QUALITY OF SERVICE IN TELECOMMUNICATION NETWORKS

Fabricio Carvalho de Gouveia and **Thomas Magedanz**
Technical University of Berlin, Franklinstr., Berlin, Germany

Keywords: Quality of Service, Telecommunications, Convergence, Congestion Control, Service Level Agreements.

Contents

Summary

Wireless mobile Internet is migrating toward an integrated system of Internet and telecommunications technologies in order to fulfill the future telecommunications requirement: ubiquitous communication, where mobile users move freely almost anywhere and communicate with anyone, anytime with any device using the best service available. This demands a rapid progress in telecommunications and the Internet technologies. This topic addresses state-of-the-art research in QoS management issues in Telecommunications. It presents problems and actual solutions in the literature, in an attempt to ensure necessary levels of service and performance to critical multimedia and data applications, including service guarantees.

1. Introduction

Mobile Technology Evolution started with analog cellular technologies e.g. Advanced Mobile Phone Service (AMPS) and Total Access Communications System (TACS). The Second Generation (2G) leaded by the digital technology Global System for Mobile communication (GSM) re-evaluated the concept of mobile telephony, with a rapid evolution of services. Then the 2,5G has prepared its way with a Packet Switched (PS) extension of the GSM network called General Packet Radio Service (GPRS) technology, providing quality in form of speed and efficient use of network. Other technologies succeeded up to the arrival of the 3G Universal Mobile Telecommunications System (UMTS) and Code Division Multiple Access CDMA2000. Thanks to this evolution of the networks towards PS technologies, data services have experienced a huge increase in terms of data transmission capabilities and new services arrived.

Today end users just need an IP access connection, e.g. via a Wireless Local Area Network (WLAN) hotspot, a Digital Subscriber Line (DSL) connection or a GPRS / UMTS network to have access to these services. The success of Skype and other VoIP / Multimedia over IP providers in face of dropping IP connectivity prices is providing evidence for this view. With the progress of this IP-fication of networks, the competition on IP-based (telecommunication) services is growing, and we can witness a changing value chain in which connectivity charging decreases in favor for applications and content charging. This heterogeneous all IP telecommunications trends demands outstanding provision of Quality of Service (Figure 1), and motivates the provision of service experience to users, in order to fulfill their expectations motivating them to use it more and recommend it to friends.

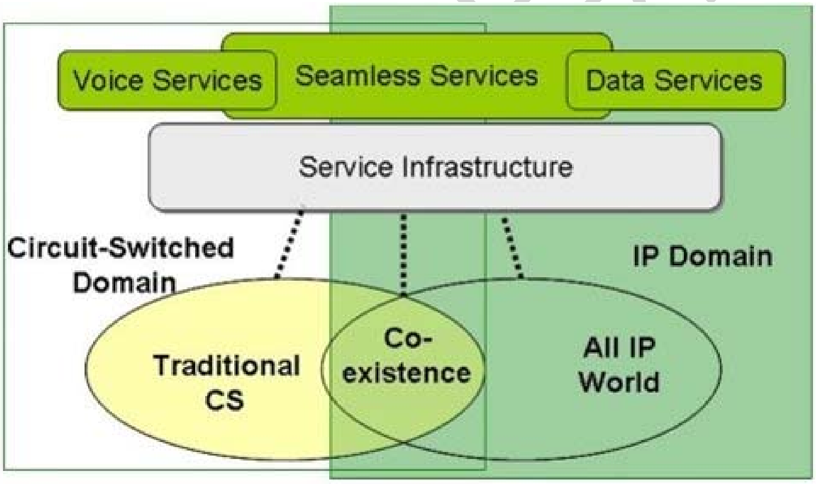

Figure 1: Convergence to all IP Telecommunication Scenario

As the number of users in the Internet and other telecommunication networks grows, it becomes very clear that real-time applications become more difficult to implement because of erratic queuing delays and packet loss. The topic QoS in telecommunication networks described here will give a broader vision of the management of QoS in this growing environment.

2. Quality of Service (QoS) Background

2.1. QoS Definition

Quality of Service (QoS) in the field of telecommunications can be defined as a set of specific requirements provided by a network to users, which are necessary in order to achieve the required functionality of an application (service). The users specify their performance requirements in form of Quality of Service parameters such as delay or packet loss (described in section 2.3), and the network commits its bandwidth making use of different QoS schemes to satisfy the request. Each service model has its own QoS parameters.

The quality of a service can be a differentiator in the business market. Its parameters and measures are necessary to provide an indication of how well a service or product is, and therefore, be an important point when selecting services offered by different vendors (or service providers). If service features or price are similar, quality becomes the differentiator for users, as well as, service providers can make use of quality to have an image of a "respected" provider.

Challenging situations that cause QoS to degrade can be summarized as follows:

- The first one is congestion, which is caused by traffic overflow (Bottlenecks);
- Delays, caused by networking equipment low performance in large loads, as well as caused by distance or retransmission of lost packets;
- Shared communication channels, where collision and large delays become common, and
- Limited bandwidth networks with poor capacity management.

2.2. Telecommunication QoS Characteristics

Telecommunication is a communication of data or voice over a distance. There exist many telecommunication technologies from the legacy networks to emerging all IP networks. To cite some examples: basic telephony services (wired and wireless), Satellite communications, Asynchronous Transfer Mode (ATM), Ethernet and 802.1 networks, IP-routed networks (e.g. Internet) and so on.

The current global Internet service is based on best-effort service. This service does not guarantee anything, even delivering the IP packets within the network. Considering a packet sent to the Internet for delivery to a destination host, the network does not guarantee any specified delivery time, delivery speed, the available bandwidth, or even if the packet will be dropped if it faces congestion. Delay is not a problem if we consider delivering of an email message, where seconds or minutes will have a small impact on the end user. But if the transmission delay in a voice-over-IP (VoIP) call is large, or delays vary too much, or too many packets are lost, the quality will become unacceptable. The Public Switched Telephone Network (PSTN) that is referred to as the basic telephony service is considered a reference for voice quality, and fulfills all expectations of users. This network is Circuit Switched (CS) based and for this reason, has a dedicated circuit (or channel) reserved for the call. It provides very good quality on one hand, but is less efficient than IP networks on the other, which use the available bandwidth more efficiently. Of course the network presents sometimes problems to deliver Quality of Service, which is referred to noise on the circuit or appropriate loudness levels.

It is very important to note that the QoS is specific to the service. Each service may be expressed by a set of parameters that are specific to it. Jitter is a parameter that is applied to packet switched networks, Cell Loss Ratio (CLR) to Asynchronous Transfer Mode (ATM) and these parameters would be meaningless in a circuit switched analog network. Other characteristic is that QoS is an end-to-end issue. This means that all entities in the path between the parties are concerned to make the service possible and all the segments are involved in the process of QoS guarantee.

2.3. QoS Parameters

To provide and sustain QoS, resource management must be QoS-driven. To allocate resources, the resource management system must consider different parameters:

- resource availability;
- resource control policies, including Service Level Agreements (SLA);
- QoS requirements of applications, which are quantified by QoS parameters (e.g. Jitter, Delay, Packet loss).

To keep track if the contracted QoS are being met, the QoS parameters must be monitored and resources reallocated in response to system anomalies. Application layer must ensure that the required QoS parameters can be satisfied (through QoS negotiation signaling) prior to the reservation of resources. If negotiation is agreed, the session starts. If a change of state happens i.e. degradation in the QoS, and the resource manager cannot make resource adjustments to compensate, the application can either adapt to the new level of QoS or degrade to a reduced level of service. The measurement of QoS is based on parameters like delay, jitter, packet loss, throughput and many others, depending on the application and management scheme.

The general QoS parameters most considered in packet switched networks are summarized below:

Delay: this parameter is intrinsic to communications, since the end points are distant and the information will consume some time to reach the other side. Delay is also referred as to latency. Delay time can be increased if the if packets face long queues in the network (congestion), or crosses a less direct route to avoid congestion. The delay can be measured either one-way (the total time from the source that sends a packet to the destination that will receive it), or round-trip (the one-way latency from source to destination plus the one-way latency from the destination back to the source). Round-trip delay is used frequently, because it can be measured from a single point using the “Ping” command. The round trip delay is a relatively accurate way of measuring delay, because it excludes the amount of time that a destination system spends processing the packet. The “Ping” command performs no packet processing. It only sends a response back when it receives a packet. In the case for having a more accurate delay measure, then it is needed the measure in both points of the network. The final result is the minimum delay time possible in that link for sending a packet from a source to the destination.

In transmission via satellite networks the delay is even in the best situation high (about

260ms one-way), because the long distance increases the number of electronic components and processing time built into the equipment at each end and in the satellite itself that will add delay to the transmission total time, despite the signals in the space traveling at light speed (electromagnetic waves). If a telephone call using satellites is taking place, this equipment induced delay will affect directly the quality of the call. For delays greater than 250ms there is the problem of talk overlap (when a speaker repeats his sentence when the other party is responding).

Jitter: Jitter is the delay variation and is introduced by the variable transmission of delay of the packets over the network. This can occur because of routers' internal queues behavior in certain circumstances (e.g. flow congestion), routing changes, etc. This parameter can seriously affect the quality of streaming audio and/or video. To handle jitter, it is needed to collect packets and hold them long enough until the slowest packets arrive in time, rearranging them to be played in the correct sequence. Jitter buffers can be observing when using video or audio streaming websites (e.g. YouTube) and are used to counter jitter introduced by the internet so that a continuous playout of the media transmitted over the network can be possible. When clicking in a link to play the video, buffering starts before the media stream actually does. This procedure causes additional delay, but is necessary in the case of jitter sensitive applications.

Packet Loss: happens when one or more packets of data being transported across the internet or a computer network fail to reach their destination. Wireless and IP networks cannot provide a guarantee that packets will be delivered at all, and will fail to deliver (drop) some packets if they arrive when their buffers are already full. This loss of packets can be caused by other factors like signal degradation, high loads on network links, packets that are corrupted being discarded or defect in network elements. Wireless networks have higher probability of loss that is introduced by the air interface (e.g. interference caused by other systems, multiple obstacles (buildings, environment) in the path, multipath fading, etc.). Some transport protocols such as Transfer Control Protocol (TCP) make delivery control by receiving acknowledgements of packet receipt from the receiver. If packets are lost during transfer, TCP will automatically resend the segments which were not acknowledged at the cost of decreasing the overall throughput of the connection.

Throughput: Throughput is the amount of data which a network or entity sends or receives data, or the amount of data processed in one determined time space. It has as basic units of measures the bits per second (bit/s or bps). The throughput can be lower than the input tax due to losses and delays in the system. Throughput is a good measure of the channel capacity of a communications link. A good example of throughput measure is performed by a bandwidth meter (which is used for measuring the real transfer rate that a DSL connection has). The Bandwidth Meter estimates the current throughput of a DSL connection by calculating the rate at which a test file is delivered to the computer for a particular Server. Paying for a 1000kbps DSL connection means that the bandwidth available for this connection is up to this 1000kbps in the access network. A bandwidth meter tool reliably measures the speed with which a user can download information from particular servers. However, it may not reflect users' experience downloading particular pages on the Internet. There are many factors which affect the rate at which webpages and files download, regardless of the connection type.

These factors include:

- The load on the server a user is accessing.
- The remote server's "distance" from the measured system (in terms of network hops (routers)).
- The speed of the computer.
- The number of programs running on the computer.
- The configuration of the network.

Service Level Agreements (SLA): SLA is a written agreement defining business goals between two parties, for instance it can be used between users and the network access operators, between network operator and service provider, between network operators, and between service providers. The SLA defines in a customer friendly manner, tangible and easy constraints to be observed and proven, the expectations for all the involved parties in the delivery of the service. SLAs are used to achieve QoS end-to-end, because it defines the way the services are to be provided, used, what is permitted, price for the services offered, Penalties for not meeting service requirements, added options of support, incentive prizes for overcoming service levels, etc.
There are three types of SLAs:

- Customer SLA: describes the services (product offering) supplied to the customer. This is a contract between the customer and the operator and is written in terms that the customer can understand, and is on the end-to-end delivery of the components of the product.
- Supplier SLA: describes how the various performance objectives will be met for the resource and service components that are provided by a third party. It is usually derived from the contractual agreement between the service provider and the operator.
- Internal SLA: it focuses on resources and has the purpose of managing the set of services and service components that are defined by one of the above two types of SLAs. It is subject to agreements between organizational functions within the operator's business.

Before entering the management schemes some examples will be described on how these parameters manifest themselves to the user. If we consider that downloading a file (using a phone or other equipment) will take 5 minutes, the download is not going to be oversensitive to delay. This is due to the fact that delay is in the order of milliseconds (normally in the range 100 to 1000ms), and incrementing this amount of time in the total download time (in this example 5 minutes) is imperceptible to the user. Jitter also will not have an impact (considering moderate levels of jitter). The packets may be received in a different order than they were dispatched, but their order will be corrected by the destination equipment. This process is transparent to the user and has no effect on the download, which will be completed only when the last packet arrives. Packet loss has a big impact on the download application since they need to be resent for maintaining data integrity. A congested network will drop more packets and even contributing more to the problem where the resend of packets will cause more congestion. Therefore download is known as a packet loss sensitive application.

Other applications like video conference or voice, on the other hand, have different characteristics. Delay is very important as said before and must be below 250ms to be acceptable. Jitter will cause jerky pictures or muffled sound, since for the interpretation of video and audio the data should be in the right order. Packet loss, on the other hand, could have a low impact. Considering that in a video transmission 99% of the packets are received quickly and in the right order, the data should be enough to reconstruct the video and audio streams without notice from the user. Then real time applications like video conference and online gaming are delay and jitter sensitive. These examples prove that each kind of application demands different QoS metrics for having the user satisfied with the service quality.

-
-
-

Bibliography

Braden D., (1994) *Integrated Services in the Internet Architecture:* IETF RFC 1633 [Describes in more detail the IntServ architecture for the Internet]

Braden D (1997) *Resource reservation protocol* IETF RFC 2205 [Shows the functionality of the RSVP protocol]

Camarillo G., (2004) *The 3G IP Multimedia Subsystem, Merging the Internet and the Cellular Worlds*, 480 pp. John Willey & Sons Ltd West Sussex, England [It describes and motivates the merging of Internet and cellular networks]

Chen Jyh-C, Zhang T, (2004) *IP-Based Next-Generation Wireless Networks* 413 pp. John Wiley & Sons, Hobokey, New Jersey, USA [This presents QoS management, architectures and protocols in wireless networks].

EU IST BRAIN Project www.ist-brain.org [Describes the 4G network architecture]

Grossman D., (2002) *New terminology and clarifications for DiffServ*. IETF RFC 3260 [Describes in more detail the DiffServ architecture]

ITU-T recommendation (2002) Traffic Control and Congestion Control in IP-based Networks [traffic and congestion control standard]

Oodan A., Ward K., Savolaine C., Daneshmand M., Hoath P. (2003). *Telecommunications Quality of Service Management*, 596 pp. Bodmin, Cornwall, UK: [Presents a very broad analysis of QoS in Telecommunications].

Priggouris G.,Hadjiefthymiades S., Merakos L. (2000) Supporting QoS in the General Packet Radio service pp. 8-17, IEEE Network [Describes the QoS management In GPRS]

Salsano S. (2002) QoS Control by Means of COPS to Support SIP-based Applications IEEE Network [Shows the Policy based architecture application for call admission control procedures]

Savolaine C. G., (2000) Quality of Service Issues for Voice in an IP Environment, IEEE Globecom 200 Conference. [Analyses the performance QoS parameters in VoIP systems]

Third Generation Partnership Project (3GPP) – www.3gpp.org [Contain the standards from UMTS/GPRS and GSM]

Third Generation Partnership Project (3GPP) (2005) *End to end quality of service (QoS) signalling flows*. 3GPP TS 29.208 Version 6.4.0 [Presents the QoS flow reservation in IMS]

Third Generation Partnership Project 2 (3GPP2) (2000) *Wireless IP architecture based on IETF protocols*. 3GPP2 P.R0001, Version 1.0.0 [The protocols of IETF are used in the CDMA technology]

Valdar A. R., (2001) Next Generation Networks, International Conference on InfoCommunication Trends, Budapest. [Describes the issues on NGNs]

Biographical Sketches

Fabricio Carvalho de Gouveia received his graduation in Electrical Engineering from the University Regional of Blumenau (Brazil) in 2000 and his M.S. in Telecommunications from the Federal University from Parana (Brazil) in 2003. He is employed as a Junior Scientist at the research center for Next Generation Network Infrastructure at FOKUS, where he is working towards his Ph.D in the Field of Next Generation Networks (NGN). His major research topics are QoS provisioning for NGN. He has hold speeches at technical conferences related to his research topics and published research papers in his fields of expertise.

Dr. Thomas Magedanz received his M.S. and his Ph.D. in computer sciences from the Technical University of Berlin, Germany in 1988 and 1993, respectively. Thomas Magedanz is head of the next generation network infrastructures (NGNI) competence centre at the Fraunhofer Institute FOKUS, Germany. Furthermore, he is professor at the Technical University of Berlin in the field of next generation networks. He is an internationally recognized expert in the areas of Intelligent Networks (IN) and Open Service Delivery Platforms. He is a member of the IEEE, editorial board member of several journals, and the author of more than 200 technical papers/articles. He is the author of two books on IN standards and IN Evolution and is a regularly invited tutorial speaker at major international telecom events and conferences. He has participated in numerous national and European R&D projects related to Next Generation Intelligent Networks and Open Service Delivery Platforms.