;/

Call Identifier: SMART 2012/0046

Full title: Study on European Internet Traffic: Monitoring Tools and Analysis

Deliverable D7: Gap Analysis of Internet Traffic Measurement

Date of preparation: Draft version for review, March 20, 2015

List of participant organisations:

Role	Participant name	Participant short name	Country
Tenderer	Hebrew University of Jerusalem	HUJI	Israel
Supporting partner	Alcatel Lucent Bell N.V.	ALB	Belgium
Supporting partner	Martel GmbH	MAR	Switzerland
Supporting partner	Universidad Autónoma de Madrid	UAM	Spain
Supporting partner	Universite Pierre et Marie Curie – Paris 6	UPMC	France

Editor of this deliverable: Professor Scott Kirkpatrick

Organisation: Hebrew University of Jerusalem

e-mail: kirk@cs.huji.ac.il

Contents

1 Overview

The Internet is a vast and complex system, subject to a wide range of measurements, to meet the needs of end-users, operators, analysts and regulators. To structure our discussion of these measurements and our identification of the gaps that exist today, we first present a schema of the Internet as a whole. Then we proceed through that schema and sketch the gaps we have found between today's practice and the needs of tomorrow. We devote a chapter following to each area, where we go into more detail.

The illustration below serves as our guiding schema. On the left we see the end-users, on the right the Content Providers, and in between the Internet. This representation is intended to be broadly representative of the main features of the network, rather than comprehensive.

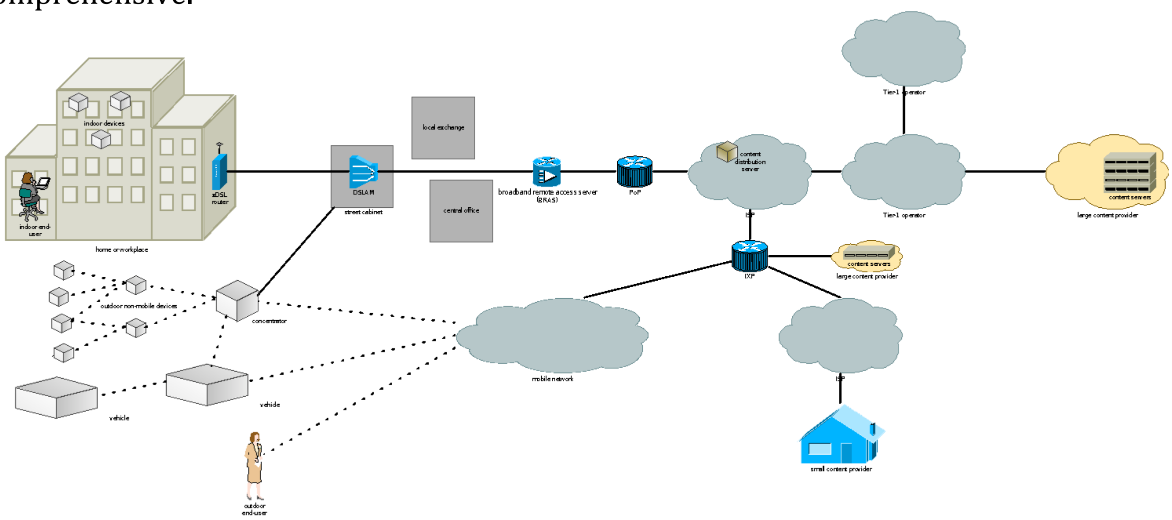


Figure 1: Guiding schema for the deliverable

End-users include people: a person in a home or office, a person moving about in the street. But they also include objects: objects indoors in the home, workplace or factory, and objects outdoors, including stationary sensing and signalling devices and mobile objects such as vehicles.

The Content Providers can be large datacentres in locations chosen for their inexpensive electricity, providing access through the backbone network of a tier-1 operator. They can also be smaller datacentres in locations chosen for proximity to end-users, in which case access might be provided through an Internet Exchange Point (IXP). Or they might consist in servers that are located directly in the premises of Internet service providers (ISPs), to get even closer to the end-users and to help the ISPs save on their upstream connectivity costs. There are smaller Content Providers, located in less privileged locations. Furthermore, Content Providers can also be individuals, or objects, communicating in a peer-to-peer fashion, in which case we locate them on the left hand side of the diagram.

Two main forms of access network are shown: fixed line and wireless. We represent fixed line access by an ADSL network, which could be copper or fibre. We could equally well have shown a cable network. And we show wireless access through a cellular telephone network, which has become the principal form of access to the Internet. We do not show other wireless technologies such as satellite phones.

Our examination of existing measurement approaches and the gaps to be bridged will cover the several different elements of this overall picture, as different issues arise in each element.

1.1 Fixed-line access networks

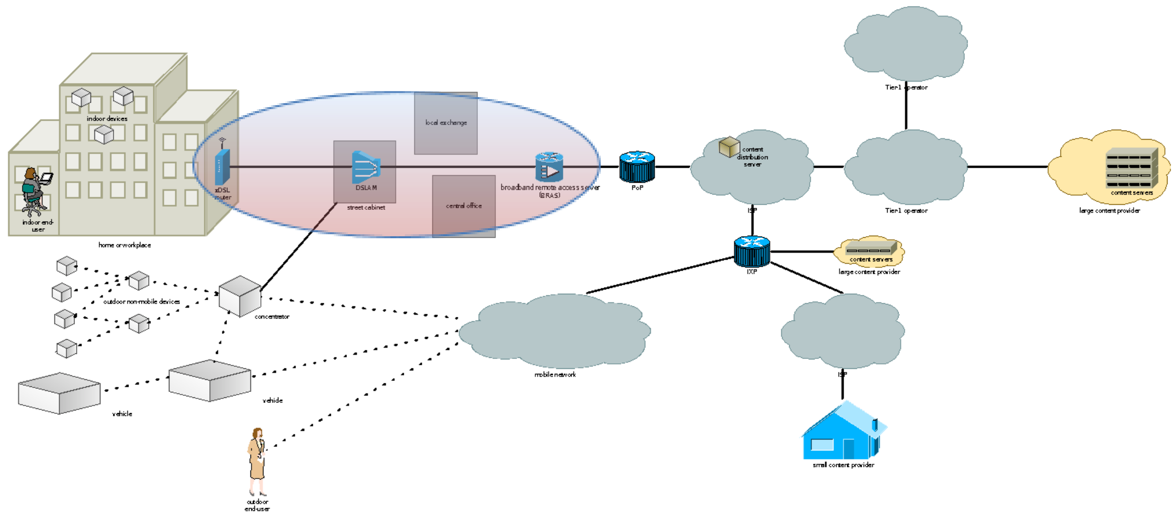


Figure 2: Diagrammatic representation of fixed-line networks

Much of the focus of regulators is on the performance of fixed-line access to the home. On the user side, they may be measuring from:

- a software agent on the user's computer
- a hardware box, such as a SamKnows box, located in the user's home
- an operator-provided xDSL or cable modem
- the street cabinet
- an operator-provided device that is located at an "ersatz" customer premise, specially created by the regulator and/or the operators that it regulates to stand in as a 'typical' customer premise

The location of the observation point at the user side should be as close to the point at which the signal enters the user's premises as possible. The ideal location would be to run measurement routines on the xDSL- or cable- modem itself.

The end-point for measurements on the operator's side might be:

- a DSLAM (Digital Subscriber Line Access Module), located in a street cabinet, local exchange or central office

- a BRAS (Broadband Remote Access Server)
- a point located somewhere else in the core of the operator's network
- at an IXP, through which the operator connects to the outside world
- the servers of major Content Providers, or other end-users

Different endpoints within the Internet serve different purposes. Operators will concentrate on monitoring the equipment they control, while regulators and end-users may be more interested in understanding the quality of the connection provided between the customers and popular Internet services, at least those supplied within the country or area for which the regulator is responsible. As always there are trade-offs between generality and the ability to certify observations and achieve prompt problem resolution.

1.2 Wireless access networks

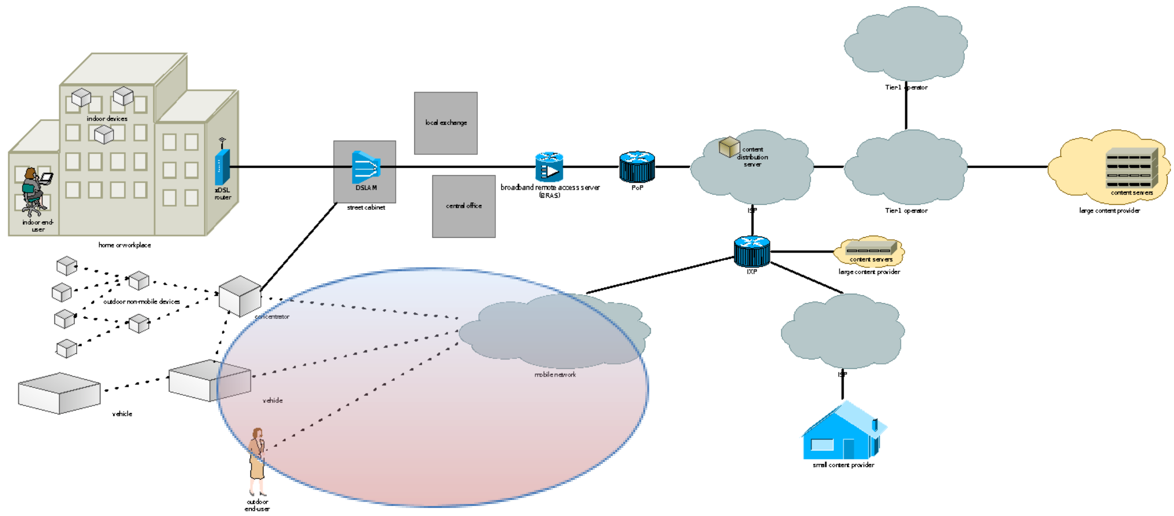


Figure 3: Diagrammatic representation of wireless access networks

With regard to cellular networks, regulators have largely concentrated on mapping signal strength in the outdoor environment. Much less has been done to look at indoor access to cellular wireless (though now that many people have given up their fixed-line phones, they are increasingly using their mobile phones from home, and there is considerable usage of mobile phones from offices, shops, and factories, as well as from underground, in cities that have metropolitan systems that offer cellular access). MONROE, a new EC project, places multi-carrier signal strength monitors in fixed outdoor locations to provide regulators and the public with continuous signal quality readings.

In addition to signal strength, regulators (and others) are interested in measuring network and application layer parameters that affect data traffic and thus access to Internet services. There is, increasingly, software for deployment on mobile devices to do this. Cellphones, in order to remain active participants in the carrier network, must communicate at regular intervals with the nearest cell tower. This provides an opportunity to gather measurement information from each active phone's location.

WeFi is an example of a company which is making use of this background "conversation" to add simple measurements of data network performance to the information which is obtained.

1.3 Inside the home network

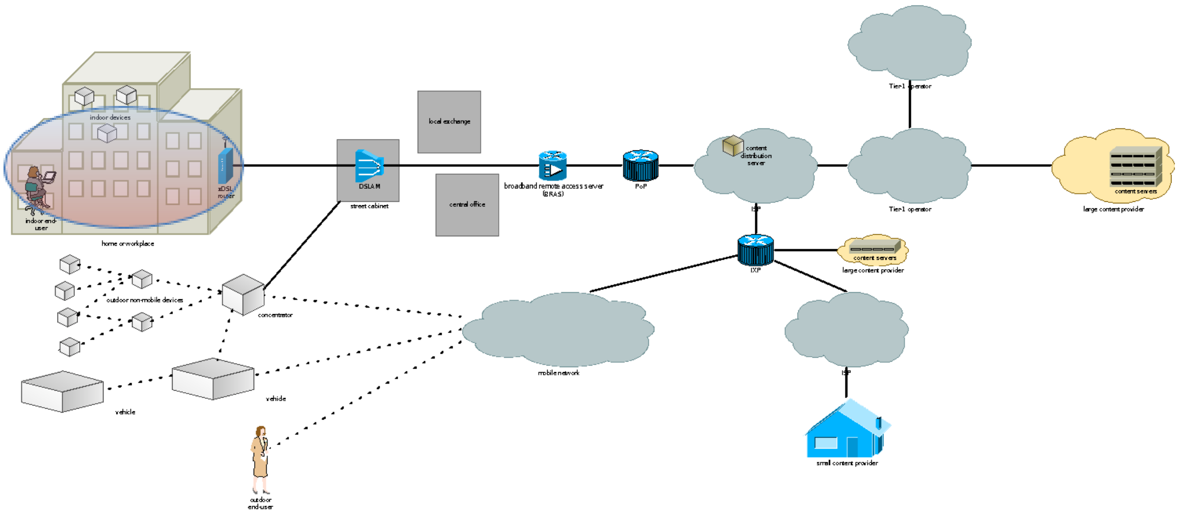


Figure 4: Diagrammatic representation of home networks

Internet service providers are painfully aware that the Quality of Service experienced by their customers depends upon factors outside their control. They are in the middle, with providers of content and upstream networks responsible for part of the overall experience, and with the vagaries of the home network responsible for another part. If a user has poor Wi-Fi connectivity and interference from a microwave oven, if a user is running applications that saturate the upstream link, if the user's own computer is running slowly because of its background activity, he or she will have a poor experience.

Operators do have the possibility to instrument part of the home network, as their xDSL modems (or cable modems) tend to double as routers and as Wi-Fi base stations, and they can also suggest that users run quality monitoring software on their computers, and so they, and researchers, are trying to get a better grasp on the impact of these factors on overall quality. In a subsequent chapter we review some of the active measurement tools now available in software and in hardware for measurements from the users' networks, and their limitations. Hardware based tools tend to be run at predetermined intervals, to provide uniform coverage. They must, as a rule, be careful to make only modest demands on the external bandwidth for which the user is paying, and which may have total usage caps applied. Simple software tools, such as OOKLA's SpeedTest or Glasnost (which checks to see if specific applications, such as VoIP or torrents, are being discriminated against) can be more elaborate, as they are initiated by the end-user. But as a result, they do not provide uniform coverage and do not by themselves monitor for problems.

1.4 The operators' upstream links

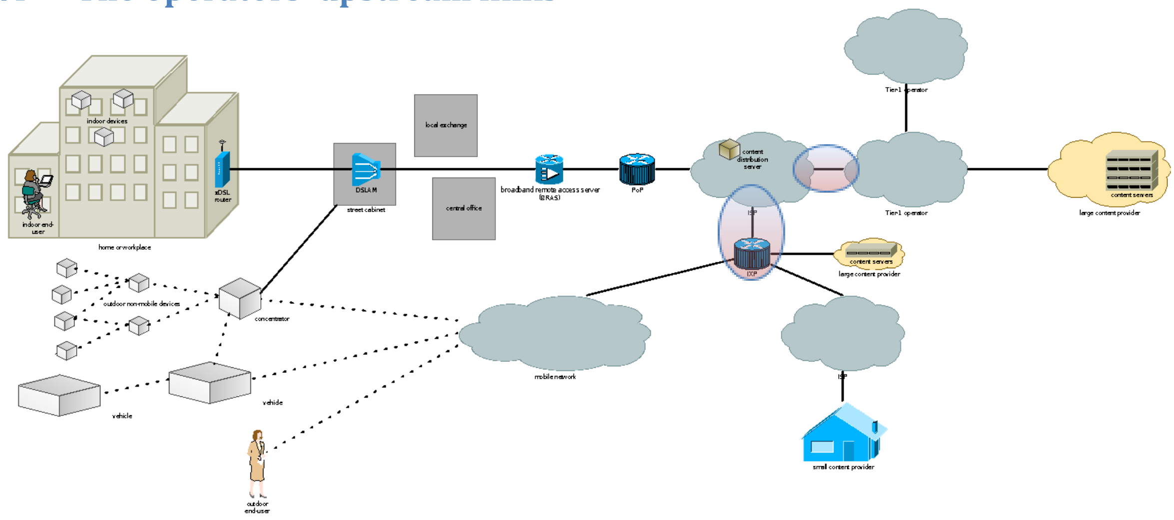


Figure 5: Diagrammatic representation of the upstream links

Much of network economics is driven by the cost of Internet Service Providers' upstream links and the business arrangements by which that cost is recovered. The cost of connectivity, paid to tier-2 and tier-1 operators, is one of the main costs for ISPs, and is the motivation behind the efforts of many operators to move out of the niche of serving as a tunnel between Content Providers and users, either by providing their own competing services or by charging for privileged access to users. These costs are at the heart of the Network Neutrality debate. One way that operators can reduce their upstream costs is through peering at IXPs. But we know little about the bandwidth of upstream links and about their impact on user experience. It is difficult to identify IXPs, their locations, their economics, and their impact. One goal is to measure performance details from which we can infer these business relationships. But who has “standing” to do so, and which of these details are relevant to regulators, which to users?

Content distributors, such as Akamai, minimise the costs of data transfer by caching the most popular information, or the parts of a services response that must instantly available, in many locations as close as possible to the end-users. This replication, and what may be new (but proprietary and thus unknown) business arrangements between the participants, has led to a sense that the Internet is becoming less strictly hierarchical, and “flatter” in its topology. Measuring and characterising this evolving topology is an ongoing research activity.

The quality of user experience is affected, often in invisible ways, by these upstream links. Famously, Iliad in France has throttled YouTube at this level, but much more takes place unseen. In the US, video suppliers Netflix and Amazon have had extended peering disputes with Verizon, Comcast and Cogent, middle-distance carriers which provide access bandwidth to large numbers of home and many smaller ISPs as well. A recent study[1] measured the differences in round trip times from a few observation points to the near and far end of links between Comcast and Cogent, finding clear evidence for daytime congestion.

[1] M. Luckie, A. Dhamdhere, B. Huffaker, Y. Hyun, KC Claffy and D. Clark, “Internet Interdomain Congestion” (IMC 2015 paper; presentation available on CAIDA website)

As regulatory requirements evolve in the US, how do we support them with measurement capabilities? Who is working for whom? Operators and content distributors can help, but will not if the results might reflect badly (or unfairly) on themselves. The FCC in the US has declared that it will in the future (and once litigation subsides) regulate data carriage as a public service under a restricted version of the Title II that covers public broadcasters. The all-important details of what aspects will be regulated and what characteristics and charges will be monitored have just become public, will undoubtedly be litigated, and are yet to be understood and absorbed by all who are affected.

1.5 Content Providers

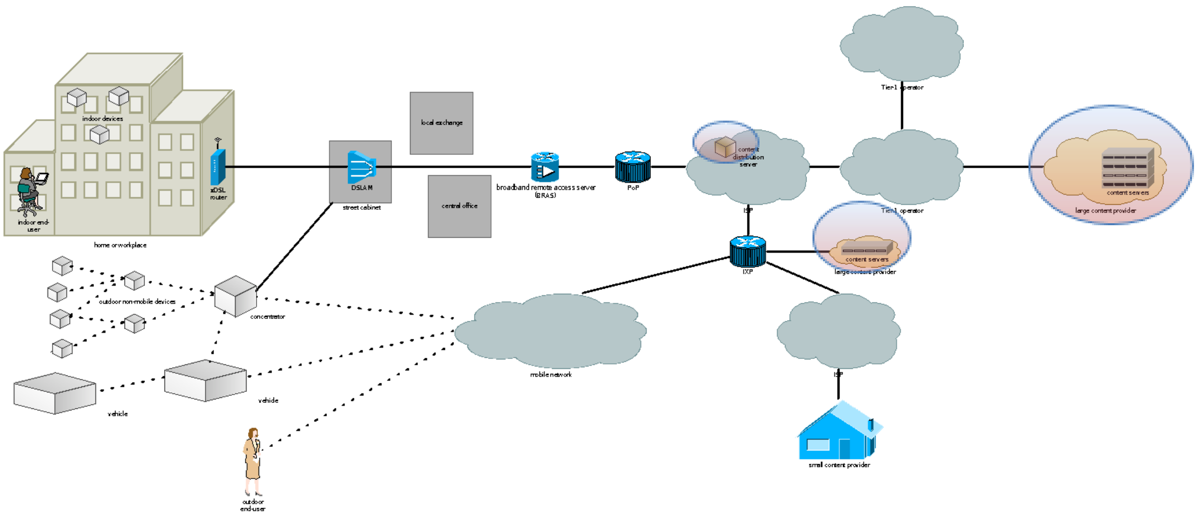


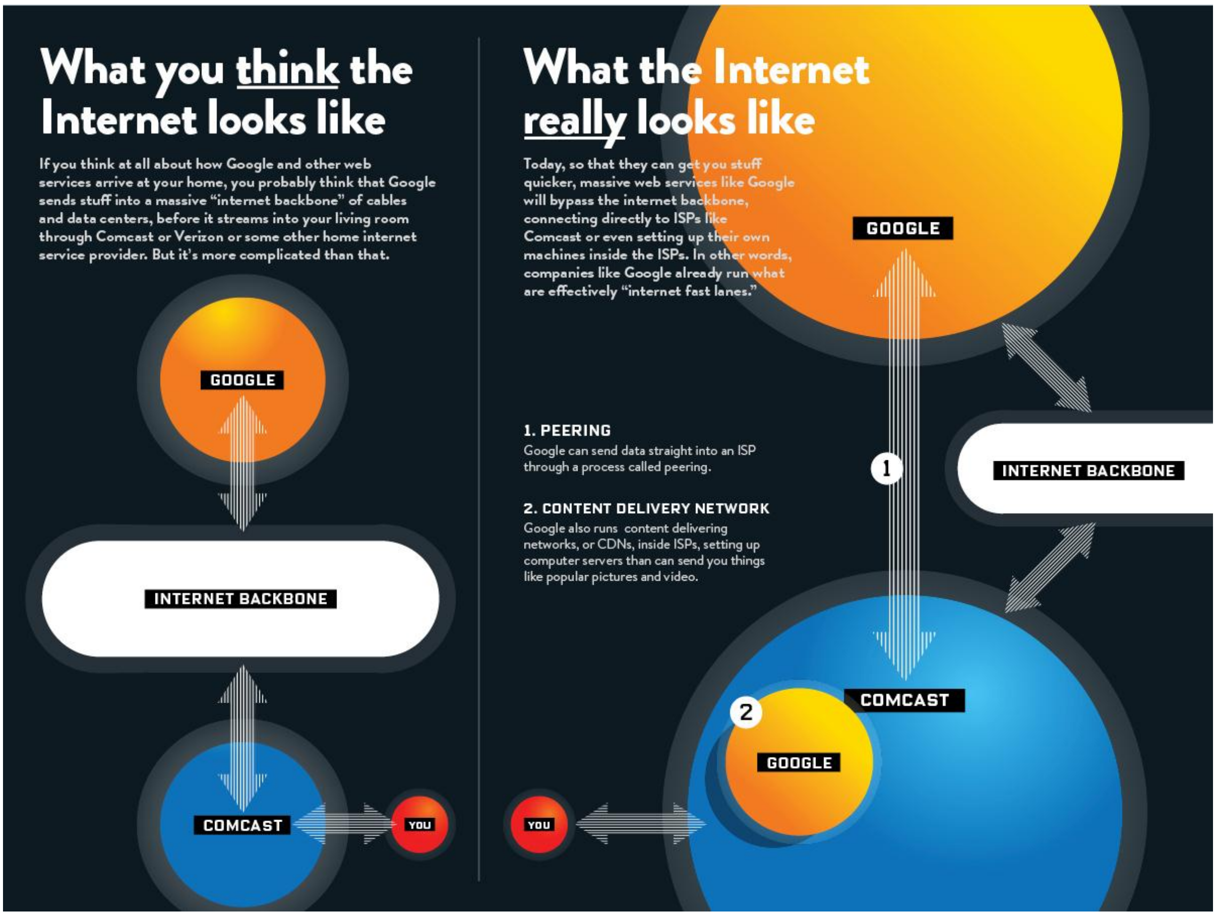
Figure 6: Diagrammatic representation of Content Provider networks

Much of content is served content, as opposed to peer-to-peer provided content. Also, much of this comes from large Content Providers. The proximity of this content to the user can have a tremendous impact on the quality of the user's experience.

Operators try to determine where the content is coming from and how to ensure that the user's experience is a good one. This becomes increasingly complex as the greatest fraction of music, news, and even video is being delivered to smart phones and other mobile devices around cities, in cars, and even in rural areas.

Operators have the option of hosting large Content Providers' servers, which can help them save on upstream bandwidth but which means that they are giving up significant control of their own networks. And they open themselves to Network Neutrality issues.

More and more, large replicated content is being offered through IP-level anycast, making it all that much harder to understand where the content is being served from and to evaluate overall quality of experience. In fact the hierarchical model of the Internet and its tiers is under great stress as new Internet businesses proliferate. The following picture conveys the seriousness and size of the challenge:



1.6 The Internet of Things

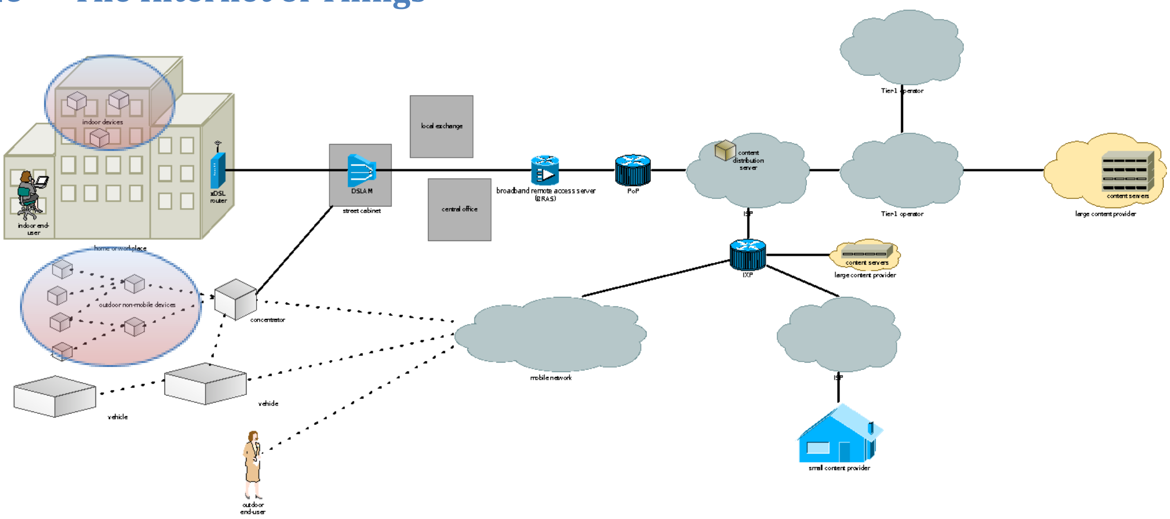


Figure 7: Diagrammatic representation of the Internet of Things

The Internet of Things (IoT) is attracting enormous attention. Satellite workshops have sprung up at every major networking conference, and research initiatives in IoT are offering substantial funding opportunities at European, North American, national and regional levels. In part, the attraction is due to the very large challenges of scale involved in trying to harness information gathered or required by billions of smart devices and the feeling that this is the next open frontier in information technology. Who can resist conference announcements that promise (as IoT 2014 did):

Pervasive connectivity, smart devices and demand for data mean that the Internet of Things is growing by leaps and bounds. Computing power is dropping in price, new sensors are being developed, and as people buy into Internet of Things technology, economies of scale lend themselves to the creation of ever more data-centric businesses. Instrumenting and connecting devices has massive potential to deliver value, but there is need for a coordinated effort for rolling out the next generation of self-reporting devices.

A second aspect bringing credibility and urgency to IoT is that there are well-developed niches or “silos” in which large numbers of inexpensive low-powered sensors, control devices, or simply unpowered echoing “radio bar codes” are already being put to effective use in applications such as logistics (barcodes), transport, manufacturing, smart cities, environmental monitoring, and aids to vehicle navigation — both manned and unmanned. Some of the silos support viable businesses which have already given rise to de facto standards for collecting and distributing the information that flows within them. The challenge of bringing these already vital areas together through a “narrow waist” like that of the middle layer of the Internet seems even greater than the challenge faced by the designers of what became TCP/IP in integrating the proprietary networks of the 1960s and 1970s. In addition, the “killer application” that will make this challenge a necessary step for the different areas of activity to join in solving is not obvious today.

Since much work on IoT today is still architectural, theoretical, and tested through presentation, simple prototypes and demos, the measurement tools and requirements are still local to each silo. It seems most likely that integration of tools and methods and the development of wider area communications will grow within each silo perhaps spreading to similar application areas, unless a conscious effort is found to discover areas where this can proceed more directly.

1.7 Measurement techniques

Measurement techniques can be either active (those which add traffic to be measured) or passive (those which monitor existing traffic). The former transmit packet trains to infer link parameters such as latency, jitter, bandwidth or loss, or detect routing paths by using the venerable (and not entirely accurate) traceroute tool. The latter extract a copy of the link traffic via a SPAN port (mirror) or a splitter. Then, the traffic is analyzed to seek for Quality of Service (QoS) statistics such as response time or percentage of requests served successfully.

The following table provides a taxonomy of the different measurement techniques available.

PARAMETER TO MEASURE	WHAT	WHO PROVIDES
Network level		
Packet loss	Measurement of the number of packets lost, most likely due to congestion or wireless link noise.	Routers through SNMP counters.
Latency	One-way or bidirectional delay between two endpoints.	Active measurement equipment or agents, which can be as basic as a probe with a ping.
Bandwidth	Available (or used) bandwidth in a given pipe.	Routers through SNMP counters or Netflow records.
Jitter	One-way or bidirectional jitter between two endpoints.	Active measurement equipment or agent, which can be as basic as a probe with a ping.
Transport level		
Throughput	Bandwidth consumed by a transport (TCP) session.	Routers (through Netflows) or specialised passive traffic probes.
Retransmissions	TCP retransmission, due to loss, jitter, or slow endpoints.	Specialised passive traffic probes.
Zero window announcements	Announcements from endpoints that suffer saturation.	Specialised passive traffic probes.
SYN/RST/FIN counters	TCP parameters that indicate connection start, end, reset, etc.	Specialised passive traffic probes, firewall, IPS/IDS systems.
Application		
HTTP/SMTP/TNS/DNS/DRDA/LDAP performance parameters	Different performance parameters from HTTP/SMTP/TNS/DRDA/LDAP protocols, for example HTTP response codes.	Specialised passive traffic probes, measurement agents, server logs.
Multimedia quality assessment parameters	Different performance parameters related to over-the-top services such as VoIP and Video over IP, for example rate in frames per second.	Specialised passive traffic probes and measurement agents.

We note that the communication network and services can be measured from many different angles that span from the physical layer to the application layer. In a measurement project, all these different parameters are necessary and, most importantly, the combination and correlation of them is of outmost importance.

For example, when the QoS is deemed poor in the response time the analyst has to dig into the data to find the root causes. Chances are that the packet loss in the network (for instance a wireless link) is high and the response time is also high due to TCP retransmissions and not because the application is slow.

Thus, the measurement data, which must be combined and correlated to be meaningful, must come to the analyst in a homogeneous format. However, one of the main gaps deleted is, precisely, the many different formats that can be found for the measurement data, very few of them standard. The situation is even worse when dealing with server and application logs, which are written in natural language in some cases.

Furthermore, there is also a lack of certification in the measurements. Namely, it becomes very challenging to determine if a given data is accurate or not. As a result, the conclusions from the analysis can be quite misleading, especially as the network speeds grows and finer time resolution is required. For example, suppose that a given company desires to verify if the operator fulfils a SLA of 2 Mbps in a given link. The company might install one of the many measurement instruments available in the market and measure 1.9 Mbps. Due to the fact that the tolerance of the measurements is not certified by any measurements institute, the evidence that the SLA has not been fulfilled would not hold up in court.

Finally, our gap analysis shows that privacy and anonymisation are major issues that impede the widespread deployment of measurement techniques in the Internet. On the one hand, the privacy of the user must be respected and data has to be treated in an aggregated manner. On the other hand, the analysis loses accuracy and robustness as the measurements are aggregated.

For example, if the QoS of a mobile Internet service is bad and variable, it becomes necessary to locate the bad coverage spots, by identifying users with poor service. However, geolocation of users is strongly constrained in many countries. It is most broadly used in navigation aids for drivers, where the individuals' location is necessary, but only given to that individual. Traffic information is shared, but only as aggregates without any identities. Therefore, there has to be a trade-off between the user privacy and the ability of the operator to measure and diagnose its own network, which benefits all its customers.

1.8 Standardisation

More and more End-to-end telecommunication services require resource interconnection between operators and interoperability between multiple vendors. It brings strong needs for standards for interoperability of measurement systems, comparison of the measurement results, and easily comparable Quality of Service (QoS) and Quality of Experience (QOE) parameters both in wired and wireless networks.

As easily comparable and adequate information about the QoS of retail Internet Access Services (IAS) is crucial, major SDOs (Standards Developing Organisations) have been working on QoS/QoE parameters to measure network performance, particularly ITU-T and ETSI provides QoS/QoE related standards that Electronic Communications (BEREC) and Electronic Communications Committee (ECC) refer. Yet, it is quite challenging to compare in an objective way. There are no clear standards in place yet which can support network operators in selecting QoS extension for LTE services and applications. Especially for mobile networks, KPI and key quality indicators should be applicable across all vendors and operators. In addition, it is necessary to put more effort to find the right KPI target values in a way to fulfil popular/emerging application-specific QoS requirements.

IETF and BBF are collaborating to provide large scale measurement interoperability. There is no final standard yet, and it needs to be a common way to collect and understand the results of tests across different devices to enable correlation and comparison between any networks or service parameters from different vendors.

In the first workshop of this study, the need for interoperable Layer 2 measurement was identified, but we are aware of no related standardisation. In addition, standardisation activities related to network measurement of emerging networks, such as the networks of novel devices in the IoT, or SDN networks, built with new, more modular and more self-aware software that offers application interfaces to measurement information are only at the concept stage. Standardisation for the emerging networks should be encouraged.

1.9 Conclusions and recommendations

Measurement proves to be a vital and active area, but some hard problems surface. These are the gaps we are looking for in this report, areas in which new innovations and techniques are called for if we are to solve these problems. To summarise the most important gaps, we call attention to the following:

Interoperability and coverage for end to end combining of measurement data

Interoperability between the data-generating components of the high speed backbone that come from different suppliers is a big obstacle today. To permit end to end solutions, sharing of operational information across domains, should permit problem resolution but the data are not in easily combined forms. Standards for the data involved don't seem to be happening quickly enough. Teams which try to build such systems stumble on the lack of adequate documentation, so the first step in bridging this gap is improved documentation of the data products generated by each of the incompatible systems.

Most discussion and tools for end-to-end measurement address IPv4 and L3 (the TCP/IP environment). Yet L2 is being more directly utilised in new SDN active environments, as well as for high speed control. No standards and few tools exist. Cross-domain or end-to-end problem resolution still only a dream; this is an important gap which may remain for a while. It should form a focus for research efforts stimulated by the EC.

Certification of measurements

Certification is essential for measurements in the backbone and infrastructure. Only then will it be possible to have actionable performance contracts or SLAs. When measuring at the edge of the Internet, it allows regulators to determine if end-users are receiving the data bandwidths, reliability, and availability they paid for. This starts with good experimental design, but goes beyond by requiring some control over precision and accuracy of measurements. This gap is not as large as some others, and should be addressed in the near future for home and office broadband services. In mobile it is more difficult. No standard definitions of mobile QOS (technical standards) or QOE (user requirements) are defined yet, much less accepted.

Privacy

Privacy discussion needs to be made more concrete, since geolocation of individuals, aggregated in various useful ways to produce traffic information, has become a necessary part of our daily lives. Many more aspects of the connected world we live in can be expressed as aggregates, and should be shareable. But the possibility of identifying individuals whose characteristics make them unique in some way will persist. A better understanding of how often this might occur and when it is acceptable is needed. Exploration of the privacy exposure and its impact should be a standard part of every investigation of new technology in the end-user space that the EC supports or encourages.

Automation of measurement and Big Data issues in its analysis

Greater degrees of automation are required to provide automatic problem detection, repair and rerouting or load balancing in complex networks. Increasingly ubiquitous active measurement facilities are part of the path across this gap, but the data that would be generated if every intelligent device at the edge and in the branching points of the Internet participated in measurement and monitoring raises challenging Big Data problems. At the same time to ability to obtain ample "crowd-sourced" observations essentially anywhere and almost instantly should be explored.

The distributed structure of the structure of the Internet increases the complexity of the problem. One example is "IP anycast" - the fact that BGP gracefully allows many locally advertised definitions of the optimal route to a particular resource, such as the local provider of YouTube.com, to coexist. Checking is minimal, so this can go wrong, leading to stolen or misplaced traffic and lost messages. Monitoring to assure correct function and to improve placement to optimise utilisation of the Internet's "pipes" is only done by the Content Providers today, not by the operators. Academic researchers find this an obstacle to understanding the evolving structure of the Internet, operators a challenge to managing it. Regulators and users cannot really tell how their critical services are being provided without better monitors. This calls for a systematic effort to increase the breadth of Internet monitoring for connectivity and link performance, coupling it with better machine learning-derived tools for extracting insights from the data that will emerge. And of course this must be done in a way that puts blame on providers only when it is proven, and inescapable.

2 Access Networks and the Evolving Internet

In order to generate sustainable growth in the telecommunication market, it is important for ISPs to understand the quality experienced by customers and the impact and operation of new devices and technology. In order to obtain reliable benchmarks, some ISPs use vendor-provided hardware measurement platforms that connect directly to the home gateway.

While the test capabilities of such probes are good, they are too expensive to deploy on a mass scale to enable the detailed understanding of network performance (e.g. to the granularity of a single backhaul or single user line). Besides, there is no easy way to operate similar tests on other devices (e.g. set top box) or to manage application level tests (such as IPTV) using the same control and reporting framework. ISPs also use speedtests and other diagnostic tests from user devices, but they are not able to perform continuous testing and the uncontrolled device and home network means that results are not comparable.

Regulators are also important customers of network monitoring solutions. They use the results for the development and enforcement of broadband policies. They request that the datasets are able to compare multiple broadband providers, diverse technical solutions, geographic and regional distributions, and marketed and provisioned levels and combinations of broadband services. This requires that the measurement approaches meet a high level of verifiability, and accuracy and fairness to support valid and meaningful comparisons of broadband performance.

End-users are another major customer set with respect to network performance monitoring. Performance test results from ISPs or regulators can be a good source for end-users to choose their ISP. In addition, they wish to compare advertised performance and the real performance that they receive. A number of simple consumer PC or smartphone-based tools are available to them to evaluate overall network performance and check their suppliers' claims. Also, they may perform diagnostics prior to calling their ISP to report a problem.

In order to satisfy the user demands, Both academic and industry communities developing network monitoring tools focus on providing reliable and real-time monitoring tools which can be easily integrated into network management tools to identify, isolate and fix network problems.

2.1 Fixed-line access networks

SamKnows is a global leader in broadband measurements and particularly they have partnered with regulatory agencies such as the FCC's Office of Engineering and Technology and Consumer and Governmental Affairs Bureau which has used SamKnows Whiteboxes since 2011 to "Measure Broadband in America" and produce annually the same-named report. Other regulatory agencies using SamKnows technologies are Ofcom, the UK telecom regulator, iDA Singapore's telecom regulator, ANATEL, the Brazilian regulator, and CRTC, Canadian Radio-Television and Telecommunications Commission. SamKnows is also working with BT to implement tests to monitor their network. PCPro, a UK monthly computer magazine has named SamKnows as one of the best broadband speed tests: http://www.pcpro.co.uk/features/381937/.

The probes are Linux-based hardware probes which attach to the home router and are capable of running a suite of broadband performance measurements. SamKnows has deployed over 40,000 measurement boxes, mainly in Europe, the US, Canada, Brazil and Singapore. Active tests are run on these boxes according to a predefined testing schedule applicable to that country and/or ISP.

Measurements metrics which can be tested are:

- Download speed: Throughput in Mbps utilising one or more concurrent TCP connections.
- Upload speed: Throughput in Mbps utilising one or more concurrent TCP connections.
- Web browsing: The total time taken to fetch a page and all of its resources from a popular website.
- Video streaming: The initial time to buffer, the number of buffer underruns and the total time for buffer delays.
- Voice over IP: Upstream packet loss, downstream packet loss, upstream jitter, downstream jitter, round trip latency.
- UDP latency: Average Round Trip Time of a series of randomly transmitted UDP packets.
- UDP packet loss: Percentage of UDP packets lost from the UDP latency test.
- UDP latency/loss under load: Average Round Trip Time and packet loss of UDP packets whilst the line is heavily loaded with downstream or upstream traffic.
- UDP contiguous loss: Events of two or more consecutively lost UDP packets to the same destination.
- ICMP latency: Round trip time of five regularly spaced ICMP packets.
- ICMP packet loss: Percentage of packets lost in the ICMP latency test.
- DNS resolution: The time taken for the ISP's recursive DNS resolver to return an A record for a popular website domain name.
- DNS failures: Percentage of DNS requests performed in the DNS resolution test that failed.
- FTP throughput: Throughput in Mbps at which a file can be downloaded from - or uploaded to - an FTP server.
- Peer to peer: Throughput in Mbps at which a file can be downloaded from a peer-to-peer network.

- Email Relaying: The time taken to relay an email via the ISP's SMTP servers and reach a target mail server.
- Availability: Total time the connection was deemed unavailable for any purpose, which could include a network fault or unavailability of a measurement point.
- Consumption: A simple record of the total bytes downloaded and uploaded by the router.

The European IP Network RIPE (NL) has deployed thousands of "Atlas" probes worldwide. RIPE Atlas probes are non-commercial and are available for experimentation purposes, according to a credit system; users gain credits by hosting probes, and with the accumulated credits can perform experiments which they define. If an experiment is considered by RIPE to be particularly interesting, the normally needed credits can be waived.

These are small, USB-powered hardware devices that hosts attach to an Ethernet port on their router via a network cable. They conduct different measurements and relay the results to the RIPE NCC, where it is aggregated with data from the rest of the RIPE Atlas network. Measurement metrics are relatively simple:

- Ping/Ping6.
- Traceroute/Traceroute6.
- DNS/DNS6.
- HTTP/HTTP6.
- SSLCert/SSLCert6.

While there are several companies providing network measurement solutions, a good alternative for conducting such measurements is the use of open-source tools. Measurement Lab (**M-Lab**) is an open, distributed server platform for researchers to deploy Internet measurement tools. The M-Lab tool lets users perform the following tests:

- Network Diagnostic Tool (NDT): Test the connection speed and receive a sophisticated diagnosis of problems limiting speed.
- Glasnost: Test whether certain applications or traffic are being blocked or throttled on the broadband connection.
- Network Path and Application Diagnostics (NPAD): Diagnose common problems that impact last-mile broadband networks.
- Pathload2: View how much bandwidth your connection provides.
- ShaperProbe: Determine whether an ISP is performing traffic shaping.
- BISmark Gateway: Host a router device to test Internet connectivity over time.
- WindRider: Detect whether a mobile broadband provider is performing application or service specific differentiation.
- SideStream: Collect statistics about the TCP connections used by the measurement tools running on the M- Lab platform.

- Neubot: Perform periodic tests to measure network performance and application-specific traffic throttling.

Among the M-Lab tools, the BISmark (Broadband Internet Service Benchmark) gateway is a joint project between Georgia Tech (USA) and the University of Napoli Federico II (Italy). The BISmark platform is an OpenWRT-based platform used for measuring ISP performance, as well as traffic inside the home. It has deployed programmable gateways in hundreds (typically around 200 are active at any time) of home routers (Netgear WNDR 3800 routers are used as standard, but BISmark software can be installed on other makes) in about 30 countries. The BISmark website states that among requesters, participants in regions close to measurement lab servers are given priority.

BISMark project goals are measuring access link performance and application performance, and representing the performance to users. It measures directly at the gateway device and performs periodic measurement on the "last mile" and end-to-end. The reasons to use gateway solutions for measurement are (i) to observe all traffic passing through the network, (ii) to be able to isolate individual factors affecting network performance (wireless, cross traffic, load on measurement host, end-to-end path, configuration and hardware, and (iii) to be able to isolate user behaviour.

The measurement metrics are:

- Active measurements, against a fixed set of measurement servers, of:
  - Latency: Last mile, end-to-end (ping),
  - Throughput, jitter, packet loss (upstream and downstream),
  - Capacity, available bandwidth,
  - DNS lookup time and faults.
- Passive measurements (with the consent of the users):
  - ARP and DHCP logs,
  - Aggregate flow statistics (incoming and outgoing).

Started in 2010, BISmark is designed to be extensible, with privacy, security and reliability as first-order concerns. Measurements and software upgrades are controlled centrally.

In Europe, there are many projects on network measurement and among them are two ongoing EC projects related to broadband measurement: Leone and mPlane.

The **EC project Leone** concentrates on solutions for large-scale broadband measurement overcoming the current shortcomings that measurement results are hard to compare and systems are hard to scale and integrate into existing network management tools. The project aims to improve the current shortcomings of large-scale broadband measurement by:

- An extensible architecture framework.
- The definition of new metrics.
- Introducing new analysis methods.

- Improving the network management tools of network/service providers.

It deploys SamKnows probes at customers' premises that carry out measurements at configurable intervals and report the results back to a collection infrastructure. Based on its extensible framework, it has proposed and developed new metrics in the following layers: (i) Network Layer focusing on user perception and parameters that directly influence network properties; (ii) Transport Layer by determining connection establishment time and transport protocol performance and (iii) Application Layer that specifically relates to YouTube videos and Website performance that are predictors of end-user Quality of Experience (QoE). It has also implemented features on top of measurements taken by SamKnows probes, such as:

- Visualisation of the measurements.
- Integration of the measurements and analysis results into existing network management tools.
- Analyses to determine the causes of degradation of QoE; tools attempt to combine the multiple measurements, in order to isolate which part of the Internet or service infrastructure has failed or needs to be upgraded.
- Automatic repair of the problem, where possible.

Leone uses only fixed-line broadband probes and conducts only active measurements. The framework has been designed with the following three assumptions:

- A measurement system must be under the direction of a single organisation.
- Each Measurement Agent (MA) is only associated with a single Controller at any point in time.
- Only MAs - and not measurement peers - can initiate measurement tasks and communicate with Controllers and Collectors.

The Leone Framework has three basic elements: Measurement Agents (MAs), Controllers and Collectors. The relationship and corresponding protocols are illustrated in the Figure below.

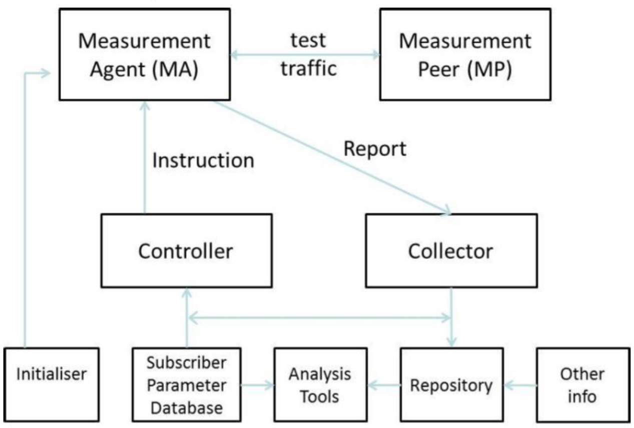


Figure 8: The Leone Framework proposal to IETF

For more detail information on Leone, please refer to Appendix A.

The **EC project mPlane** (intelligent measurement plane for the Internet) consists of a modular measurement infrastructure to perform active, passive and hybrid measurements; it operates at a wide variety of scales and dynamically supports new functionality. mPlane is about large scale network measurements, measurement data statistical analysis and mining for troubleshooting support embedding measurement into the Internet as an additional capability. In order to build a distributed, open, standard measurement infrastructure for the Internet, mPlane is focusing on providing the following technologies:

- Probes – get the data: they build on existing tools/methodologies (ping, traceroute, tstat, etc.) and offer a flexible, programmable, open platform to run and collect passive, active, hybrid measurement
- Repositories – store and preprocess the data: they collect measurements in a standard way, pre-process large amounts of data in efficient ways, and grant access to interested parties (ISP, Content Providers, end-users, regulation agencies, etc.) subject to authorisation rules.
- Intelligent reasoner – dig into the data: this component mines automatically the data and extract useful information, drills down to the root cause of a problem, and allows structured, iterative, and automated analysis

While Leone is more directly focused on measurement of access network performance, as the regulatory verification scenario described in its Deliverable 1.1, mPlane claims that it has a wider set of target scenarios leading to a requirement to integrate heterogeneous measurement components across multiple scales. The followings are some of mPlane use cases:

- Traffic anomaly detection and root cause analysis in large-scale networks
- Quality of Experience for web browsing
- Mobile network performance troubleshooting
- Verification and certification of service-level agreements
- Multimedia Content Delivery Troubleshooting
- Cloud services troubleshooting
- Estimating Content and Service Popularity for Network Optimisation

mPlane architecture aims to simplify network monitoring practices by (i) common / standard interfaces, (ii) automated intelligence (the development of the Reasoner is a key result that will allow structured, iterative, automated analysis), and (iii) support of large scale (innovative probe technology to intelligent algorithms for distributed data analysis).

The following Figure shows the general arrangement of entities in the mPlane architecture.

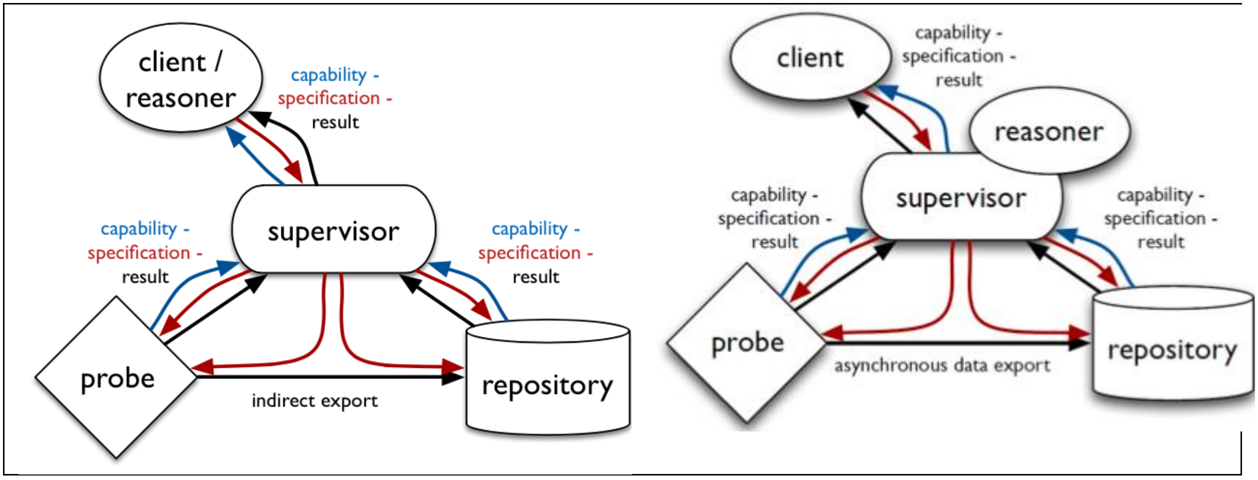


Figure 9: General arrangement of entities in the mPlane architecture

For more detail information on mPlane architecture, please refer to Appendix B.

In the US, a number of major telecom companies such as AT&T, Comcast, Verizon and Time Warner Cable have individual Speedtest platforms which test the speed between a user's computer and various servers around the country.

There are also many software-based tools that can be installed on end devices to test broadband speeds (e.g. Dasu, Grenouille, Neti@Home, BSense, Netalzyr), but they include the impact of home network devices and therefore do not give a true representation of the service provided by the ISP.

2.2 Wireless Access Networks

As mobile traffic grows and the usage of smart gadgets using Internet services is increasing, diverse mobile apps to test broadband performance have been introduced. The major leading companies or organisations (e.g. SamKnows, Ookla, RIPE) in broadband performance measurement also introduce mobile apps for the new demand. In the EU, there are organisations which have measurement tests for mobile apps such as the French ZDNet.fr 4G Monitor which shows the level of network quality of service (3G , 3G+ and 4G) for the leading French mobile operators such as Bouygues, Orange, Free and SFR.

There are many mobile apps currently on the market which provide a variety of measurement performance and analytics for mobile Internet performance. The following is a sample of mobile app leaders and their specialisation as well as some projects in the development phase.

SamKnows developed a mobile measurement app for Android and iOS platforms in 2011, with the purpose of measuring mobile Internet performance via dongles (USB modems). Unlike the fixed-line Whiteboxes and Blackboxes, the tests are not tied to a volunteer hosting the device; the Mobile dongles are used for performance measurement from any device. They are available currently in the US and Brazil.

Measurement metrics which can be tested are:

- Download speed: Throughput in Mbps utilising one or more concurrent TCP connections.
- Upload speed: Throughput in Mbps utilising one or more concurrent TCP connections.
- UDP latency: Average Round Trip Time of a series of evenly spaced transmitted UDP packets.
- UDP packet loss: Percentage of UDP packets lost from the UDP latency test.
- UDP Jitter: The interval between the minimum Round Trip Time and the average Round Trip Time.

RIPE is developing a RIPEstat mobile application for iPhones and iPads to provide the same type of information as for the wired tool.

One of **BISmark**'s passive measurements tests the wireless Access Point configuration. Also, researchers are developing techniques that isolate the source of performance bottlenecks to either the access link or the home wireless network, as well as tools that help to understand the nature of wireless networks. Since the home router sits between two common sources of performance issues - the access link and the wireless network - it is ideally suited for identifying and isolating problems between them.

WeFi is a market leader and pioneer in mobile network management – both for cellular and Wi-Fi coverage. Through online tools, embedded APIs, and downloadable apps, WeFi offers a complete suite of broadband solutions that enable service operators and Content Providers to deliver optimal user experiences to their customers.

WeFi has built the world's largest map of Wi-Fi hotspots, enabling the best connection to be found anywhere in the world.

With over 200 million Wi-Fi hotspots in their global database, WeFi performs the following user analytics:

- Manage network congestion and data offloading.
- Mine subscriber behaviour and statistics.
- Offload users to the best available networks.
- Manage content delivery speeds and bitrates.
- Intelligently deploy hotspots for increased coverage.

WeFi's target clients include Mobile Network Operators, Multi Screen Operators and Content Providers.

In a recent (May 2013) GSMA (Groupe Speciale Mobile Association) publication "Mobile Wireless Performance in the EU & the US"[2] and the November 2014 Ofcom (Office of Communications in the UK) research document "Measuring Mobile Broadband

[2] E. Bohlin, Kevin W. Caves, and Jeffrey A Eisenach, Mobile Wireless Performance in the EU & the US, GSMA, May 2013.

performance in the UK – 4G and 3G network performance"[3], the salient market and technology trend is the outperformance of 4G and 4G LTE networks over 3G networks.

According to the GSMA publication, US carriers have been rapidly and extensively deploying LTE ahead of their EU counterparts. US carriers such as Verizon Wireless, AT&T, Sprint and T-Mobile have all invested $billions in transitioning to LTE. The US is deploying LTE at a much faster pace than the EU; in 2013, 19% of U.S. connections were on LTE networks compared to less than 2% in the EU. Growth in investment in the US is translating into faster data connection speeds: US speeds are now 75% faster than the EU average, and the gap is expected to grow. In addition, US consumers use five times more voice minutes and twice as much data.

The mobile wireless market in the EU is seen as inefficient compared to the US market due to market fragmentation which not only limits market choices but discourages incentives for investment, which would facilitate a more integrated mobile wireless ecosystem and improve consumer welfare. The 3 steps recommended to authorities and policy makers in the GSMA publication to close this gap are (i) rationalising and harmonising spectrum policies, (ii) permitting efficient consolidation, and (iii) refocusing regulation on investment.

The following section highlights some of the leading performance measurement testing companies; foremost among them is Ookla (US) used to test mobile broadband services over 3G and 4G networks.

Ookla is the global leader in web-based network diagnostic applications with products including Speedtest.net, NetGauge and NetMetrics.

Speedtest.net is a web service that provides free analysis of Internet access parameters, such as connection data rate.

The service measures the bandwidth (speed) and latency of a user's Internet connection against one of many geographically dispersed servers located around the world. Each test measures the data rate for the download direction, i.e. from the server to the user computer, and the upload data rate, i.e. from the user's computer to the server. The tests are performed within the user's web browser using the HTTP protocol. Since service inception in 2014, over 6 billion speed tests have been completed.

The site also offers detailed statistics based on test results. This data has been used by numerous publications for the analysis of Internet access data rates around the world.

NetGauge is used for Internet Performance and Line Quality Testing and is a completely customisable platform which gives business the tools to measure network throughput, latency, packet loss and scan firewall ports. NetGauge's methodology provides accurate data on the following:

- Download – The tool can measure Gigabit+ speeds to gauge how much inbound traffic a connection can consistently handle, determining its Maximum Sustainable Throughput (MST).

[3] Ofcom, Measuring mobile broadband performance in the UK -4G and 3G network performance, November 13, 2014.

- Ping - By performing true socket-based and HTTP tests NetGauge can determine connection quality (latency) down to the millisecond.
- Firewall - NetGauge can scan for open ports to ensure they're set up properly or secured against intruders.
- Upload – The MST measurement technology is engineered to test a connection's outbound bandwidth, determining how quickly it can upload data.
- Jitter – NetGauge can diagnose network jitter by viewing when and how often ping times are fluctuating.
- Packet Loss - A series of UDP packets is sent from the client to the server and it is measured how many arrive.

NetMetrics is a massive database of Internet, mobile, fibre and satellite network test results that offers an unparalleled resource for organisations seeking to understand real-time global network performance.

2.3 Content providers and Net Neutrality Issues

Internet Infrastructure provides the backbone for data transmission across the globe, enabling us to efficiently stream video, connect to social media, share data in the cloud and communicate. Industry experts predict that Internet traffic will grow to more than 10 billion terabytes by 2020. To meet this demand for bandwidth, next-generation Internet Infrastructure equipment is being deployed to deliver higher speed, higher performance, and more flexible networks. Internet infrastructure should evolve to meet the Content providers requirements in terms high capacity, bandwidth and associated security and privacy concerns. Net Neutrality plays a major role in providing fair access to small to high volume Content Providers.

NETWORK NEUTRALITY

As its name indicates, Network Neutrality is about creating a neutral Internet. The basic principle driving Network Neutrality is that the Internet should be a free and open platform, almost like any other utility we use in our home (like electricity). Users should be able to use their bandwidth however they want (as long as it's legal), and Internet service providers should not be allowed to provide priority service to any corner of the Internet. Every web site (whether it's Google, Netflix, Amazon, or UnknownStartup.com) should all be treated the same when it comes to giving users the bandwidth to reach the Internet-connected services they prefer. Proponents of Network Neutrality do not want to give the ISPs too much power because it could easily be abused. However, big (Tier 1) ISPs prefer to distribute bandwidth differently depending on the service. They'd prefer to create tiers of Internet service that are more about paying for priority access than for bandwidth speeds.

Network Neutrality, as preferred by Government and users, is expected to provide an open Internet that will:

- Stimulate ISP competition.
- Prevent unfair pricing practices.

- Promote innovation and the spread of new ideas.
- Drive entrepreneurship.
- Protect freedom of speech and expression

NET NEUTRALITY (in the US context)

Network Neutrality is the principle that calls for the Internet to remain free and open - with no "fast lanes" that would allow some Content Providers to take priority over others.

In the US, President Barack Obama has come out in favour of Network Neutrality. The White House released a short video (Nov. 2014) in which the president called on the Federal Communications Commission to "implement the strongest possible rules to protect Network Neutrality."

The large Internet companies like Netflix and Google are also interested in Network Neutrality, since they might have to pay Internet Service Providers (ISPs) to get on a fast lane if such a thing existed, (that's called "paid prioritisation."). Network Neutrality is also favoured by lots of small Internet companies - the kind that might not have the means to pay for prioritisation - and dozens of public interest groups, too. When the FCC asked for comments on Network Neutrality, it received an astonishing 3.7 million replies, a vast majority urging the FCC to embrace it.

Even some ISPs say they agree with the goals of Network Neutrality. After Obama's video was released, Comcast, the biggest of them all, said that it agreed with almost everything the president called for.

Obama was quite specific about what he hoped the FCC would do: apply Title II of the 1996 Telecommunications Act to the ISPs like Comcast, AT&T, Verizon and Time Warner Cable. Title II would reclassify these companies as akin to public utilities - like the old telephone company - and would regulate them as such.

Although the president insisted that many of the more onerous parts of Title II - like price regulation - could be held in abeyance, the ISPs dread the thought of being regulated under Title II . They would prefer to be regulated under another part of the Telecommunications Act, Section 706, which calls for a lighter touch.

Then there is the question of what, exactly, Network Neutrality entails. Does it include only "the last mile" - that is, the relationship between the ISP and the Internet user? Or does it also include "interconnection" - the point at which a content company like Netflix joins the ISP's network and begins its journey to the customer? Currently, Netflix pays a fee to four big ISPs to gain uncongested access to their networks. Not surprisingly, Netflix says that Network Neutrality means it shouldn't have to pay this fee. Comcast and its ISP brethren disagree.

One reason federal Network Neutrality rules have been so difficult to achieve is that, in the past, when the FCC has tried to regulate the ISPs without using a Title II designation, the courts have essentially ruled that the FCC lacks the authority to apply rules that would ensure Network Neutrality.

On Feb 26, 2015, the FCC voted to approve new (but not yet public) rules for regulating Network Neutrality, applying Title II to regulate broadband data carriage as a common carrier service.

Network Neutrality is demonstrably a good thing, and it needs to be enshrined in law, not just done in good faith. The real problem is with the law itself: It was never meant to regulate broadband. The details of the FCC's decision are still emerging.

NET NEUTRALITY (in the European context)

The European Commission is in favour of Network Neutrality. In April 2014, the European Parliament has voted to protect Network Neutrality, limiting the power of telecoms providers to charge third parties for faster network access. Internet providers will still be allowed to offer certain specialised service at a higher price - video on demand and business-critical data-intensive cloud applications, for example – as long as these aren't supplied to the detriment of others. Currently, only the Netherlands and Slovenia have Network Neutrality laws in place and some countries, such as the Italy and UK, are deeply unenthusiastic.

"Most member states have confirmed their support to EU rules on Network Neutrality set at a level of principles, leaving more scope for BEREC [the Body of European Regulators for Electronic Communications] guidelines and national enforcement." Instead of a definition of Network Neutrality there could be a reference to the objective of Network Neutrality. However there is no consensus across member states in this context.

Italy has proposed to remove the definitions of "Network Neutrality" and "specialised services" (14 Nov. 2014), allowing broadband and telecom companies to manage traffic across their networks (and potentially offer faster speeds to companies that are willing to pay a premium) as long as they provide a minimal level of access for all online content. The European commission has expressed the concern over this proposal to give network providers the ability to offer different speeds to different sites.

While the EU focuses on Network Neutrality, six UK ISPs have been required to block access to more torrent sites, as well as few copyright infringement sites.

2.4 Internet infrastructure

In October 2013, a report entitled "Quality of Broadband Services in the EU", documented the performance experienced by 9,467 households across 30 countries (28 EU member states, Iceland and Norway), with a variety of wired fixed technologies (xDSL, Cable and FTTx).

In terms of tests, SamKnows monitors the following indicators in this study:

- Web browsing.

- Voice over IP.
- Download speed.
- Upload speed.
- UDP latency.
- UDP packet loss.
- DNS resolution.
- Video streaming.

The findings were as follows:

- Average peak download speeds: xDSL - 8.13Mbps, Cable - 52.21Mbps and FTTx - 47.74Mbps.
- The average download speed: 30.37Mbps during peak hours, 31.72Mbps during all hours, representing 75.6% of the advertised headline speed.
- The average upload speed: 8.07Mbps, representing 91.1% of advertised upload speeds.
- xDSL based services achieved 68.3% of the advertised download speed.
- Cable and FTTx achieved 89.5% and 82.7%, respectively.

In the most recent "FCC 2014 Measuring Broadband in America" report, ISP Quality of Service findings indicated:

- US ISPs now provide 101% of advertised speeds.
- On average, across all US ISPs, *sustained upload speed* was 107% of the advertised speed, closely matching the result in the 2013 Report of 108%.
- *Latency* can be a major factor in overall performance of Internet services. In the tests, latency was measured as the Round Trip Time from the consumer's home to the closest speed measurement server within the provider's network and back.
  - Across all terrestrial technologies during peak periods, latency averaged 34.9ms,
  - During peak periods, latency increased across all terrestrial technologies by between 12% and 19%.
- *Burst speed* techniques increased short-term download performance by as much as 29.3% over sustained speeds during peak periods for Mediacom, and by more than 10% for five other providers. The benefits of burst techniques are most evident at intermediate speeds of around 8 to 15Mbps and appear to tail off at much higher speeds.

Web Browsing, Voice over Internet Protocol (VoIP), and Streaming Video from the "FCC Measuring Broadband America 2014" Report

- *Web browsing*. In specific tests designed to mimic basic web browsing - accessing a series of web pages, but not streaming video or using video chat sites or applications - the total time needed to load a page decreased with higher speeds.

However, the performance increase diminishes beyond about 10Mbps, as latency and other factors begin to dominate. At high speed, a user is unlikely to experience much if any improvement in basic web browsing from increased speed – i.e. moving from a 10Mbps broadband offering to a 20Mbps offering. However, higher speeds may provide significant advantages in a multi-user household, or where a consumer is using a specific application that may be able to benefit from a higher speed.

- *VoIP*. VoIP services were adequately supported by all of the ISPs discussed in the report. However, VoIP quality may suffer during times when household bandwidth is shared by other services. The VoIP measurements utilised for this report were not designed to detect such effects.
- *Streaming Video*. The results published in this report suggest that video streaming will work across all technologies tested, though the quality of the video that can be streamed will depend upon the speed. For example, standard definition video is currently commonly transmitted at speeds from 1Mbps to 2Mbps. High quality video can demand faster speeds, with full HD (1080p) demanding 5Mbps or more for a single stream. Consumers should understand the requirements of the streaming video they want to use and ensure that their chosen broadband service will meet those requirements, including when multiple members of a household simultaneously want to watch different streamed videos on separate devices.
- *Variability of Performance.* In the previous “FCC Measuring Broadband America 2013” report, a new category of charts was added to track the variability of performance of a service provider. This is based on the percentage of users across a range of advertised speeds that experience, on average, performance levels at that speed or better. This information, commonly called a Cumulative Distribution Function, shows how speed is distributed across the population of consumers included in this survey. As in the 2013 report, the results demonstrate that consumers should be reasonably confident that the performance they receive from their ISP will be consistent with the results reflected in this report.
- *Satellite Broadband*. Results are included in this report for one satellite-based broadband service provider (ViaSat). Satellite-based broadband Internet services differ from terrestrial-based services in several key ways. In particular, because satellites broadcast wirelessly directly to the consumer, no terrestrial infrastructure has to be deployed. As a result, satellite technologies have a more uniform cost structure.

3 Measurement techniques

Measuring the Internet and in general the communication network is of fundamental importance, especially for businesses and also for residential users. For the former, it is perhaps one of the most important tools for the day to day activity of the company. For the latter, we note that Internet penetration is increasing and more and more services are being offered through the Internet, such as TV, for example.

In a previous section, we have classified the many different measurement techniques available. In this section, we provide a list of gaps that are threatening the widespread deployment of measurement techniques, from measurement interoperability, to certification, privacy and measurements in the Internet of Things (IoT). The section serves to understand what are the limiting factors of the measurement techniques, before we resort to standardisation aspects.

3.1 Measurement interoperability

Although there are many “measurements” of the Internet today, they have been developed, deployed and operate independently, so measurement results are hard to compare and systems are hard to scale and integrate into existing network management visualisation tools. There is a clear gap in the harmonisation or interoperability of data coming from different heterogeneous sources.

Current networks and IT systems feature different types of equipment from many different vendors. Such equipment provides measurement data in different formats, which can be broadly grouped by:

- QoS measurements: these are counters that reflect packet loss, available bandwidth and others such as jitter or latency.
- Logs: these are text files that in some cases have a structured and well-known format (Apache) and in some others have a totally unstructured format (application logs).
- Traffic traces: these are binary files in PCAP format.

For example, the following shows a line from a structured log in the presentation layer of a large commercial web portal:

```
11.156.8.178 - - [16/Dec/2014:01:01:09 +0100] "GET / HTTP/1.1" 200
140 JSESSIONID="-" 0,249 ANON_80
```

As it turns out, the log is structured, and the response time can be clearly identified, along with the web resource being invoked. Other system level logs, such as the following from a IBM WebSphere Application Server, also shows a structured format. As shown, it is a text file that can be parsed to extract the most relevant performance features, such as memory occupation or the time for the garbage collector to complete its duty.

```
<AF[1]: Allocation Failure. need 528 bytes, 0 ms since last AF>
<AF[1]: managing allocation failure, action=1 (0/510025904) (26843472/26843472)>
<GC(1): GC cycle started Tue Dec 2 14:18:42 2014
<GC(1): freed 416410544 bytes, 82% free (443254016/536869376), in 142 ms>
<GC(1): mark: 112 ms, sweep: 30 ms, compact: 0 ms>
<GC(1): refs: soft 0 (age >= 32), weak 13, final 4578, phantom 0>
<AF[1]: completed in 142 ms>
<GC[1]: Expanded System Heap by 65536 bytes
<GC[1]: Expanded System Heap by 65536 bytes
<GC[1]: Expanded System Heap by 65536 bytes
```

The following figure shows an example of what can be obtained from the previous log file. It shows the memory occupation from a given WAS in a real commercial network example.

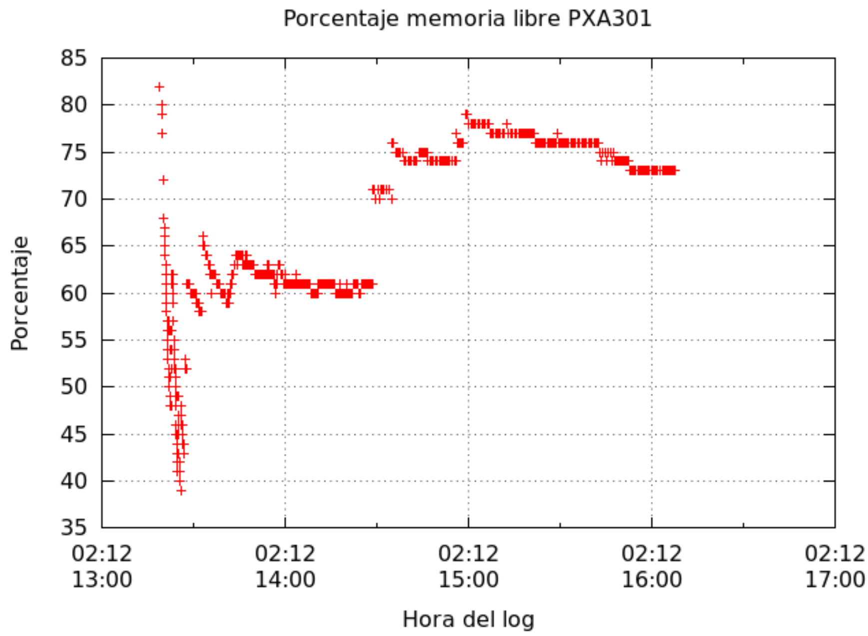


Figure 10: Memory occupation from a given WAS in a real commercial network

Clearly, the WAS shows a memory occupation peak that will affect the application performance.

In contrast, the following shows an anonymised example of unstructured log.

```
2014-12-02 14:34:59,774 INFO
es.ANON.iris.business.albaranes.mgr.AeMaquinasFranquearMgr - ****
NOTIFICACION a MAQUINAS: <?xml version="1.0" encoding="UTF-8"
standalone="yes"?>
<admision xmlns="http://ejb.scmf.ANON.es">
    <apli>IRIS</apli>
    <fechaHoraPeticion>02/12/2014 14:34:57</fechaHoraPeticion>
```

```
    <datos>
        <fechaHoraAdmision>02/12/2014 14:34:56</fechaHoraAdmision>
        <codEnvio>AE190001338140000000512</codEnvio>
        <tipoOper>A</tipoOper>
        <codiRed>1917001</codiRed>
        <numMaquina>DIPB301014</numMaquina>
        <producto>
            <codProducto>S0003</codProducto>
            <numEnvios>1</numEnvios>
            <subproductos>
                <detalle>
                    <codValorAnadido>000</codValorAnadido>
                    <numEnvios>1</numEnvios>
                    <importe>403</importe>
                </detalle>
            </subproductos>
        </producto>
        <importeAlbaran>403</importeAlbaran>
        <importeFranqueado>403</importeFranqueado>
    </datos>
</admision>
```


As it turns out, this is not so easy to parse in order to obtain meaningful performance evaluation parameters. One has to know some more detail about the application to understand if the previous text contains useful information about an application or not.

Approaches for solving data heterogeneity

A common approach to solving this issue is to come up with a standard that provides a unified format. There have been many attempts in this respect and yet the data format heterogeneity happens and it is impeding the development of new and automated techniques for performance evaluation. We note that a common standard for data is very hard to achieve in practice, because heterogeneity is inherent to the data and the standardisation effort actually tries to remove it.

For instance IPFIX is an IETF protocol, as well as the name of the IETF working group defining the protocol. It was created based on the need for a common, universal standard of export for Internet Protocol flow information from routers, probes and other devices that are used by mediation systems, accounting/billing systems and network management systems to facilitate services such as measurement, accounting and billing. The IPFIX standard defines how IP flow information is to be formatted and transferred from an exporter to a collector. Previously many data network operators were relying on Cisco Systems' proprietary NetFlow technology for traffic flow information export.[4]

Alternatively, a more flexible and realistic solution is to provide a good documentation of the data. Such documentation must be provided in the same directory of the data, as an integral part of it. It can also be provided in the same data file, as the first part of the file and into comments. This is not an elegant solution but it is truly effective and helps to circumvent many data heterogeneity problems.

Experts´opinion

[4] http://en.wikipedia.org/wiki/IP_Flow_Information_Export

We have identified this measurement data heterogeneity issue in our workshops many times. For example, in the first workshop Alessandra Scicchitano from SWITCH pointed out that vendors provide layer 2 measurement tools but unfortunately they are not compatible with each other. This means that if in a network there are switches from different vendors, these tools are not usable.

In the second workshop Jorge López de Vergara from UAM explained that nearly 80% of the interviews in our survey reflected that a major issue is the lack of data measurement interoperability.

Who is affected by data heterogeneity?

Large operators and organisations that have different heterogeneous systems in operation and find it hard to obtain a consistent unified view of the overall performance of their infrastructure/services. Regulators can also benefit from such interoperability, as they can make available to consumers comparable information about measurements of the quality of the operators' networks.

Current telecommunication operators features a large deal of measurements probes, both active and passive. On the other hand, they also have SIEM systems (Security Information and Event Management) that collect data from different log repositories and SNMP-based tools that also produce a large amount of data. The issue is how to effectively interpret and combine such data coming from different systems, with different formats and semantics.

On the other hand, there is a trend towards reduction of CAPEX and OPEX in operators, which makes it necessary to cost-effectively treat such data. To this end, current distributed database technologies such as HIVE open up promising avenues for the storage and processing of large amounts of data. The Apache Hive data warehouse software facilitates querying and managing large datasets residing in distributed storage. Hive provides a mechanism to query the data using a SQL-like language called HiveQL. HIVE uses off-the-shelf storage and it is easy to scale, thus providing a cost-effective alternative.

Therefore, a gap in the analysis of heterogeneous sources of monitoring data using distributed database technologies exists in operators. Furthermore, such data is not only useful for operational intelligence but also to offer new data analytics products to customers.

3.2 Certification of measurements and confirmation of performance

There is no easy way to certify a network measurement and this means in practice that it is not possible to enforce SLAs legally. An operator may claim that an MPLS tunnel latency is under some threshold and the measurements performed by the customer may show that is actually higher. However, there is no way for the customer to **prove it** and the operator may always say in reply that the measurement device is not precise enough.

For the citizen it is also impossible to prove that the operator is not providing, say, a given FTTH bandwidth which has been promised in the contract. There are many measurement devices but not **certified measurement devices.** Actually, UAM holds a patent[5] in this area that was used to build the certified probes currently in use by the Spanish Ministry of Industry to certify measurement performed by the Ministry in the different operators.

Standardisation efforts

The current standardisation effort focuses on providing methodologies for performance evaluation procedures but not in certification. For example, the ETSI EG 202 057-4 V1.2.1 (2008-07) standard deals with Speech Processing, Transmission and Quality Aspects (STQ) and also user related QoS parameter definitions and measurements. In part 4 (Internet access) it provides a number of techniques to measure the quality of an Internet access link, based on web downloads, namely the time it takes to download a file. The standard specifies the download file size, which depends on the access link bandwidth.

Even though the procedure to assess quality of service is clearly specified in the previous standard, no mention whatsoever is made to the precision of the measurement. For example, chances are that the clocks involved are skewed and in that case all measurements have significant noise, regardless of the measurement procedure. Much of the available software for measurements runs in user space and it is likely written in Java or Flash. Such languages lack the necessary precision to perform measurements in high-speed access links such as FTTH, because the user space does not provide real time capabilities. Therefore, the measurements can be severely skewed and become invalid for certification purposes[6].

The same does not apply to other sectors such as utilities, in which certification is an issue of primary concern. Actually, all electrical meters must be certified before they are put into production. Thus, the user is protected against possible billing mistakes, for example. In the telecommunications markets there is no way to legally sustain that an operator is not fulfilling a given SLA.

Who is affected by lack of certification?

Regulators, citizens, and large companies that make intensive use of the network. Nowadays, businesses are more and more dependent on the Internet. For example, teleworking is gaining increasing momentum, which entails that quality of service is an important factor for productivity of the teleworker, that connects to the company through the network. Therefore, companies and citizens must have the means to certify

[5] APPLIANCE FOR THE CERTIFIED MEASUREMENT OF THE BANDWIDTH OF A NETWORK ACCESS AND METHOD FOR THE CALIBRATION THEREOF), Javier Aracil Rico, Javier Ramos de Santiago , Jorge López de Vergara Méndez, Luis de Pedro Sánchez, Sergio López Buedo

Patent PCT/ES2010/070269, Spain, 28/04/2010

[6] On the effect of concurrent applications in bandwidth measurement speedometers. J. Ramos, P.M. Santiago del Río, J. Aracil, J.E. López de Vergara, Computer Networks, Vol. 55, Issue 6, pp. 1435-1453, April 2011

that the operator is not providing adequate QoS. Eventually, lack of QoS implies lack of revenues.

3.3 Privacy legislation about measurement data

Believe it or not, it is still unclear if an IP address is a personal data or not, even in presence of dynamic IP addresses. This is the case for all operators. As a result, it is very hard for an operator to effectively use the network traffic data for marketing purposes, for example. There is a gap in the state of the art in determining which data is subject to anonymization and which not.

As a result, a very conservative approach is applied to anonymise as much personal data as possible. Our experience in working with different operators in this topics shows that even ephemeral (dynamic (DHCP)) IP address are subject to anonymization. This is by far too conservative because there is no way to relate the ephemeral IP address to the user identity. The fact that much of the information needs to be unnecessarily anonymised entails that large processing capabilities are necessary to obtain the data. The cost of big data exploitation projects grows larger because of this requirement and the deployment of new services and commercial offers, based on big data analysis, becomes harder.

Much of the current operators' production is outsourced and the data anonymization requirements poses severe challenges to share the necessary operations' data with the outsourcer. In fact, many services are not deployed because of legal constraints regarding data anonymization. Therefore, there is a clear gap to be covered in the data anonymization issue, which is to clarify what data is regarded as personal and non-personal.

Target users: Operators and, in general, organisations willing to use their data.

Recent advances in "big data" analytic techniques have created crowd-sourcing opportunities that require letting many different methods be applied to large amounts of traffic data, in order to see what clues emerge, and which methods are most effective. These require stronger methods of anonymization (plus usually an NDA setting forth the permitted uses of the data and controlling its presentation and release to others), but even with personal information removed, there are ways of isolating unique behaviours and using external data to recover identities of individuals.

Orange, which is the dominant carrier in many Francophone African countries, sought community input on how its traffic data could aid development in these countries, running a series of "Data for Development (D4D) Challenges." The first challenge was issued in 2012, with the results presented at the NETMOB conference in Boston in May 2013. The datasets offered (under a simple NDA) were based on anonymised Call Detail Records (CDR) of phone calls and SMS exchanges between five million of Orange's customers in Ivory Coast between December 1, 2011 and April 28, 2012. The datasets were: (a) antenna-to-antenna traffic on an hourly basis, (b) individual trajectories for 50,000 customers for two week time windows with antenna location information, (3) individual trajectories for 500,000 customers over the entire observation period with sub-prefecture location information, and (4) a sample of communication graphs for 5,000 customers. The second challenge, using data from Senegal, is now in progress, due to conclude in May 2015.

The Orange D4D data has proved to be instructive. Useful and accurate observations were made about public transport usage (with suggestions for improvement), about detecting boundaries between wealthier and poorer districts, and about controlling spread of disease. This last study, which received first prize in the competition, proposed ways in which information campaigns could be a more effective means of disease containment than physical quarantine.

Orange was nervous about the possibility of identifying customers from the D4D public data sets. One group at Cambridge, [7] using preliminary versions of the D4D datasets, showed that the connectivity data set (4), which sampled only 5000 of the 5,000,000 customers, could be reconstructed into a connectivity and traffic graph which covered the whole country. Kirkpatrick and Bickson (D4D submission) later showed, using data from another country, that sampling a few steps away from one subscriber in 1000 could give adequate coverage to allow large scale network characteristics to be explored. But Orange took these results as evidence of a privacy exposure. As a result, the final D4D data sets were further crippled. In data set (4), a mobile phone seen in the neighbourhood of any of the sample "ego" sites was given a different ID number for its appearance around each site. This prevented linking the site information from each of the sampled sites, making their use for traffic studies very limited.

Another traditional approach to anonymization is to suppress the information which permits finding unique cases, or blur the resolution of the information made available for the same reason. This is usually expressed as "k-anonymization," the idea being that the data is suppressed or smoothed so that no query for a set of the measured attributes will return less than k results. Unfortunately, this strategy is difficult to do with rich many-featured data sets, and impossible to scale to very large sets of data.

Furthermore, a carrier wants to identify individuals within their large data sets with specific problems or characteristics (bad service or performance, potential to not pay their bill or to change carrier) in order to take specific actions with those individuals. But researchers not working on the carriers' particular problems might wish to understand aggregate performance of the network and the factors which affect it. They need accurate statistical and aggregated characteristics of the network and its traffic. Perhaps the individual data points contained in the data set can be modified by adding noise to them, so they are harder to distinguish or identify, choosing the noise contributions in such a way that the averages of interest are not affected.

There is a recent stream of work on "differential privacy" which quantifies the problem and explores corrupting the data with noise to further reduce the chance of identifying individuals. This work is described in several books and articles[8],[9].

[7] Kumar Sharad and George Danezis, "De-Anonymizing D4D Datasets," in Privacy Enhancing Technologies Symposium (PETS 2013), Bloomington Indiana (July 2013)

[8] C. Dwork, "Differential Privacy: A Summary of Results," Springer, (2008)

This formalisation of the problem characterises the probability that the presence or absence of a single individual in the data set cannot be detected by combining a query of the data set with information from a second data set. Although it seems possible to allow useful queries to be made with decreased privacy exposure, a different noising strategy is required for each query, so the method appears to be expensive in practice. This approach was originally conceived for use with social data, such as census reports. It is not so clear how queries that depend both on activities and location (or relative location) in the network can be "noised" slightly, while still leaving the results meaningful for analysis of performance or characterising the interactions of human mobility and communications preferences.

So there is a big gap between the most naive expectations of privacy, where no personal information can be discovered even by a malicious analyst, and today's ad hoc methods of anonymization. And expectations of privacy are changing with every generation of social network practices. A second area in need of better understanding and some definitions is the conditions under which data can be safely and usefully handled by non-malicious analyst, working in cooperation with carriers, regulators, and customers, to give each the information that they need.

3.4 IPv6 performance evaluation

There is a lot of open-source software for measurements, either passive or active, of IPv4 traffic but surprisingly less for IPv6. However, there is a growing share of the traffic which is IPv6, which has been adopted by Google, Yahoo, Apple, etc, for example. There is a gap in measurement instruments specifically tailored to IPv6.

Actually, a number of papers have recently shown that TCP performance on top of IPv4 differs from that on top of IPv6[10]. However, these are measurements at the IP level only and not at the network level (for example routing or latencies). Clearly, there are a lot more tools at the network level for IPv4 than for IPv6 and we foresee a demand in this area.

Who is affected by measurements of IPv6?

Google reports that approximately 4% of users accessing through IPv6, which will surely grow in the near future. Therefore, we expect that users affected by IPv6 measurements will cover the entire Internet eventually.

[9] C. Dwork and A. Smith, Differential Privacy for Statistics: What We Know and What We Want to Learn," Journal of Privacy and Confidentiality, (2010)

[10] Measuring YouTube from Dual-Stacked Hosts Saba Ahsan, Vaibhav Bajpai, Jörg Ott, Jürgen Schönwälder Passive and Active Measurement Conference (PAM 2015), New York, March 2015 http://www.netlab.tkk.fi/~jo/papers/2015-03-PAM-YouTube-Dualstacked.pdf

Measuring TCP Connection Establishment Times of Dual-Stacked Web Services Vaibhav Bajpai, Jürgen Schönwälder 9th International Conference on Network and Service Management, (CNSM 2013) Zürich, October 2013.http://vaibhavbajpai.com/documents/papers/proceedings/dualstack-tcp-cnsm-2013.pdf

3.5 Internet of Things

The IoT area is still emerging and both how to apply sensor techniques and mobile communication to the Internet is still difficult.

In the second SMART workshop Maria Teresa Herrera Zamorana from Telefonica mentioned how even in a very rural area with little mobile traffic, saturation can occur. A local farmer had put many sensors in his garlic field to monitor and supervise his crop. However each unit was connected to the mobile network which has only one Base Station in the area. The solution to create higher capacity in the mobile network is most surely not cost effective and other solutions are needed.

One major mobile phone manufacturer in Europe has in a workshop indicated that creating architectures for IoT in building small and large scale applications like eCare in the home and large scale surveillance using IoT is demanding. These uses will require measurement and design tools in order to make it possible to create applications that are cost effective (e.g. charging in the mobile network and global traffic volumes) and also performance wise. Reliability in partially working equipment and security will pose new challenges.

3.6 Other issues

In addition to the areas mentioned above, which are all good and need to be developed, there are other areas that need to be addressed:

- **End-to-end network troubleshooting.** When there is a network problem, is very hard to know where in the network the problem arises. Operators express frustration at this, being blamed for problems that they believe may be caused by over-the-top providers, or within customer premises. The NRENs also express frustration at not being able to diagnose what is happening at their customers' sites. Over-the-top providers blame operators. Home users and other customers don't know who is responsible, and are largely helpless in ameliorating their situation. Regulators are not well positioned to help the users. Their measurements are often restricted to being within the operators' networks, and even just a part of the network, although we are seeing the start of measuring some services.
- **Wireless measurements.** Most measurements of the quality of cellular systems are based upon outdoor surveys of signal strength. But customers are very often indoors. How do the measurements actually align with user experience? Since more and more of Internet access is over hand-held mobile devices, this is an area of increasing importance.
- **Quality in third party premises.** People are paying for network access in airplanes, in hotels, in trains, in cyber cafes. They may get it for free in metros, in parks, and elsewhere. But how good is the service that they are getting? In these cases, there is another party standing between the end-user and the service provider. Sometimes that party acts directly to stop the user availing him- or herself of another service (see Marriott's efforts to eliminate private hot-spots, at

least in the US). In addition to measuring the quality, can we measure attempts to interfere?

- **Censorship, discrimination.** Can we detect how communications are being suppressed? Can we detect when we are being treated differently based upon our location? Can Content Providers detect where users are in order to apply different national rules regarding intellectual property rights. This is a collection of different questions that perhaps deserve to be separated out.
- **Systems for sharing network data.** Can we systematise how to publish and how to subscribe to sources of network data? We are operating in an “Internet” of data sources and data sinks, with multiple autonomous agents interested in sharing data, but based upon policies, perhaps in exchange for payment, and certainly not with everyone. There are some attempts at cross-domain data sharing, notably perfSONAR.

4 Standardisation

Various technologies have developed over time and standards have evolved in different directions for the different technologies. In order to get reliable benchmarks, some ISPs use vendor provided hardware measurement platforms that connect directly to the home gateway, and it is very difficult to extract and correlate the performance information between different systems and vendors.

End-to-end telecommunication services require resource interconnection between operators and interoperability between multiple vendors. Regulators want better measurement of the quality of experience of consumers. It brings strong needs for standards for interoperability of measurement systems, comparison of the measurement results, and easily comparable Quality of Service (QoS) parameters.

Network measurement data is important for diverse stakeholders – ISPs, regulators, end-customers, government agencies concerned with the security/integrity of the networks (separate from policy makers or regulators), researchers and developers, and Content Providers and/or content delivery service providers. According to the needs, standard development organisations and industry forums have been working on measurement methodologies, systems, protocols, parameters and metrics for a long time.

4.1 QoS/QoE parameters

Service Providers and network operators have trusted brands, the maintenance of which is critical to their business. The challenge is to make new technology work in a way that meets customer expectations for quality, availability and reliability, while still offering network operators the flexibility needed to adapt quickly to new technology. Quality of Service (QoS) parameters are a key factor in the roll-out of new technology. New network services in conjunction with the use of smartphones require new QoS measurement methods, reference data and load profiles in order to guarantee the quality of new services. These encompass web browsing, data transfer, video streaming, video live traffic, video sharing and LTE RCS location services.

As easily comparable and adequate information about the Quality of Service (QoS) of retail Internet Access Services (IAS) is crucial, major SDOs have been working on QoS /QoE parameters to measure network performance. Yet, it is quite challenging to compare in an objective way. Electronic Communications Committee (ECC) states three main problems on QoS in its report (ECC-REPORT195[11]) as followings:

- ISPs are measuring (if at all) different sets of Quality of Service (QoS) parameters;

[11] ECC Report 195, Minimum Set of Quality of Service Parameters and Measurement Methods for Retail Internet Access Services, approved in April 2013

- Non-harmonised definitions and methodologies applied for the measurement of the QoS parameters give non-comparable values among different ISPs even in # case of similar QoS parameters;
- Consolidated information regarding QoS values from different ISPs is available in just a few countries across Europe.

In order to guide common parameters on network performance measurement, European Commission published its official Journal of the EU[12] containing norms and specifications for networks and services in electronic communication in March 2007. Chapter VII of the Document details the recommended quality parameters of electronic services from the point of view of the end-user. These parameters are to be found in ETSI Guides: ETSI EG 202 009, ETSI EG 202 057-1 and in Recommendation ITU-T G.1020, ITU-T Y.1541. Thus, these standards are used for the guidance of broadband performance monitoring in the Body of European Regulators for Electronic Communications (BEREC) and Electronic Communications Committee (EEC).

Including the standards recommended from the EC, some more examples of user related QoS/QoE from ITU-T and ETSI are explained below.

ITU-T has a long history on network measurement, especially on QoS / QoE parameters. There are approximately 175 Recommendations related to network QoS. ITU-T SG12 is the main group on these standardisation activities but also SG2, SG 9, SG11, and SG 13 are also involved in. ITU-T provides definition of QoS and QoE in ITU-T Recommendation E.800 that first provided the definition of Network Performance (NP) as not directly visible to the users and Quality of Service (QoS) as offered to the user. ITU-T Recommendation G.100/P.10 defines Quality of Experience (QoE). BEREC's "Guidelines for QoS in the scope of Network Neutrality[13]" refers quality-related terminologies based on ITU recommendations (particularly ITU-T Recs. E.800, E.802 and Y.1540).

A few examples of related ITU-T standards are as follows:

- ITU-T Rec. P.564 (11/2007), Conformance testing for voice over IP transmission quality assessment models: it specifies the minimum criteria for objective speech quality assessment models that predict the impact of observed IP network impairments on the one-way listening quality experienced by the end-user in IP/UDP/RTP-based telephony applications (3.1-kHz narrow-band in the main body, 7 kHz wideband in Annex B). Models compliant with this Recommendation predict mean opinion scores (MOS) on the ACR listening quality scale. It is expected that the primary applications for such models are monitoring of transmission quality for operations and maintenance purposes, and measurements in support of service level agreements (SLAs) between service providers and their customers. P.564-conformant models may be deployed both in endpoint locations and at mid-network monitoring points.
- ITU-T Rec.G.1020 (11/2003) Series G: Transmission Systems and Media; Digital Systems and Networks; Quality of Service and Performance; End-User

[12] COMMISSION DECISION of 11/XII/2006 List of standards and/or specifications for electronic communications networks, services and associated facilities and services, replacing all previous versions (Official Journal EU L 86/11 of 27.03.2007) – Chapter VII

[13] BEREC Guidelines for quality of service in the scope of net neutrality, BoR (12) 131 , November 2012

Multimedia QoS Categories: is to define packet network and terminal performance parameters that better reflect the perceived quality of the target applications. It is largely focused on quality impairments resulting from delay variation and packet loss which are peculiar to IP and other packet-based technologies.

- ITU-T Rec.Y.1540 (03/2011) Internet protocol data communication service – IP packet transfer and availability performance parameters: defines parameters that may be used in specifying and assessing the performance of speed, accuracy, dependability and availability of IP packet transfer of international Internet Protocol (IP) data communication services.
- ITU-T Rec.Y.1541 (12/2011) Network performance objectives for IP-based services: it specifies the objectives of end-to-end network performance or performance between user network interfaces (UNI's) based on Y.1540 parameters. It provides technical parameters for the differentiation of IP network traffic classes, encompassed by a huge number of appendices explaining application scenarios and background.
  - ITU-T Rec.Y1541 Amendment 1 (12/2013) New Appendix XII – Considerations for low speed access networks
  - IETF RFC 5976, Y.1541-QOSM: Model for Networks Using Y.1541 Quality-of-Service Classes: describes a QoS-NSLP Quality-of-Service model (QOSM) based on ITU-T Recommendation Y.1541 Network QoS Classes and related guidance on signalling.
- ITU-T Rec. Y.1542 (06/2010), Framework for achieving end-to-end IP performance objectives: it considers various approaches toward achieving end-to-end (UNI-UNI) IP network performance objectives. It describes some ideas regarding how the end-to-end performance objectives specified in Y.1541 can be achieved in multiple-carrier environments. It mainly describes two different approaches:
  - Impairment Allocation, which assigns a subset of end-to-end performance objectives to each provider on the path, thus achieving the total end-to-end performance objectives, and
  - Impairment Accumulation, which accumulates the sum of the performance budget commitment from each provider on the path and evaluates whether the end-to-end QoS requirements are fulfilled.
- ITU-T Rec. P.863 (09/2014) Perceptual Objective Listening Quality Assessment (POLQA): POLQA is the next-generation voice quality testing technology for fixed, mobile and IP based networks. It covers a model to predict speech quality by means of digital speech signal analysis. POLQA voice quality test included unified communications, next generation networks, 3G and 4G/LTE. It offers an advanced level of benchmarking accuracy and adds significant new capabilities for wideband and super-wideband (HD) voice signals along with support for most recent voice coding and VoIP transmission technologies. (http://www.polqa.info).

ETSI is also an important SDO in providing standards on QoS/QoE parameters and its measurement. ETSI has been particularly active in Interoperability events on speech quality.

ETSI E2NA and NTEC Technical Bodies are responsible for fixed networks and for migration from switched circuit networks to packet-based networks with an architecture that can serve in both. E2NA is responsible for QoS studies in Next Generation Networks.

ETSI STQ is ETSI's technical committee for Speech and multimedia Transmission Quality. Upcoming new services and technologies hold new challenges for end-to-end media quality as perceived by the user and for Quality of Experience (QoE), both subjects being addressed by STQ. New interests for future study include: (1) Impact of poor transmission quality on intelligibility, (2) e-Inclusion for people with disabilities and for ageing, Loudness vs. Loudness Ratings, (3) Concepts of 'Network Neutrality' and ‘differentiated traffic management' and their potential impact on QoS as perceived by the user, and (4) Synchronisation and interworking in packet based networks.

- ETSI TS 102-250-1 ~ TS 102 250-6, Speech Processing, Transmission and Quality Aspects (STQ); QoS aspects of popular services in GSM and 3G networks:
    - Part 1: Identification of Quality of Service aspects.
    - Part 2: Definition of Quality of Service parameters and their computation.
    - Part 3: Typical procedures for Quality of Service measurement equipment.
    - Part 4: Requirements for Quality of Service measurement equipment.
    - Part 5: Definition of typical measurement profiles.
    - Part 6: Post processing and statistical methods.
- ETSI EG 202 057-1 ~ EG 202 057-4, Speech processing, Transmission and Quality Aspects (STQ); User related QoS parameter definitions and measurements:
    - Part 1: General.
    - Part 2: Voice telephony, Group 3 fax, modem data services and SMS.
    - Part 3: QoS parameters specific to Public Land Mobile Networks (PLMN).
    - Part 4: Internet access.
- ETSI EG 201 769, Speech processing, Transmission and Quality Aspects (STQ); QoS parameter definitions and measurement; parameters for voice telephony service required under the ONP Voice Telephony Directive 98/10/EC.
- ETSI TR 101 578, Speech and multimedia Transmission Quality (STQ); QoS aspects of TCP-based video services like YouTube.

ECC-REPORT195[1] provides a table of (technical and administrative) QoS parameters with definition and reference of standards. Here, to emphasise where the reference standards are used, parameter names and reference standards are copied from the table. Full table including definitions of the terminologies is in the section 4 of ECC-REPORT195.

Parameter	Standard reference	Notes
Technical parameters		
Data transmission speed	**ETSI EG 202 057-04**	For upstream and downstream
Mean Data Rate (FTP/HTTP/Email)	ETSI TS 102 250-2	Dedicated to mobile Internet access services
Percent IP service unavailability (PIU)	ITU-T Rec. Y.1540	
Service availability	-	Taken from a national regulation
IP packet transfer delay (end-to-end) (IPTD)	ITU-T Rec. Y.1540	
Delay	-	Taken from a national regulation
Delay (one way transmission time)	**ETSI EG 202 057-04**	
Ping Round Trip Time	ETSI TS 102 250-2	Dedicated to mobile Internet access services.
End-to-end 2-point IP packet ***delay variation***	**ITU-T Rec. Y .1540** and more details in **ITU-T Rec. Y.1541** Annex II.	Network performance parameter
IP packet loss ratio (IPLR)	**ITU-T Rec. Y .1540**	
Unsuccessful data transmission ratio	ETSI EG 202 057-04	
Data Transfer Cut-off Ratio [%] (FTP/HTTP/E-mail)	ETSI TS 102 250-2	Dedicated to mobile IAS
IP packet error ratio (IPER)	**ITU-T Rec. Y .1540**	
Login time	ETSI EG 202 057-4	
Successful log-in ratio	ETSI EG 202 057-4	
DNS host name resolution failure ratio	ETSI TS 102 250-2	
DNS host name resolution time	ETSI TS 102 250-2	
Parameters applicable for mobile Internet access services: Service non-accessibility, Setup time, IP-Service access failure ratio, IP-Service setup time, Session failure ratio (FTP/HTTP/E-mail)	ETSI TS 102 250-2	
Administrative / Non-technical parameters		
Supply time for residential Internet access	ETSI TS 102 250-2	
Fault report rate per fixed access lines	ETSI TS 102 250-2	
Fault repair time for fixed access lines	ETSI TS 102 250-2	
Response time for operator services	ETSI TS 102 250-2	

Frequency of End-user complaints	ETSI TS 102 250-2	
End User complaints resolution time	ETSI TS 102 250-2	
Bill correctness complaints	ETSI TS 102 250-2	
Prepaid account credit correctness complaints	ETSI TS 102 250-2	
Subjective evaluation		
Global user satisfaction	ITU-R BS.1116-1, ITU-R BT.500-13, ITU- T Rec. P.800, ITU-T Rec. P.910	Level of satisfaction of the users expressed in MOS (Mean Opinion Scores)

Note: With its analysis of each of the technical parameters, the parameters and corresponding references in **bold fonts** are recommended to be used within the Minimum Set.

4.2 Large-scale measurement interoperability

Various technologies have developed over time and standards have evolved in different directions for the different technologies. There are numerous standards on network measurements from many SDOs. Even only looking at the IETF, several Working Groups (WGs) such as IPFIX (IP Flow Information Export) WG, PSAMP (Packet Sampling) WG, MB (Benchmarking Methodology) WG, PMOL (Performance Metrics for Other Layers) WG and CCAMP (Common Control and Measurement Plane) WG have published numerous RFCs on protocols and metrics for network measurement.

As network services are evolving in a direction where end-to-end telecommunication services require resource interconnection between operators and interoperability between multiple vendors, new requirements for the overall architecture and solutions for an interoperable measurement infrastructure get stronger. Among early efforts in this area, ITU-T published **Y.1543 (**Measurements in IP networks for inter-domain performance assessment) in November 2007 that specifies the measurement methodology for a network with a multi-provider environment. It specifies a set of IP performance parameters and methods of measurement applicable when assessing the quality of packet transfer on inter-domain paths. This recommendation specifies the measurement options, so that performance measurements conducted by operators in their administrative domains can be easily combined to estimate the end-to-end network performance or the inter-domain QoS. However, the work is focusing on providing QoS parameters rather than providing interconnecting protocols for measurement.

In order to provide a common information model for IP traffic measurement, ETSI has established an Industry specification group, MOI (Measurement Ontology for IP traffic) in 2009. This group identifies that there are many systems to monitor network traffic, providing measurements about delay, jitter, capacity, packet loss, etc. that use different data structures, provide data in different units and sometimes use different algorithms. There are several incompatible information models f or network measurements. Therefore, a common information model of network measurement is needed and a common information model of network measurement parameters and units has to be agreed. This group has published the following three documents:

- ETSI GS MOI 010 (05/2010), Measurement Ontology for IP traffic (MOI); Report on information models for IP traffic measurement: It constitutes an analysis of information models for IP traffic measurement. This will include the basic definitions and state-of-the-art study, as well as the main guidelines to specify a complete set of vocabulary of classes and relations to describe Internet measurements, supporting QoS parameters and offering privacy protection, by studying existing schemas that are currently used to describe such information.
- ETSI GS MOI 002 (07/2012), Measurement Ontology for IP traffic (MOI); Requirements for IP traffic measurement ontologies development: identifies the requirements that should characterise ontology for the semantic conceptualisation of information related to IP traffic measurements.
- ETSI GS MOI 003 V1.1.1 (2013-05), Measurement Ontology for IP traffic (MOI); IP traffic measurement ontologies architecture: It provides the high level structure description of an ontology for interfacing and data exchange with IP traffic measurement devices. It is based on preceding work item GS MOI 010 and meets the requirements presented in GS MOI 002.

In order to provide overall architecture and protocols for interoperable measurement systems, IETF LMAP-WG has launched in 2012, and Broadband Forum (BBF) also started its standardisation on the interoperable large-scale measurement in 2012. While LMAP framework has been designed in the LMAP WG, IPPM-WG has started to define related metrics for LMAP and BBF solutions. More details of these efforts are listed below.

- The following three on-going standards are the current official work in LMAP-WG:
- A framework for large-scale measurement platforms (LMAP), draft-ietf-lmap-framework-10 (01/2015): Measuring broadband service on a large scale requires a description of the logical architecture and standardisation of the key protocols that coordinate interactions between the components. The document presents an overall framework for large-scale measurements. It also defines terminology for LMAP. It is noted that LMAP assumes that the measurement system is under the direction of a single organisation, and each MA may only have a single Controller at any point in time.
- Information Model for Large-Scale Measurement Platforms (LMAP), draft-ietf-lmap-information-model-03 (01/2015): This Information Model applies to the Measurement Agent within a Large-Scale Measurement Platform. As such it outlines the information that is (pre-)configured on the MA or exists in communications with a Controller or Collector within an LMAP framework. The purpose of such an Information Model is to provide a protocol and device independent view of the MA that can be implemented via one or more Control and Report protocols.
- Large-Scale Broadband Measurement Use Case, draft-ietf-lmap-use-cases-05 (11/2014): Measuring broadband performance on a large scale is important for network diagnostics by providers and users, as well as for public policy. Understanding the various scenarios and users of measuring broadband

performance is essential to development of the Large-scale Measurement of Broadband Performance (LMAP) framework, information model and protocol. This document details two use cases (ISP use case and Regulator use case) that can assist in developing that framework.

IETF IPPM-WG has been working on IP performance metrics, methodologies and protocols between test equipment, focusing on active measurement. Several IPPM standards are referred in the design of QoS parameters defined in ITU-T and ETSI. Among IPPM's standardisation activities, the followings are current on-going official documents directly related to LMAP mechanisms:

- A Reference Path and Measurement Points for Large-Scale Measurement of Broadband Performance, draft-ietf-ippm-lmap-path-07 (10/2014): It defines a reference path for LMAP and measurement points for commonly used performance metrics. Other similar measurement projects may also be able to use the extensions described here for measurement point location. The purpose is to create an efficient way to describe the location of the measurement point(s) used to conduct a particular measurement.
- Registry for Performance Metrics, draft-ietf-ippm-metric-registry-01(10/2014): Despite the importance of Performance Metrics, there are two related problems for the industry. First, how to ensure that when one party requests another party to measure (or report or in some way act on) a particular Performance Metric, then both parties have exactly the same understanding of what Performance Metric is being referred to. Second, how to discover which Performance Metrics have been specified, so as to avoid developing new Performance Metric that is very similar. The problems can be addressed by creating a registry of performance metrics. This document therefore creates a Performance Metrics Registry. It also provides best practices on how to define new or updated entries in the Performance Metrics Registry.

The Broadband Forum (BBF) develops specifications for broadband wire-line solutions. In order to add more performance tests on its earlier work of TR-069 and TR-143, the BBF started WT-304 in 2012. It includes loss and jitter test, tests with emulated streaming, browsing, etc. The followings are related three standards:

- TR-069: Data models to monitor the Customer Premises Equipment (CPE) using defined diagnostic mechanisms.
- TR-143: Enabling Network Throughput Performance Tests and Statistical Monitoring.
- BBF WT-304, Broadband Access Service Attributes and Performance Metrics. This standard is started in autumn 2012 and is targeting large scale test control and reporting based on TR-069 and TR-143 which only defined throughput and response time tests, by adding more performance tests such as loss and jitter and tests with emulated streaming, browsing and so on.
    - It defines a set of common Broadband Service Attributes (BSA) along with acceptable performance measurement methods/architectures that recommend how to measure those service attributes.

- o It aims for a more flexible capability that can, for example, measure particular segments of the network, measure across multiple networks, schedule continuous tests and allow on-demand triggering of tests.

IETF LMAP and BBF WT-304 have shared framework and protocols. The LMAP and BBF WT-304 Framework have three basic elements: Measurement Agents (MAs), Controllers and Collectors. MAs initiate the actual measurements, which are called Measurement Tasks in the LMAP terminology. The Controller instructs one or more MAs and communicates the set of Measurement Tasks an MA should perform and when. The Collector accepts Reports from the MAs with the Results from their Measurement Tasks. For communication of these entities, there is Control Protocol that runs from a Controller to instruct Measurement Agents what performance metrics to measure, when to measure them, how/when to report the measurement results to a Collector; secondly, a Report Protocol is for a Measurement Agent to report the results to the Collector. The following diagram shows the scope of standardisation among the three groups:

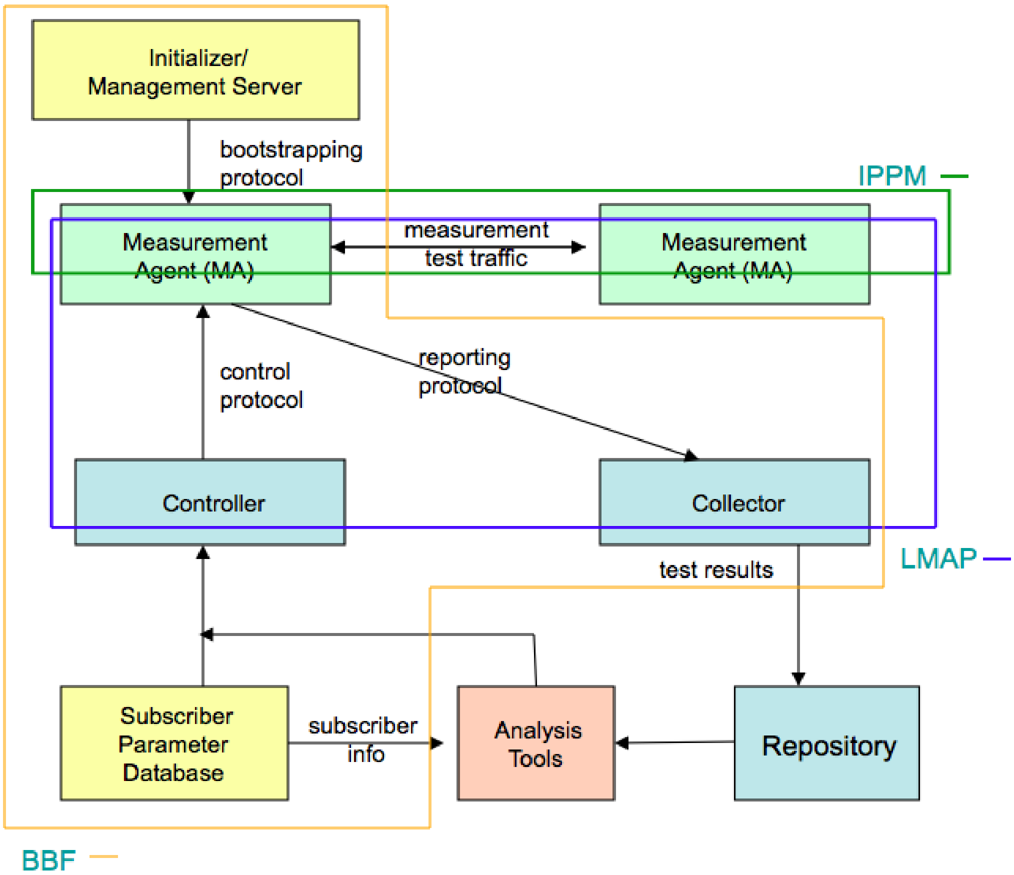


Figure 11: Scopes of IETF LMAP WG, IETF IPPM WG, and Broadband Forum

IETF LMAP and BBF are planning interoperability plugfest when protocol standards are mature to develop protocol specific test case (i.e., authentication, etc.). The first interoperability plugfest is planning in April 2015 with testing controllers and MAs, and the second one is planning in June 2015 with testing controllers, MAs and Collectors.

4.3 Wireless and Mobile networks

IEEE P802.16.3, Architecture and Requirements for Mobile Broadband Network Performance Measurement (Draft working document) identifies Mobile-specific considerations as followings:

- Measurements will typically be related to a specific user device, rather than to a router on a LAN.
- A single user device can typically operate with multiple disparate network technologies.
- A single user device may connect with multiple operators.
- A user device experiences widely varying signal and network conditions.
- Due to variability, far larger statistical samples may be required to draw generalised conclusions.
- Significantly more metadata (including, for example, location information) is required to characterise the scenario of a specific sample.
- It may be necessary to trigger testing based on a set of environmental circumstances, such as location, rather than relying upon scenarios such as LAN quiescence as a trigger.
- Active testing may be relatively more constrained due to practical issues, including data plan limits and battery consumption.
- Underlying software on many mobile devices is relatively closed, and underlying network data is often relatively difficult to access.

Considering such specific requirements, major SDOs have been working on QoS/QoE issues and performance measurement on wireless and mobile network.

The followings are selected examples of QoS/QoE in wireless and mobile networks:

- IEEE P1907.1, Standard for Network-Adaptive Quality of Experience (QoE) Management Scheme for Real-Time Mobile Video Communications:
  - Definition of an End-to-End Quality of Experience (E2E QoE) metrics and management scheme for real-time video communication systems.
  - Utilisation of correlation of both subjective and objective E2E QoE with received real-time video data (stream header and/or video signal), application-level QoS measurements, and network-level QoS measurements.
- 3GPP™ Technical Specification Group SA2 covers Quality of Service requirements for access to both packet and circuit switched GSM and 3G networks.
- 3GPP TS 22.105 defines QoS parameters for bearer services and teleservices, and sets expectations for measures such as delay, delay variation, loss, etc. for each of the service types identified
- 3GPP TS 23.107:"Technical Specification Group Services and System Aspects; Quality of Service (QoS) concept and architecture": it develops the QoS framework for the UMTS bearer service.

- 3GPP TS 23.207:"Technical Specification Group Services and System Aspects; End-to-end Quality of Service (QoS) concept and architecture".
- 3GPP TR 26.944, "End-to-End multimedia services performance metrics" defines metrics in three categories: Quality of Experience (QoE), End-to-end Service Quality of Service (ESQoS), and System Quality of Service (SQoS)
- ETSI TR 103 114, Speech and multimedia Transmission Quality (STQ); QoS Parameters and measurement methodology for Smartphones

Related to Long-Term Evolution (LTE), currently ETSI and ITU-T are working on QoS/QoE issues in LTE networks:

- ETSI STF 437, QoS of connections from current technologies to LTE for delay sensitive applications (ETSI Held an interop event on end-to-end QoS assessment for VoLTE and RCS.) The scope of the project is:
  - Address LTE related QoS problems for delay-sensitive applications.
  - Shortcomings of standards and implementations.
  - Possible solutions (actions for standards and implementations).
- ETSI TS 103 189 "End-to-end QoS assessment for VoLTE and RCS Interop Event" was held in November 2013. In four test setups (acoustic to acoustic, acoustic to electrical, electrical to acoustic, electrical to electrical), various metrics are tested in a category of:
  - End-to end test based on instrumental assessment of speech samples (POLQA model deployed).
  - End-to-end test based on instrumental assessment of video samples (PEVQ model deployed).
  - Other end-to-end tests in voice channel.
  - End-to-end tests based on functional parameters.
  - Network performance parameters.
- ITU-T Rec. E.804 (02/2014), QoS Aspects for popular Services in Mobile Networks: It provides sets of QoS parameters from an end-user perspective for the operational aspects of mobile communication. It adopted ETSI TS 102 250 for mobile quality of service benchmarking testing.
- ITU-T Draft Rec. G.VoLTE, End-to-end performance for managed voice over LTE networks (planning to finish in 2015): It provides some typical end-to-end scenarios are described, involving cases with LTE access at both sides of the communication, or with a different access technology at one side (wireless or wireline access).
- ITU-T Draft Rec. P.VTQ-M, Framework for parametric models for voice quality in mobile networks (planning to finish in 2016).

The followings are selected examples of performance measurement in wireless and mobile networks:

- IEEE 802.3ah, Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications Amendment: Media Access Control Parameters, Physical Layers, and Management Parameters for

Subscriber Access Networks: it was later included in the overall standard IEEE 802.3-2008 defines Ethernet in the first mile describing mechanisms for monitoring and troubleshooting Ethernet access links. Specifically it defines tools for discovery, remote failure indication, remote and local loopbacks and status and performance monitoring.

- ITU-T G.8013/Y.1731 (11/2013), OAM functions and mechanisms for Ethernet based networks: it defines performance monitoring measurements such as frame loss ratio, frame-delay and frame-delay variation to assist with SLA assurance and capacity planning. For fault management the standard defines continuity checks, loopbacks, link trace and alarm suppression (AIS, RDI) for effective fault detection, verification, isolation and notification in carrier networks.
- IEEE P802.16.3 started in autumn of 2012 is focusing more on general issues for mobile broadband performance measurement. The scope is standardising metrics and test procedures for provision framework for characterising and assessing the performance of various mobile broadband networks. Also, it aims to standardise protocols and data formats to allow a measurement server to coordinate and manage test operation and data collection.
  - It addresses end-to-end measurements to characterise the performance of mobile broadband networks from a user perspective. It is not limited to any particular air interface.
  - It describes the need to collect metadata associated with a measurement, such as the device's location, the cell ID and maybe radio resource control parameters – and perhaps even a measurement could be triggered based on the value of some metadata.
  - The remaining battery power and network cost/allowance are examples of environmental conditions for test execution.
- IETF LMAP claims that there is no difference between fixed service and mobile (cellular) service used for Internet access. Hence, in LMAP mechanisms, like measurements will take place on both fixed and mobile networks. Fixed services include technologies like Digital Subscriber Line (DSL), Cable, and Carrier Ethernet. Mobile services include all those advertised as 2G, 3G, 4G, and LTE. A metric defined to measure end-to-end services will execute similarly on all access technologies. Other metrics may be access technology specific. Details in IETF LMAP is in the section of "Large-scale measurement interoperability."

4.4 New measurement areas for emerging networks technologies (Virtual networks, IoT, etc.)

There are new measurement areas for emerging networks such as virtualised networks, IoT, etc. There are active standards efforts on IoT from diverse SDOs and Forums, but the issues are focused on architecture, protocols (mostly for connectivity and application support), application scenario, service requirements and mechanisms, etc. Both of the standards technologies and market deployment of IoT are not yet mature to discuss about traffic measurement for IoT. However, considering the standardisation activities to support heterogeneous and low power IoT networks and nodes are very active leading by industry, it is a matter of time that we need to consider the network traffic measurement and management for the traffics from the millions of IoT devices, particularly connected with IPv6. But like the other topics, the different groups working on standards to connect devices to each other as part of the Internet of Things will eventually need to work together or the industry will need to decide on a select few.

Standardisation of the Virtual Networks is also similar with IoT. Many are focused on architecture and protocols, and not much work has been launched in measurement of the emerging networks. The followings are identified work:

ITU-T has published a standard related to smart ubiquitous networks and an ongoing work on NGN:

- ITU-T Rec. Y.3042 (04/2013), Smart Traffic Control and Resource Management Functions for Smart Ubiquitous Network: It specifies the smart traffic control and resource management functions for Smart Ubiquitous Network (SUN) and includes motivation and objectives of smart traffic control and resource management; requirements for smart traffic control and resource management for SUN; high-level architecture and functional architecture; and control and management mechanisms.
- ITU-T Draft Rec. Y.QMF, Architectural Framework for QoE Management in NGN.

IETF has an ongoing work on

- IETF Address Resolution Delay Metric in Software-Defined Networking (SDN)(draft-pan-ippm-sdn-addr-resolv-perf-00): It defines a metric that characterises the provision performance of SDN. It is noted that this draft is not an official work item of the group at the moment that this report is written.

4.5 Other activities

In the first workshop, the necessity of interoperable layer 2 measurement was discussed. In the current survey, no related standardisation activity was identified.

However, there is an activity related to layer 2 measurement information:

- IETF RFC 7133, Information Element for Data Link Layer traffic measurement: This document describes Information Elements related to the data link layer. They are used by the IP Flow Information Export (IPFIX) protocol for encoding measured data link layer traffic information.

4.6 Discussion on Gaps in standardisation

Gaps in QoS/QoE

There are some agreed labels that could be easily interpreted by the public would give benefits for customer choices and, in return, could lead to an increase in user satisfaction. Sometimes non-technical solutions (for standardised services) may help to increase users' QoE. As an example, UK Ofcom's "easy to understand" labels are introduced in Appendix B.

Some efforts related to defining application specific QoS are identified (e.g., IETF RFC7285, Application-Layer Traffic Optimisation (ALTO) Protocol, ITU-T Draft Rec.

G.QoE-Gaming, QoE factors in gaming applications, ITU-T Draft Rec. G.MFWT, Measurement framework for web-site traffic characteristics, etc.) for emerging applications, but it is still necessary to put more effort to find the right KPI target values in a way to fulfil popular/emerging application-specific QoS requirements.

Large-scale measurement interoperability

It needs to be a common way to collect and understand the results of tests across different devices to enable correlation and comparison between any network or service parameters. In addition, it is critical that the test points are accurately defined and authenticated. The transport of Controller to Measurement Agent and MA to Collector traffic must be protected both in-flight and such that each entity is known and trusted to each other.

Mobile networks

While the major SDOs published or are publishing standards related to wireless and mobile network measurement, there are no clear standards in place yet which can support network operators in selecting QoS extension for LTE services and applications. Particularly, current mobile services are provided in multi-vendor environment, and it requires continuous attention to build common parameters which can be directly compared from different vendors. Especially for mobile networks, KPI and key quality indicators should be applicable across all vendors and operators.

New measurement area for emerging network technologies

It is not much identified on standardisation activities related to network measurement of emerging networks. Standardisation for the emerging networks should be encouraged.

Other activities

In the first workshop of the study, the need of interoperable Layer 2 measurement is identified. At the best knowledge of the current surveying, no related standardisation activity is observed.

Other issues

There are wide set of telecommunication industries of QoS related performance such as network operators, service providers, network equipment manufacturers. However, standardisation actors are somewhat limited. It is important to bring industry leaders to the international SDOs and use their resources for developing globally recognised standards for building interoperable measurement systems. The problem is that standardisation process is very long and return of investment on standardisation in global SDOs cannot be easily seen. Thus, there is often no visible incentive on business for the short term.

The European Commission's official Journal[14] containing norms and specifications for networks and services in electronic communication that gives recommended quality

[14] COMMISSION DECISION of 11/XII/2006 List of standards and/or specifications for electronic communications networks, services and associated facilities and services, replacing all previous versions (Official Journal EU L 86/11 of 27.03.2007) – Chapter VII

parameters from selected standards is one method to encourage development and use of standards. It maybe helps to set up some regular collaboration meeting between policy groups and industry leaders (or industry forum leaders) for more involvement from industry stakeholders on the related standardisation activities.

Appendix A. Overview of the ongoing EC projects related to traffic measurement

A.1 Leone Project

1. Problem statement

Although there are many "measurements" of the Internet today, they have been developed, deployed and operate independently, so measurement results are hard to compare and systems are hard to scale and integrate into existing network management visualisation tools.

2. Leone Framework

2.1. Assumptions

- A measurement system must be under the direction of a single organisation.
- Each Measurement Agent (MA) is only associated with a single Controller at any point in time.
- Only MAs - and not measurement peers - can initiate measurement tasks and communicate with Controllers and Collectors.

2.2. Leone measurement

- Protocol: HTTP with information encoded in JSON.
- Architecture: The Leone Framework has three basic elements: Measurement Agents (MAs), Controllers and Collectors. The relationship and corresponding protocols are illustrated in the Figure below.

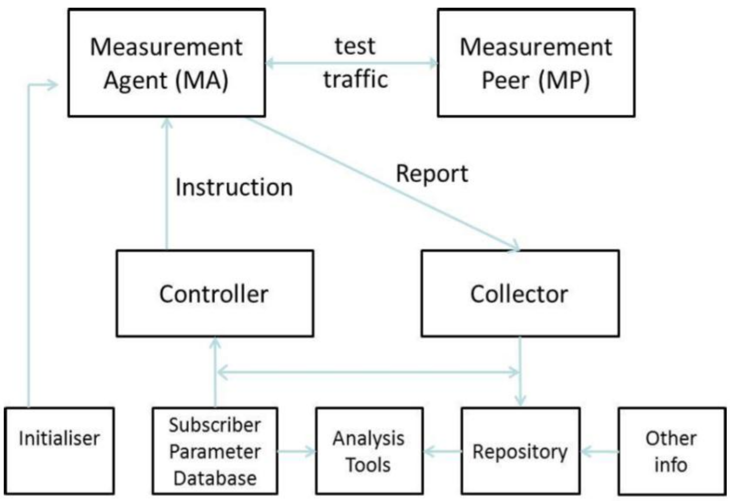


Figure A1: The Leone Framework proposed to IETF

- Registry:
  - A public registry of well-defined metrics (e.g. UDP packet latency).
    - The problem is that the metric definition often leaves too many degrees of freedom for the actual implementation – for example it does not define whether the packets are TCP, UDP, ICMP or something else.
  - Leone creates a registry of measurement points and path components:
    - To provide an unambiguous way to describe the scope of the path over which a measurement is made, since general terms like “end-to-end” are open to several interpretations,
  - Leone metrics focus on user perception and parameters that directly influence network properties.
- Subscriber Parameter Database (SPD)
  - This information is important for the analysis of the measurement data. For example, a regulator may want to compare the measured speed with the rate in the subscriber’s broadband contract, whilst an ISP may need to know the subscriber’s modem type, local aggregation node and exchange, in order to determine which other subscribers may be affected by a fault.
  - Data could include:
    - Subscriber information such as product, usage caps, traffic management policy and the subscriber’s time zone.
    - Network information such as access technology, line length, equipment type, exchange id and geo-location (especially for mobiles).
    - Network status information such as a DSL modem’s actual rate, line errors, interleaving and network utilisation.
  - Open issues
    - A scenario of an ISP-run measurement system reports results to a third party such as a regulator: it has been claimed that data privacy considerations may be easier if only the MA (and not the ISP) sends subscriber information.
    - Whether the interfaces with the SPD should be standardised.

3. Comparisons of Leone framework with existing measurement systems

3.1 Broadband Forum

- The concept from the Broadband Forum is similar to the way that Leone proposes that an MA should communicate with the Initialiser and Controller.
- The testing capabilities are currently limited to an on-demand upload/download throughput test and the configuration of the device to respond to network ping tests in order to measure network latency.
- Using the Auto Configuration Server to manage a test schedule and to continuously demand single tests across large numbers of devices is not scalable.

- Performing tests from the network towards the CPE will be hampered by NATs and firewalls, where the MA is not implemented on the Home Gateway. Furthermore, there is no method to ensure that tests are only run in the absence of user traffic.

3.2 SamKnows

- The Leone target architecture extends the current SamKnows capabilities by assuming that MAs can be embedded in multiple devices, and that Controllers use a single standard way of controlling the tests and collecting results.
- The Leone architecture also addresses one of the main implementation barriers of using SamKnows which is the accurate enhancement of measurement data with device, line or product information through the definition of the SPD.
- Leone envisages a more extensive and varied deployment of MAs than in the current SamKnows deployments. Critical extensions are to authenticate the line the MA is using and to define the network path that is being measured.

3.3 BISmark

- BISmark probes run similar active tests on a per-user schedule and report back to a central server along with capturing several passive measurements such as application throughput and usage.
- The inclusion of passive measurements means that implications for user privacy are greater,
- These concerns also drive the Leone architecture away from allowing multiple Controllers, enabling the MA to have a single end-user agreement for the testing that takes place including the purpose of the data collection.

3.4 Atlas

The Atlas features are:

- User-defined measurements are deployed on a limited number of probes.
- The bandwidth used is limited.
- A probe has no access to user traffic.
- A user has agreed to the measurement probe being placed in their own network and is aware of the arrangements for other users.

Leone assumes that MAs will be deployed on devices that do have access to user traffic and that can use the full capabilities of the network when required.

In addition, the heterogeneous nature of the MAs means that the test schedule must be carefully managed and tested for different types of device to ensure that the device and line are capable of operating the test schedule successfully.

These reasons, combined with the issues of data protection, push Leone towards having a single Controller responsible for any particular MA.

4. Leone´s analysis of use cases

4.1 ISP use case

- Purpose:

- Understanding the quality experienced by customers,
- Understanding the impact and operation of new devices and technology,
- Design and planning – proactive activities to improve the network,
- Identifying, isolating and fixing network problems,
- Benchmarking and competitor insight.

- Existing capabilities and shortcomings
  - In order to obtain reliable benchmarks, some ISPs use vendor-provided hardware measurement platforms that connect directly to the home gateway.
  - While the test capabilities of such probes are good, they are too expensive to deploy on a mass scale to enable the detailed understanding of network performance (e.g. to the granularity of a single backhaul or single user line).
  - There is no easy way to operate similar tests on other devices (e.g. set top box) or to manage application level tests (such as IPTV) using the same control and reporting framework.
  - ISPs also use speed and other diagnostic tests from user owned devices, but they are not able to perform continuous testing and the uncontrolled device and home network means that results are not comparable.

4.2 Regulators

- Purpose
  - Development and enforcement of broadband policies:
    - The need to produce datasets that are able to compare multiple broadband providers, diverse technical solutions, geographic and regional distributions, and marketed and provisioned levels and combinations of broadband services.
    - Requiring that the measurement approaches meet a high level of verifiability, accuracy and fairness to support valid and meaningful comparisons of broadband performance.

4.3 Comparison

	Regulators	**Operators**
Performance metrics	Interested in performance metrics. Would like standardised metrics (more important for regulators than operators).	Interested in performance metrics. Would like standardised metrics.
Sampling	An average is Required across a representative sample of broadband customers (per operator, per type of broadband contract).	Require the measuring of individual lines with a problem.
Timeliness	To know the (averaged) performance last quarter.	To know the performance and to instruct a test to be run at the

		moment.
Scheduling	Require the running of scheduled tests ('measure download rate every hour').	Require run one-off tests; perhaps also the result of one test would trigger the operator to run a specific follow-up test.
Pre-processing	Require standard ways of processing the collected data, to remove outlier measurements and aggregate results.	Not an important factor for an operator.
Historic data	Require the tracking of how the (averaged) performance of each operator changes on (say) a quarterly basis.	Require a detailed, recent historic data (e.g. a customer with an intermittent fault over the last week).
Scope	To measure the performance of access lines.	To understand the performance of the home (or enterprise) network and of the end-to-end service, i.e. including backbone, core, peering and transit, CDNs and application /content servers.
Control of testing and reporting	'Control' will be via negotiation with its contractor.	Require detailed control
Politics	Taking into account of government targets (e.g. UK government: "Our ambition (by 2015) is to provide superfast broadband (24Mbps) to at least 90% of premises in the UK and to provide universal access to standard broadband with a speed of at least 2Mbps."). This may affect the metrics the regulator wants to measure and certainly affects how they interpret results.	Focused on winning market share.

4.4 End Users

- An end-user may wish to perform diagnostics prior to calling their ISP to report a problem.
- Hence, the end-user could connect a MA to different points of their home network and trigger manual tests. Different attachment points could include their in-home 802.11 network or an Ethernet port on the back of their broadband modem.
- An OTT or ISP service provider may deploy a MA within their service platform to provide the end-user a capability to diagnose service issues. For instance, a video streaming service may include a manually initiated MA within their platform that has the Controller and Collector predefined. The end-user could initiate performance tests manually, with results forwarded to both the provider and the end-user via other means, such as UI, email, etc.

4.5 Findings from the use cases

- A clear need to deploy a single coherent measurement capability across a wide number of heterogeneous devices both in ISPs´ own networks and in the home environment.
- A need for a way to demand or schedule the tests and critically ensure that such tests do not affect each other and are not affected by user traffic.
- A need for a common way to collect and understand the results of such tests across different devices to enable correlation and comparison between any network or service parameters.
- It is critical that the test points are accurately defined and authenticated.
- The transport of Controller to MA and MA to Collector traffic must be protected, such that each entity is known and trusted to each other.

5. Summary of Leone Solutions

Leone works on the solutions for large-scale broadband active measurement for overcoming the current shortcomings that measurement results are hard to compare and systems are hard to scale and integrate into existing network management. The project aims to improve the current shortcomings of large-scale broadband measurement by:

- An extensible architecture framework.
- Definition of new metrics.
- Introducing new analysis methods.
- Improving management tools of network/service providers.
- Testing the proposed framework, protocols, and tools by lab trials and field trials.

Leone deploys hardware probes at the premises of the customers of different ISPs that carry out measurements at configurable intervals and report the results back to a collection infrastructure.

- SamKnows hardware probes – Leone uses only fixed-line broadband probes (“Whitebox” probes).
- Current SamKnows probes implement a number of metrics:
    - TCP download throughput, TCP upload throughput, Latency, packet loss (UDP), Latency-under-Load, Loss-under-Load, Jitter (upstream, downstream), availability, ICMP ping, traceroute, DNS (RTT and failure rate), and application tests for web browsing, plain TCP streaming, and BitTorrent throughput.

The following are the key improvements provided by Leone:

1) Extensible Framework

- The Leone framework is designed to be extensible and is able to support the evolution of applications for deploying and running measurements; for collecting, analysing, and sharing result data; for acting on those results; and for managing probes and the infrastructure.

- Leone is making standardising effort on a measurement system Information Model, control protocol and report protocol, in order to increase interoperability and comparison of measured data from different systems. It is a major contributors in IETF LMAP and Broadband Forum

2) Data Acquisition and Analysis

a) Definition of new metrics

- Focus on user perception and parameters that directly influence network properties.
- At the Network Layer, Leone characterises end-to-end delay and loss as well as path stability.
  - Newly added metrics by Leone are: PNPM Ping, Multicast – MCASTMON, SamKnows BT Vision Multicast Test, Traceroute – capturing the forwarding path.
- At the Transport Layer, Leone determines connection establishment time and transport protocol performance.
  - Newly added metrics by Leone are: TCP connection establishment time, Traffic bursts (UDPBURST), MPTCP benefit metrics.
- At the Application Layer, Leone emulates the operation of different web-based services and measures the impact of DNS and CDNs, web page load times, and video streaming performance.
  - Newly added metrics by Leone are: DNS performance - DNSPERF, Web performance - WEBPERF, CDN performance for Web pages - WEBCDNPERF, Performance of RTP video streaming, Performance of YouTube, Performance of adaptive HTTP streaming, Performance of adaptive HTTP streaming through CDN.

b) Building methodologies to analyse the measurement results.

- Output data is collected and used as input to the multi-dimensional analyses.
- One-dimensional analysis: 1) BGP Visibility scanning, 2) IPv4/IPv6, TCP, and Happy Eyeballs.
- Multi-dimensional analysis: 1) Correlating routing changes with delay variations, 2) Reliability.
- The set of metrics and analysis tools are expected to change and/or expand over time based upon the analytical insights and the operational experience Leone gains.

3) Providing tools for network/service operators

- Leone provides network or service operators with tools to maximise the utility of the information collected through large-scale measurement platforms.
- Visualisation prototype tools: developed techniques to effectively visualise traffic flows and routing information as reported by different measurement data sources.
- Extensions of management systems with additional abilities such as anomaly detection and alarm triggering.

- Automated repair techniques.

4) Integrations and Trials

- Leone runs a trial of the tools, which involves about 100 Leone probes distributed to real end-users, and integrated with the existing SamKnows global platform.
- The main purpose of the trial is to validate the measurement architecture and demonstrate the feasibility of the Leone network management framework.
- Requirements:
  - Probes located in a smaller geographic region, connected via a single ISP with different access network characteristics (e.g. different flavours of DSL),
  - Coverage of a larger number of different ISPs,
  - Coverage of a larger geographic area,
  - IPv4 and IPv6 access to Internet services,
  - IP multicast support,
  - Probes should be able to utilise multiple access technologies concurrently.
- Lab test with 30 probes distributed in partners - three use cases
  - Use case 1: install new test,
  - Use case 2: delete test,
  - Use case 3: test scheduling.
- Field trial
  - Managed by two service providers (TI, BT),
  - Testing the new measurement and analysis techniques that Leone is developing in the project,
  - At the moment, 22 persons are recruited in TI who have connections of ADSL 20Mbps, ADSM 7Mbps, ADSL 640Kbps, and use services of Internet only (15), VoIP (2), IPTV (5), Smart TV (3), Over the top premium video (2). The BT trial has a similar amount of participants,
  - Handles privacy and data protection alleviating privacy risks from Leone´s large-scale measurement trials followed the guidelines of EU´s data protection and directive, and each trialist has to sign the trial “End User Agreement”.

A.2 mPlane Project

1. Problem statement

In the Internet different technologies are combined to offer a plethora of services, and we sorely miss the technology to understand what is happening in the network and thus to optimise its performance and utilisation, especially when something goes wrong. Thus, we need an intelligent system that collects, analyses and provides visibility to support better management: an oracle that provides answers.

Since so far most effort has been dedicated to the probes, repository and “supervisor”; stated in other terms these components provide the mean for the reasoner to operate; at the beginning the framework aimed at concentrating all data post-processing in this entity the work by cases lead though to a focus in terms of cases (data) instead of underlying algorithmic commonality. Note from a machine learning perspective, the mPlane reasoner shall be seen as the oracle involved in active learning. When a massive amount of input and unlabelled data is available and labelling is costly, active learning can perform better (in particular faster with hopefully limited biased sampling) with less training if it is allowed to choose the dataset examples from which it learns. The (active) learner may pose "queries" usually in the form of unlabelled data instances to be labelled by an "oracle" (e.g., a human annotator, expert system) that already understands the nature of the problem. This approach is well-motivated for applications, where unlabelled data may be abundant or easy to come by, but training labels are difficult, time-consuming, or expensive to obtain.

Another element to account for (again from a machine learning perspective) is that the reasoner has not yet extended its modules so as to enable transfer learning among different algorithms and learning tasks (inductive transfer learning) and target domains (transductive transfer learning). Transfer learning is motivated by the fact that one usually can intelligently apply knowledge learned previously to solve new problems faster or with better solutions. The fundamental motivation for Transfer learning stems from the need for lifelong machine learning methods that retain and reuse previously learned knowledge.

2. mPlane architecture

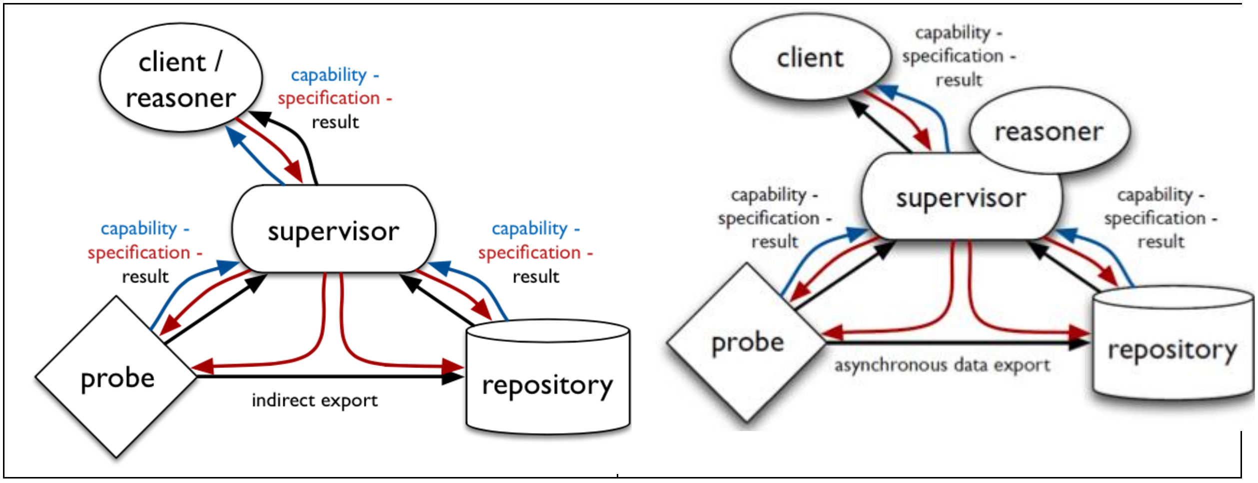


Figure A2: General arrangement of entities in mPlane architecture

In the mPlane architecture, anything that publishes capabilities and makes services available according to them using the protocol described in this document is a component and anything that uses those capabilities is a client.

The measurement components can be divided into two categories: probes and repositories. Roughly, probes perform measurements, and repositories provide access to stored measurements, analysis of stored measurements, or other access to related external data sources.

The supervisor is responsible for collecting capabilities from a set of components, and providing capabilities based on these to its clients. Application- specific algorithms at the supervisor aggregate the lower-level capabilities provided by these components into higher-level capabilities exposed to its clients. The supervisor basically interfaces between the measurement data processing and the measurement data collection (probes and databases)

Within an mPlane domain, a special client known as a reasoner may control automated or semi- automated iteration of measurements, working with a supervisor to iteratively run measurements using a set of components to perform root cause analysis. While the reasoner is key to the mPlane project, it is architecturally merely another client, though it will often be colocated with a supervisor for implementation convenience.

The following Figure shows workflow among the mPlane components:

- Capabilities define the tasks a component can perform.
- Specifications consist of a description of which measurement have to be performed, how, and when.
- Components announce their capabilities when registering to the supervisor.

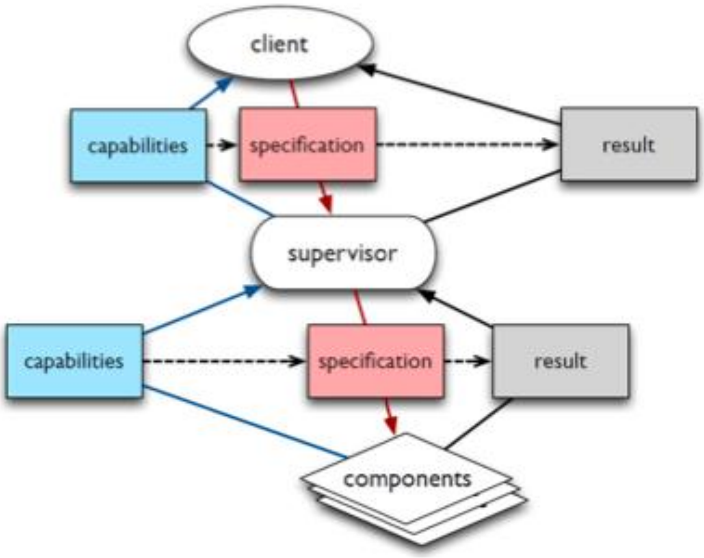


Figure A3: mPlane workflow

For inter-domain measurement, each domain collects and owns its measurements. Different mPlanes are under the control of different players (ISP, CDN, etc.). Multi-domain measurements are handled as communications among supervisors. The following Figure shows an example of mPlane inter-domain measurement.

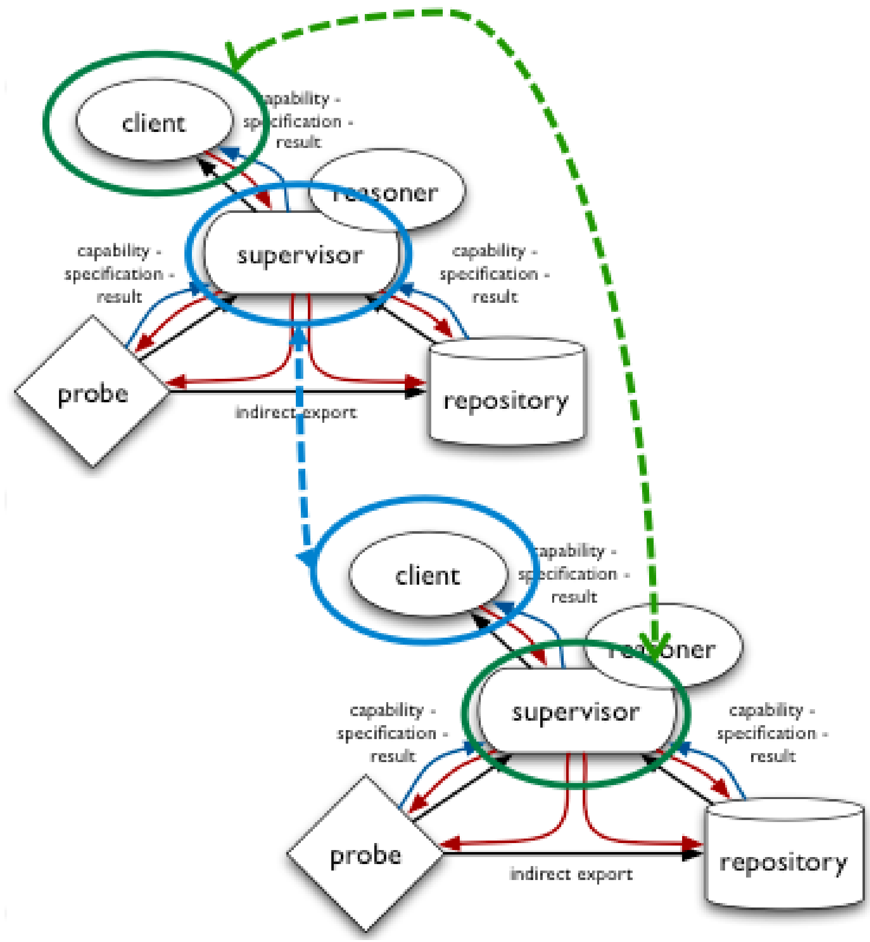


Figure A4: mPlane inter-domain measurement

In summary, mPlane aims at simplifying network monitoring practices by:

- Reasoning and Analysis focused on iterative measurements - Troubleshooting support
- Open source release of software - Tstat, Blockmon, QoF, tracebox
- Maximum reuse of existing measurement tools - first software libraries available at the mPlane website

Appendix B. UK Ofcom – "Easy to understand" labelling

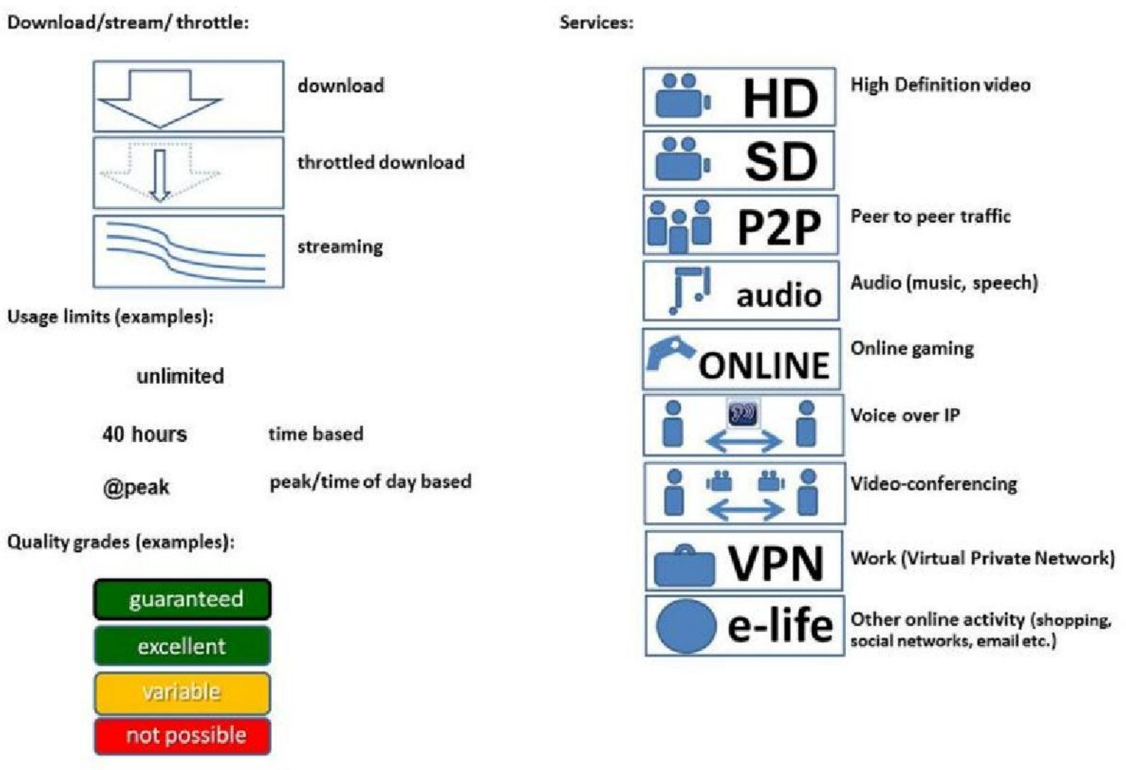


Figure B1: Labelling proposed by UK Ofcom

The following figure shows illustrative QoE summaries for 3 hypothetical ISPs.

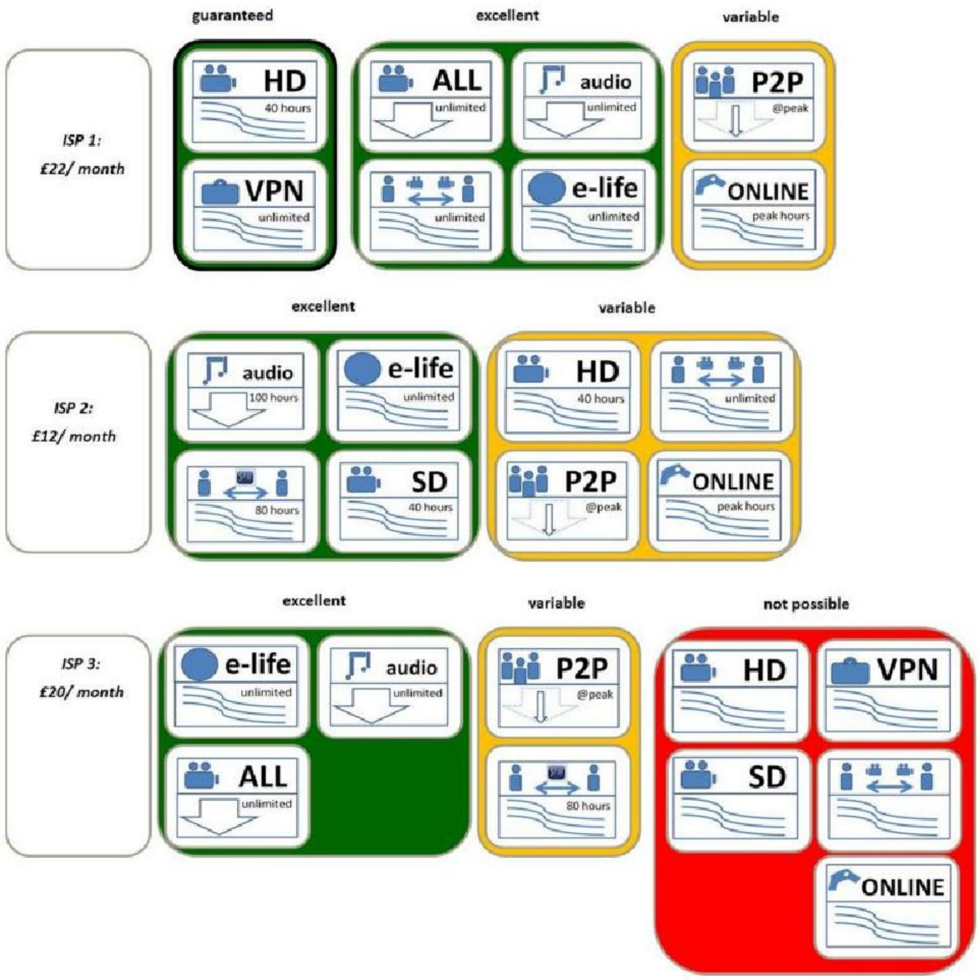


Figure B2: Illustrative QoE summaries for 3 hypothetical ISPs

The following Figure shows illustrative transparent traffic management status representations.

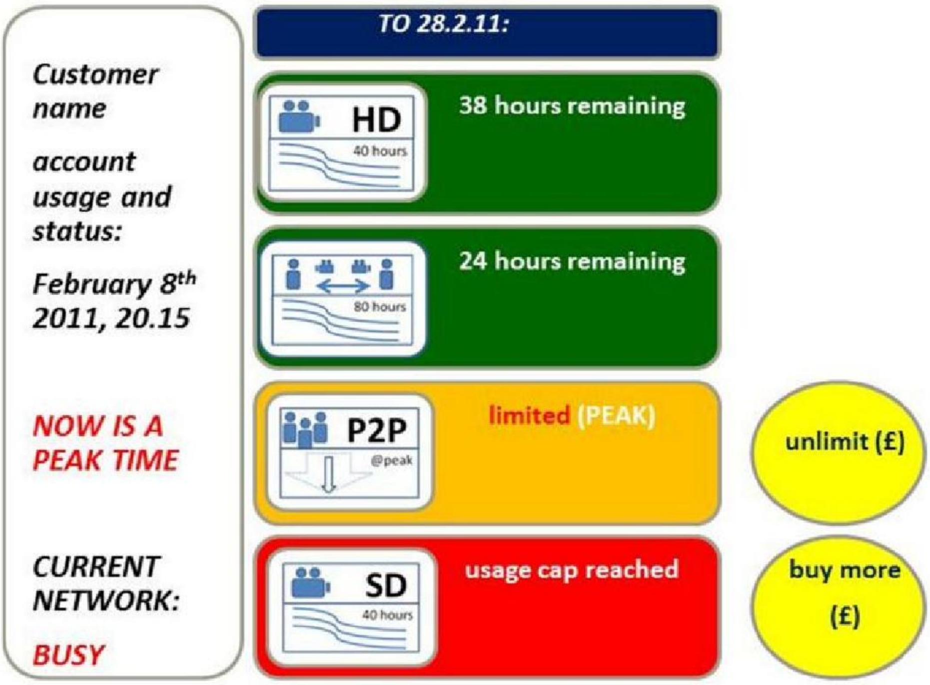


Figure B3: Illustrative transparent traffic management status representations