network performance monitoring

From the above example it becomes quite clear that it is a key necessity to maintain the network performance for the BQ traffic mode at an acceptable level, while the network performance of the BG mode is not so critical. Hence, the network provider must make sure that the premium traffic modes (AQ and IQ) do not get a too large share of the overall capacity, so there are sufficient resources left for the BQ traffic mode. This can be achieved by admission control of services using the AQ and IQ modes (i.e. in our case, premium video communication and premium streaming services).

Note also that admission control is a mechanism working at the flow level. Admission control is also part of the user (application) to network interaction (signalling), so that the user can be informed whether his on-demand service request can be accommodated or not. This is just as we know it from the “green button” used for making traditional voice calls on our mobile smartphone, which will give us a “busy tone” if we are not granted the connectivity needed for the voice service. The topic of how the user in general is informed in real-time about the anticipated service level of various typical services is the subject of Section 5 below.

In relation to maintaining the overall and general service level of the BQ mode, a relevant factor is the topic and trend of measurements and monitoring of applications and connectivity performance of the open Internet. Often, such measurements are performed by international research projects or businesses specializing in overviewing the performance of the Internet. We anticipate that these 3rd party actors will provide timely and up-to-date information on various network segments or geographic parts of the Internet, such as the Internet backbone and the mobile broadband part of the Internet (see e.g. the EU project MONROE[9]). In this way, these 3rd party actors (hopefully, neutral and non-biased actors) will help ensuring transparency of the general quality level of the BQ traffic mode facilitated by the open Internet.

So, consider now that one network operator attempts to push more traffic onto the (premium) AQ and IQ modes, by keeping the general performance of the BQ rather low. Another competing network operator has then the chance of offering a better BQ mode performance than the first operator and in this way attract more customers. In general, it is expected that the BQ (and the BG) mode represent the most cost efficient way of transporting IP packets, where the network performance requirements are not that strict (supporting applications where the QoE is not that critical), just as the open Internet already today is very cost efficient for such IP transport. It is also interesting to note from various studies that it is the good mix of traffic volumes across the BG, BQ, IQ and AQ modes that is the more resource optimal way of transporting IP packets, while generating the highest utility for the customers.

[9] Measuring Mobile Broadband Networks in Europe, MONROE – https://www.monroe-project.eu/

The realization of this network efficiency is in particular important in the network segments where one finds the bottleneck. Often the bottlenecks today are in the edge, the customer premises broadband gateway, and/or in the access part of the network. When creating value by differentiated and dynamic traffic management in the customer network or in the network edge, the resulting packet handling should not be tampered with throughout the internetworking path, but rather carried end-to-end by differentiated and value added connectivity services across interconnected networks that are able to realize such multi-service IP packet transport.

As already noted above, some moderate congestion and varying from time to time is to be expected for the BQ mode as facilitated by the current Internet. This congestion and packet loss of the BQ mode is part of the traffic regulation regime of a well-functioning open Internet. It appears as a reasonable guess that over time we can anticipate a greater transparency on the performance of the Internet, including its backbone parts. This will then include general performance targets on what is the general and acceptable level of packet loss and delay performance of various Internet segments.

c) Conference Hotel – Case description

As an example of how to perform dynamic traffic management, and what network efficiency gains can be achieved by connectivity differentiation, we provide a simplistic example with numerical assumptions and quantified results. The case anticipates a conference hotel offering connectivity services to various applications. From these assumptions we dimension the capacity needed and manage the traffic and connectivity services accordingly. More on motivation and context for this case can be found in Appendix Section 4 a) above.

We assume that incoming traffic is dominant (but the methods will apply for both directions). The hotel supports the four traffic modes and the following connectivity services, although in this particular case the BG traffic mode is not further considered:

- Premium Video Communication (PV), supported by the AQ mode
- Premium Streaming (PS), supported by the AQ or IQ mode
- Browsing and Basic Streaming (BR+BS), supported by the BQ mode

Moreover, we assume that the mean rates offered for each connectivity service are linked to their MOS curves as shown in Figure 5 below. The actual shape of the MOS curves will vary from service type to service type, but network performance parameters other than mean rates, like packet loss, delay and delay variation, may have significance. The inelastic traffic classes PV and PS will have a typical steep function shape describing the fact that, if the rate is below a certain threshold, the QoE will be zero (as measured by the MOS scale), while the QoE will increase rapidly when the bitrate is above that threshold. For elastic services, however, the corresponding MOS curves will be smoother, and satisfactory utility can be provided even if the bitrate is substantially lower than that for when the excellent MOS is achieved, e.g. as given by the arrows in Figure 5 below.

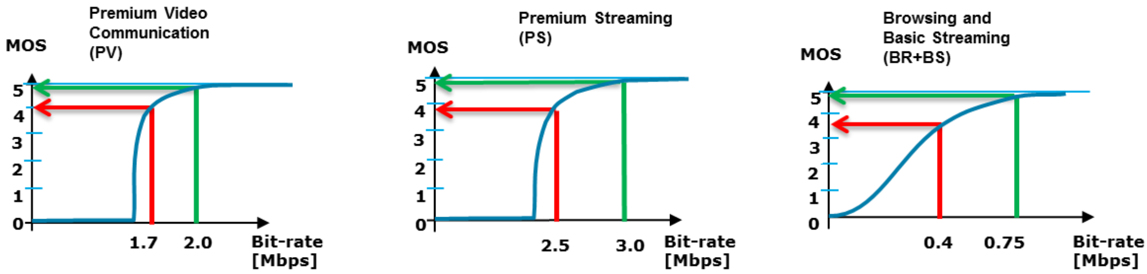

Figure 5. A sketch of MOS curves for the traffic classes PV, PS and BR+BS

On the basis of the examples shown in the figure above, we may find the rates that correspond to a particular quality measure (MOS), e.g.

- Excellent; MOS close to 5
- Acceptable or Good; MOS equal to 4 for PV, 3.8 for PS and equal to 3.5 for BR and BS

In Table 1, mean rates for the different traffic classes based on the MOS curves in Figure 5 are given for Excellent and Acceptable or Good QoE, respectively. It should be emphasized that these values are just examples which reflect potential services. Based on these examples, we discuss and point out the dynamic traffic management methods and principles in simplified terms.

Table 1. Mean rates (in Mbps) for the traffic classes PV, PS, BR+BS with Good and Acceptable quality

	Premium Video	Premium Streaming	Browsing and Basic Streaming
Excellent (MOS=5) ➔ [Mbps]	2.0	3.0	0.75
Acceptable or Good (MOS=4 for PV 3,8 for PS and MOS=3,5 for BR+BS ➔ [Mbits]	1.7	2.5	0.40

The demand may now be specified by the numbers of flows in each group; (N_{PV}, N_{PS}, N_{BR+BS}), and the corresponding mean rates; (r_{PV}, r_{PS}, r_{BR+BS}), which give the capacity required or bandwidth demand B as:

$$B = N_{PV} r_{PV} + N_{PS} r_{PS} + N_{BR+BS} r_{BR+PS}$$

Hence, if the anticipated demand is taken as given in advance and the rates are also known, this equation determines the needed capacity B. Or, on the other hand, if the (maximum) capacity limit B is given, the equation will give the numbers of flows from the different groups so that the total demand never exceeds the capacity limit B.

Below, we briefly discuss some example scenarios to illustrate the use of the bandwidth demand equation above to determine rates and number of flows under various targeted MOS constraints.

i. *Scenario S1: Dimensioning case with excellent MOS*

The first scenario reflects the case of all users having excellent MOS, e.g. offered rates are given by the first row in Table 1 above. We consider the case of total demand from 150 flows as given by the numbers below:

- Premium Video (PV): 20 flows
- Premium Streaming (PS): 30 flows
- Mix of Browsing (BR) and Basic Streaming (BS): 100 flows

By demanding excellent MOS, the total rate is found to be:
Total rate= (20x2.0+30x3.0+100x0.75) Mbit/s=205 Mbit/s, which we take as the basis of the total rate this scenario is dimensioned for. Observe that all combinations of numbers of flows satisfying

$$2N_{PV} + 3N_{PS} + 0.75N_{BR+BS} \leq 205$$

will be feasible for engineering the traffic with excellent MOS. If, for instance, the number of inelastic flows is increased by 50% (i.e. 30 PV flows, 45 PS flows), then the numbers of elastic users with excellent MOS is reduced to

$$N_{BR+BS} = (205\text{-}30\text{x}2\text{-}45\text{x}3)/0.75=13 \text{ flows.}$$

This clearly shows the need for admission control to protect the QoE of the elastic traffic. This can be done by limiting the number of inelastic flows and thereby reserving a minimum amount of capacity for the elastic traffic flows. This is further elaborated in Scenario 2, A and B below.

ii. *Scenario S2: Dimensioning case with Acceptable MOS*

S2-A: Handling increased demand for Premium Video Communication.
Suppose that the demand for Premium Video (PV) is increasing and that the numbers of PS and BR+BS users are as in scenario S1 above. By reducing the MOS requirements of all service types to acceptable or good, what is then the possible increase in the number of PV users (PV flows)?

By using the second row in Table 1 for the rates, we then find:

$$N_{PV} = (205\text{-}30\text{x}2.5\text{-}100\text{x}0.4)/1.7=52 \text{ flows.}$$

Hence, the number of PV flows is increased by 32 flows. This increase from 20 to 52 PV flows is quite substantial, and still acceptable or good MOS for all flows is maintained.

S2-B: Handling increased demand for Browsing and Basic Streaming (BR+BS).
Next, consider another case of increased demand, now by accommodating the increase in demand that comes from the elastic traffic. Again we use the rates that are assumed to accommodate acceptable or good MOS. In this case we find:

$$N_{BR+BS} = (205\text{-}20\text{x}1.7\text{-}30\text{x}2.5)/0.4=240 \text{ flows}$$

This is an increase of 140 elastic flows, from 100 to 240 BR+BS flows. It is again quite substantial, and still acceptable or good MOS for all flows is maintained.

In summary, we note that the bandwidth demand and flow equation in this case takes the form:

$$1.7N_{PV} + 2.5N_{PS} + 0.4N_{BR+BS} \leq 205$$

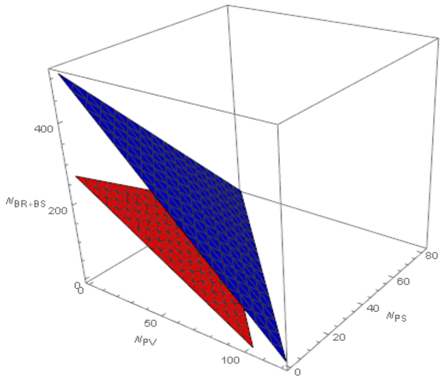

Figure 6. The regions of feasible flows (N_{PV}, N_{PS}, N_{BR+BS}): red excellent, blue acceptable MOS

In Figure 6, the regions of feasible number of flows; (N_{PV}, N_{PS}, N_{BR+BS}) are illustrated by the two planes. The red plane corresponds to excellent MOS, while the blue plane corresponds to acceptable MOS. Hence, the region in-between the two planes give the size of possible gains.

iii. Additional benefit by Traffic differentiation

When dimensioning actual capacity for a given traffic demand, one needs also to take into account the variation in the demand, as well as the variability of the rate of each flow, which implies that it is not possible to load a link to 100% since this will lead to unacceptable delays and packet losses. These variations are handled by the buffers in routers. Hence, the actual needed capacity must be larger than the demand in terms of total rate. Suppose that the demand is given as Scenario S1, with excellent MOS, and that we can tolerate some delay violation probability. The {delay bound, delay violation prob.} pair then fully defines the requirements for the traffic class, together with the mean rates given in Table 1. Table 2 below provides some example constraints assumed in the following calculations for the different service classes. Note that BR+BS are assumed to be much more tolerant to delay than what is assumed for PV and PS.

Table 2. Delay bounds and violation probabilities for the traffic classes PV, PS, BR+BS

	Premium Video	Premium Streaming	Browsing and Basic Streaming
Delay bound [ms]	1	5	50
Violation probability	0.01	0.02	0.02

By requiring that the

$$\Pr\{Delay > delay\ bound\} \leq Violation\ \ prob ,$$

we find the capacity needed when no differentiation is supported to be

C-non-diff=233.6 Mbps,

while the corresponding capacity with support for differentiation is found to be

C-diff=207 Mbps.

The actual calculation is based on assumed packet length of 800 bytes and Poisson input traffic model. Further, the differentiation is based on non-pre-emptive queueing among the four traffic classes. Hence, we conclude that the additional gain by traffic differentiation is quite substantial and should be utilised to obtain efficient network operation.

5) Service Level Awareness and User Interaction

a) Service Level Indicators in real-time

While the above discussion points at a set of quite ambitious future capabilities, we have to start with providing some ideas and a common “language” for the new concepts. Building from the four traffic or connectivity modes introduced in Section III a, Figure 7 below provides some examples of what may be a way of keeping the user informed, preferably in real-time. While the signal-strength and the 2G, 3G and 4G network indicators give us some information from which we can infer a likely application performance, we would like to move far beyond the QoE uncertainties that are characteristic of today’s Internet and applications use.

Figure 7 provides example illustrations of three different types of services. The first is a Live Event Service in situation A and situation B, respectively. In this example, the service is considered as a rather inelastic service, although this can depend on the way the application or product is marketed. In situation A the service is available for a small screen (S) sized device in the BQ mode, but is available for a medium screen (M) in AQ mode. At another point in time or location, the same service is available for screen size S and M in the BQ mode, while by the AQ mode both screen size L and XL are available. In situation A the screen size L and XL is not supported (service level unavailable).

The next type of service is a Browsing service, which is an elastic service (traffic flow). In situation C and D the illustrated service level indicator has a 5 level scale that can be considered like the MOS scale. The chosen illustrations are made to indicate that the service level, in both situation C and D, is rather uncertain. However, by selecting the IQ mode, the service level is increased by about one MOS level. How to indicate the level of uncertainty of the service level indicator is by itself a challenge.

The last service type considered is an upload service, as illustrated by situation E. Also in this example the indicated service level (in this case given in minutes) should be accompanied by an indication of certainty, but such an uncertainty indicator is not included here. The example illustration shows (see the colour codes) that the BG mode will likely result in a 20 minutes upload time, while the BQ mode will result in about 7 minutes upload time. By choosing the IQ mode the anticipated upload time is 4 minutes.

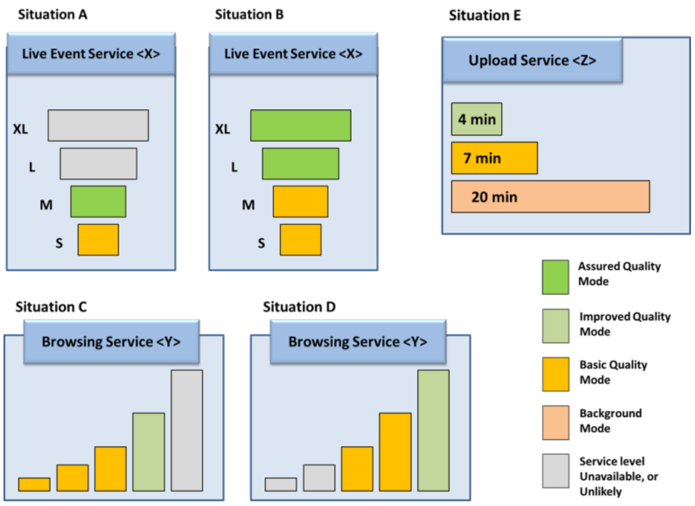

Figure 7 Service Level Indicators in real-time. Some example illustrations

b) User Interaction with the Network

The service level indicator examples above are all focused on how the user can be informed about the expected application or network performance, which can be mapped to the anticipated QoE. From this information, the user can adapt his expectations and also be informed on the available choices. While the BQ mode is the default mode and no particular action by the end-user is needed, the selection of the other connectivity modes may involve some active steps. As an example, consider how the smartphone can be configured through "my baseline connectivity settings". Here the user can see the default settings and modify them, and select the applications or application features that should use the BG mode or the IQ mode. However, usage of the AQ mode and also the IQ mode are anticipated to depend on more dynamic choices and selections through the application itself, depending on the urgency or importance of the user's task.

There are many issues that arise in relation to the above concepts and suggestions. However, this whitepaper cannot go into all of them. For instance, the true end-to-end experience will depend not only on the network performance of the network providers involved, but also on the end-user device itself, the potential customer premises local network, and possibly the datacentre network and servers hosting the application backend solution. Moreover, the way of interaction by the user with the application or the network can take several forms. In addition to the interaction with the applications locally on the device, the user can interact (via the local applications) with the application server side or with the network more directly. The upstream vs. the downstream traffic flow directions must also be considered carefully.

6) Managed and Assured Service Quality Traffic Exchange

In the backbone of the current open Internet, where the IP packets traversing the Internet are exchanged across the networks at Internet exchange Points (IXP), any marking of traffic type is discarded and the traffic is treated as one "best-effort" traffic class. At the same time, many of the

Internet services are processed by datacentres or data caching elements closer to the end-users. This "back-office" part also requires large amounts of IP traffic exchange and this traffic, exchanged between data centres, is typically carried over private managed quality IP networks.

The privately managed IP networks also accommodate business Virtual Private Network (VPN) services, and we observe that the overall growth of the privately managed network market is stronger than that of the Internet backbone[10]. Moreover, while caching and allocating Internet datacentres closer to end-users is a well-known approach to improving user experience for many applications, this approach is not feasible for more inelastic traffic flowing between or among end-points at the periphery of the Internet. That is, we still lack a multi-service backbone solution for an Open Multi-Service Internet that can efficiently accommodate all the traffic modes and connectivity classes, matching the needs of current and future applications.

Another interesting backbone solution, facilitating IP traffic and also supporting various media service interworking, is that of the GSM Association IP eXchange (aka. IPX, not to be confused with IXP, as explained above). This inter-carrier solution is designed to enable compatibility with and evolution of the Telco voice service. This backbone solution is a good candidate for carrying the flows of the AQ and IQ traffic modes. However, this approach might not be the most flexible and cost efficient solution.

Another complementary and seemingly attractive approach relates to the evolution of the current Internet backbone. The network elements of this backbone are already today equipped with multi-traffic-class support, so the CAPEX associated with enabling and evolving the current best-effort Internet backbone into a managed multi-service backbone is minimal. Some technical challenges are related to the consistent end-to-end matching of traffic class codes, as well as to the support for traffic monitoring and SLA assurance. The larger challenges, however, relate to the establishment of feasible business models on one hand and regulatory and policy concerns on the other hand. While the latter is in general a topic of this whitepaper, the business model challenge deserves a few more words in relation to such an open "Managed and Assured Service Quality (ASQ) Traffic Exchange" approach. (Cf. the notion of MQX in Section III – c, above)

The EU project ETICS (Economics and Technologies for Inter-Carrier Services) addressed these business model challenges as one of its key topics. The recommended approach is the so-called "Sending Party Network Pays" (SPNP) principle, where the carrier network sending the traffic (the sending party network) pays the receiving carrier network for transporting the traffic, as agreed by the SLA, to end-points of a given region. One important note here is that this principle is proposed to be applied among the backbone carrier networks, whereas the business model principle for the OSP should be considered from one business vertical to another and often be left for competitive agreements between the NSP and the OSP. Moreover, the ETICS project suggested that the SPNP principle for ASQ traffic exchange can, and should be, complemented with compatible retail level

[10] https://www.telegeography.com/products/commsupdate/articles/2014/04/23/new-global-network-builders-emerge/

business model and charging principles, as feasible for specific end-user services and applications. See the ETICS whitepaper[11] and deliverables for further details.

In this context, we observe that the traffic exchange for the BQ (and even BG) traffic modes will also require some quality management, although not in the same strict forms as foreseen for the IQ and AQ traffic. As mentioned above, some congestion is not a bad thing and should be anticipated as part of a well-functioning traffic regulation regime for efficient traffic filling of the link capacities. In summary, these new features should enable and support the more efficient and transparent transport of all multi-service Internet IP traffic in a reasonable and well-balanced approach, effectively sharing the same physical capacity.

7) Research and Innovation Challenges

It is foreseen that an evolved Internet approach can be triggered already in a near-term timeline, based on today's available technologies and already deployed network nodes. However, the advanced features of open multi-service networking outlined throughout this whitepaper clearly call for novel and challenging research directions. Such challenges need to be tackled as part of the bigger picture of upcoming techno-economic opportunities and challenges, driven by new related networking concepts such as cloudification, softwarization or virtualization that are important elements of the emerging 5G networks. In the following, we provide some elaboration on technology-oriented challenges along with a note on the relevance of standardization efforts in such directions.

a) Service level awareness and user experience in softwarized network infrastructures

Service level awareness indicators in real time are of paramount interest to improve and adapt service levels towards user satisfaction, while ensuring efficient use of network and energy resources. Such indicators would provide the means to track existing service levels and to identify and optimize which adjustments and novel procedures should be made.

User satisfaction is ultimately driven by perceived user experience. While network characteristics provide insights on the application performance, they cannot completely capture the user's experience. To accurately assess the application performance experienced by the users, also subjective evaluations are needed. Currently, very little is known about the general application performance of operational Mobile BroadBand (MBB) networks mainly due to: i) lack of a large-scale measurement campaign that investigates subjective application QoE in operational MBB networks and ii) lack of common knowledge regarding the mapping between network characteristics and subjective application QoE. At present, there are only a few closed-form relations between lower and higher layer QoS KPIs and perceived application QoE. It is e.g. not easy to predict, by BER and RTT measurements, if a streaming YouTube video will stall due to an empty playout buffer. An example for such a model is the well-known E-Model for voice calls (ITU-T G.107).

---

[11] https://www.ict-etics.eu/fileadmin/documents/news/ETICS_white_paper_final.pdf

Furthermore, lower layer QoS could have been collected at earlier time instances or other nodes than the end nodes of the session. Methods are required to determine what particular QoS degradation, at which points in the network, caused application QoE degradation. QoE objectives could then be defined according to the desired type of service level indicator. QoS levels should be selected according to the QoE objectives. For example, live multimedia applications with Assured Quality (AQ) will have strict delay and bandwidth requirements. For live media applications offered with IQ or BQ service level, some degradation of video quality or skipping some frames can be acceptable in order to keep up with real-time sources. However, in a video on-demand application with IQ or BQ service, the client (user) may prefer some amount of delay or pause due to buffering or outages, rather than degradation of video quality for real-time live high-quality video play-out.

Performance assessment could be carried out in real-time, leveraging on networking measures and not mandatorily requiring the collaboration of the end-user (who may be unaware). Key enablers are techniques based on heuristics, generalized traffic analysis, deep packet inspection and generation of QoE samples from networking statistics data fusion. This is a major challenge and an essential step to ensure service level aware operations of future networks.

As a consequence, mappings between QoE and QoS should be considered application dependent, but also time, space (throughout networks), content and context dependent. Such fine mapping granularity will help to better capture user demands that can be met by differentiated services. It will require a large-scale measurement campaign to investigate subjective application QoE in operational networks, and to research mappings between network characteristics and respective QoS measurements and subjective application QoE.

Moreover, user experience over operational virtualized network functions and infrastructures impose additional challenges for real-time monitoring and reaction. Underneath the concept of 'Service Level Awareness' there is the concept of 'Infrastructure Composition Awareness' and 'Dynamic Infrastructure Set up'. In fact, different operators can implement up to date network management solutions (Software Defined Networking (SDN), Network Functions Virtualization (NFV), etc.). As a consequence, different network composition will provide different performance and the situation can vary from day to day. For example, in SDN the complete set of resources of a network segment are visible to a management unit (controller). By enabling and standardizing messaging between controllers, network resource information of different network segments can be shared in some aggregated form. This enables event-driven real-time full network configurability for end-to-end services. In such a context, to engineer service level awareness in real time implies that the different network segments are really integrated at the management level and not only simply interconnected. Full network flexibility, including real-time protocol and architecture re-configurability, will be an irreplaceable value added to ensure the maximum level of QoS/QoE satisfaction.

b) Control and energy-efficient engineering of multi-class network information flows

Control and engineering of information flows need large data collections of measurements in the network to decide on suitable reactions to reach target performance, if necessary. To achieve real global coverage (from the ocean to the basement of a building, from the remote mountain peak to the narrow street of a big city surrounded by very tall skyscrapers) the network is composed of many segments (terrestrial fixed, terrestrial mobile, satellite, UAV), operated by several different operators and some of them not having an operator at all (NFC, Ad hoc networks). ITU provides high level methods on how to achieve end-to-end performance objectives by dividing the end-to-end path into segments (network domains) and defining targets for each segment. Labelling of packet flows, buffer management including dropping and prioritization and traffic shaping are used for that. The need of more sophisticated methods could require modifying and improving relevant protocols (from resource management up to application), as well as modifications of the architecture (including optimum segment selection to be implemented in real time).

The network softwarization fostered by SDN and NFV and the increased use of more general purpose IT hardware are likely to affect the energy consumption negatively. The management procedures should be able to deal with this too. The goal should be to set up a framework to manage different segments operated by different operators or even managed by the single final customer (for example the home or office local network or the NFC connection). The framework must have the capability and entitlement to manage up to and including the single customer modem, or to give orders to networks elements, with the aim to guarantee the target properties.

Multi-class network information flows also call for powerful technological means to control the performance of the differentiated information flows beyond network management. Network coding has the potential to offer such powerful means to jointly satisfy demand and reliability. Although theoretical network codes already exist with optimal performance and competitive complexity level, novel protocols across the different controllers of the network segments and virtual resources are needed to efficiently act on the reliability level of the several traffic modes. Coding across networks also has the potential to provide inherent means for the provision of security guarantees, while allowing full integration with SDN-based network design and virtualized physical resources.

The above challenges call for the definition of new ontologies of traffic logical mappings. Such ontologies would serve as a design environment for multi-class networking over the new architectural abstractions and logical concepts brought about by softwarization and virtualizations.

c) Need of standardization

Standardization bodies are currently defining the necessary steps towards the fundamental upcoming shift in networking. The foreseen integration of 5G networking, virtualization of communication and storage infrastructures and transport networks calls for fundamental advances in technology.

Efforts to define new protocols and architectures towards virtualization and softwarization are already taking place in IETF, IRTF and ETSI NFV, to name a few. While the current and future dominant role of IP networks is strongly reflected by current ITU standardization, there also exist active ETSI efforts in the same direction.

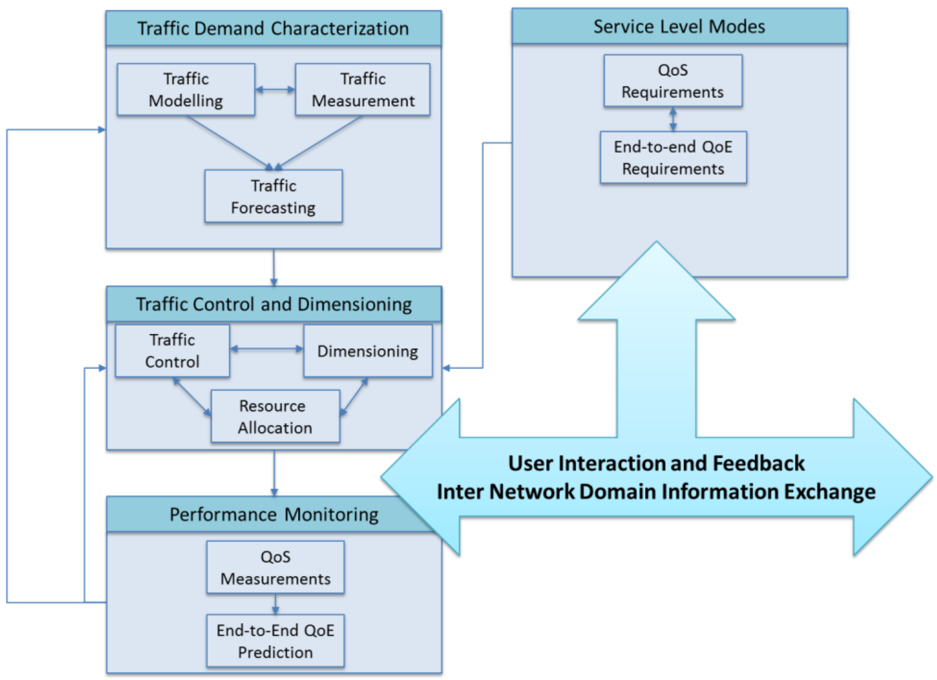

Figure 8 Functional elements of dynamic and customer centric traffic management

The processes and information flows described in the current standardization works relate to operational tasks internal to the network operator. They are summarized in Figure 8 above, which is inspired from a similar figure by ITU-T (for more details see ITU E.490.1[12].) In addition to these, Figure 8 also points to the needs of user interaction and feedback, as well as those related to exchange of information and supporting operations across networking domains. Topics in these regards have already been identified and discussed in the Appendix Sections 4, 5 and 6 above.

{ * *End of document* * }

[12] ITU, "Overview of Recommendations on Traffic Engineering", Technical Report, ITU E.490.1, 2003