White Paper

Measurement and Analysis of End-to-end Service Quality of 3G Networks and Services

Editor	Bharat Bhushan (Fraunhofer FOKUS)
Contributors	Fraunhofer FOKUS
Project	AlbatrOSS – an EU IST FP5 Project - www.ist-albatross.org
Date	February 2004
Distribution	Public
Contact	bhushan@fokus.fhg.de

Abstract

The QoS (quality of service) of a service product perceived by the end-user depends on the performance achieved by the components that constitute that product as well as performance of the network and service resources on which those components run. These resources may be spread over many domains and owned by different network and service providers. In this scenario, the end-to-end QoS will depend on overall performance achieved by all the network and service components. Measuring performance and analysing it to ascertain service product QoS is a challenging task and the subject of this paper.

Much research work to date has overlooked the end-to-end service quality issue while concentrating on the measurement and analysis of the performance of IP-based networks. The work that does address service quality issues does not relate to the work of standards bodies such as 3GPP (3rd Generation Partnership Project) and TMForum, IETF, and ITU-T, which are developing a common framework for 3G service quality measurements. This White Paper tackles this challenge by investigating a mapping between QoS perceived by end-user onto performance of individual networks and services in different domains, and relating this work to the work of standards bodies. The key problem is defined and the mechanism to measure network and service performance and transform it into end-to-end service quality is developed. This paper also covers how the measurement mechanism can be applied to SLA violation monitoring.

Keywords: QoS Management and Assurance, Wireless Broadband Networks, Next Generation Operations Support Systems, Networks and Systems Performance Monitoring

Table of Contents

1 Introduction ... 4

2 Related Work ... 4

3 Overview of the Work Presented ... 5

3.1 Problem Definition ... 5

3.2 Approach Adopted and Results ... 6

4 QoS Classes and End-to-End QoS Architecture ... 7

4.1 End-to-end QoS Architecture ... 8

4.2 QoS Delivery and Measurement Scenario ... 9

5 Information Model ... 9

5.1 Customer Contract View ... 12

5.1.1 Party ... 12

5.1.2 Party Role ... 12

5.1.3 ProductSpecification ... 12

5.1.4 ServiceSpecification ... 12

5.1.5 ServiceLevelAgreement ... 13

5.1.6 ServiceLevelSpecification ... 13

5.1.7 ServiceLevelSpecObjective ... 13

5.1.8 ServiceLevelSpecConsequences ... 14

5.2 Service Quality View ... 14

5.2.1 KeyQualityIndicatorSLSParam ... 14

5.2.2 EndToEndAvailability ... 14

5.2.3 Accuracy ... 14

5.2.4 ServiceTransactionSpeed ... 15

5.3 Network Performance View ... 15

5.3.1 KeyPerformanceIndicatorSLSParam ... 15

5.3.2 AvailabilityKPI ... 15

5.3.3 AccuracyKPI ... 16

5.3.4 ServiceTransactionSpeed ... 16

5.4 Transformation Algorithm for End-to-End Service Availability ... 16

FOKUS

5.5 Limitation of the SID Model 17
6 OSS Functions and Component Architecture 17
6.1 OSS Functions 18
6.2 Component Architecture 18
6.2.1 Monitoring Analysis Component 19
6.2.2 Network Performance Evaluator 19
6.2.3 Report Log 19
6.2.4 SLA Manager 19
6.2.5 Supporting Components 19
6.3 Evaluation and Performance Results 20
7 Conclusions, Discussions and Future Work 20
Glossary 22
Acronyms 23
References 25

1 Introduction

The need for greater interoperability between various service providers, operators and technologies means that traditional methods of measuring service performance and quality are no longer necessarily valid in a mobile environment. QoS commitments need to be made on a peer-to-peer basis and in order to make such commitments the service providers (SPs) and operators need to know what they can commit to. They therefore require network and service performance information not only from their own domain but also from those with whom they are cooperating. It is important because usage of end-to-end services involves establishing a session from the end-user to application servers, maintaining the session and application servers within QoS parameters, and terminating the session properly when it has ended. In this scenario, challenges are posed by various segments of the end-to-end network connection that take part in providing end-to-end services; these segments include the access network, the UMTS (Universal Mobile Telecommunications System) or GPRS (General Packet Radio System) core network, the IP backbone, and the intranet connecting application servers. Each of these segments may be subject to a different SLA and QoS policy and may require a technology-specific QoS management. This White Paper defines the problems and presents a mechanism for the measurement and analysis of end-to-end service quality for 3G networks and location-based services. The mechanism also includes how the measurements can be used for SLA violation monitoring.

The rest of the paper is organized as follows. Section 2 discusses related work. Section 3 investigates the problem and describes the approach adopted to carry out the research work. Sections 4 and 5 present the end-to-end QoS architecture and information model, respectively. Section 6 presents the OSS functions required defined for the mechanism. To realise the ideas presented in this paper, OSS components have been implemented within the EU IST project AlbatrOSS. Section 6 also presents a component architecture and evaluates the work to date. Section 7 concludes the White Paper.

2 Related Work

A broad survey of related work has revealed that very many technical papers have been written on SLA, QoS and 3G networks but comparatively few address the problem being addressed in this paper. Even if they do address the problem, they do not do so in its entirety. Most relevant papers are briefly described in this section.

The work described in [6] proposes a framework for wired-wireless translation and QoS management. It uses a UMTS end-to-end QoS architecture and classes for a dynamic SLA-based QoS control. According to the proposed framework, end-to-end QoS control is achieved using a mix of SLA-based and policy-based schemes. Technical paper [18] proposes another 'framework' for SLA management. However it is difficult to see how these frameworks can be used for the management of service quality of 3G networks and services.

Paper [8] and [11] propose a QoS system architecture, a fully IP-based access network, an architecture of end-systems for a wide variety of access networks including UTRAN and IEEE Wireless LAN. The main aim of the work described in [12] is to use simulation to examine the effect of Internet traffic on QoS of GPRS networks. Subjective and objective QoS measures are derived for web-based applications running on TCP. Technical paper [13] is also an interesting paper in which good insights into the performance of wireless networks are provided. But it addresses mobility support for QoS in wireless networks at the pico- and micro-cellular environment. Paper [14] is a good tutorial on [3] and accompanying 3GPP technical specifications on the QoS of IP multimedia subsystem. Paper [15] presents a fairly comprehensive survey of QoS concepts, requirements, and techniques for a mobile distributed computing environment. Papers [16] and [17] present a good survey of current projects on and tools for performance monitoring and evaluation of Internet QoS.

In summary, a good deal of research work has been done in the area of the QoS of 3G networks and services but work tends to focus on the measurement and analysis of the parameters of network performance while giving a disproportionately less regard to the overall service quality that users perceive. Those works that do address service quality lack explanation on its measurement and analysis on 3G networks. This leaves a key question open: what do those parameters of performance mean to SLAs that SPs sign with customers? Furthermore, it does not take into account the fact that service quality information is crucial in developing new service offers. The survey of related works revealed another shortcoming. The research fails to address the bearing of the work and results on technical specifications and guidelines being developed by standards bodies such as 3GPP, TMForum, ITU-T and IETF. If they do address the works of standard bodies, they lack objective explanations on the use of standards.

This paper attempts to overcome shortcomings in current research work. It applies international standards to the development of the mechanism for the measurement and analysis of end-to-end service quality on 3G networks and then test out the concepts. This is the overall goal of the work presented in this paper.

3 Overview of the Work Presented

In order to fulfil the overall goal, a broad study of relevant work, including the work of standard bodies, was done to find out what were the limitations and how the work presented in this paper could complement them. Previous section briefly summarised this study. Requirements were also drawn from this study and fed into the development of an information model and a functional architecture. The rest of this section defines the problem identified, and briefly discusses the approach adopted and the results of the work.

3.1 Problem Definition

Lack of a mechanism to measure and analyse end-to-end service quality: As IT and 3G networks converge, service provision and delivery to a dynamic service market is also witnessing a major change. Even in the simplest case of service provision, administration and charging boundaries are crossed. Given this situation, SPs need to guarantee and deliver their customers a certain level of quality and run services efficiently.

The problem SPs face is two-fold. Firstly, what is the impact of service quality and network performance on SLAs that SPs sign with customers and other SPs? Secondly, how can end-to-end service quality be measured on 3G networks? Service quality cannot be measured unless SPs know what Key Performance Indicators (KPIs) and Key Quality Indicators (KQIs) are most relevant to 3G networks and services and how they can be used to measure the service quality delivered. They need a mechanism to measure, express and to report service quality in quantities and units.

Devising such a mechanism is a complex task because it involves expressing SLAs using service KQIs. The KQI values must be mapped onto KPI values. The task does not end with the mapping, as KPI values must also be transformed into KQI values.

Lack of the use of standards: The mechanism should not be ad-hoc and must be applicable to a wide variety of services and requirements. This is where relevant standards bodies such as 3GPP, TMForum (specially WSMH - Wireless Service Measurements Handbook and eTOM framework [19]), ITU-T, and IETF come into play. Standard bodies have put some serious effort into defining and standardising a common framework for QoS of IP-based services. Their work also leverages existing work from IETF, encompassing wireless specific requirements. Standards provide guidelines but they are not specific enough when it comes to implementation. This is reasonable because standard bodies ought not to align themselves to a particular technology. However, this creates difficulties for those who need a close insight into what these standards mean in terms of implementation and how they can be used. One of the novel aspects of this paper is that it elucidates for the reader *what* standards are necessary to understand end-to-end QoS measurement and develop the mechanism.

3.2 Approach Adopted and Results

In order to tackle the problems defined and develop the mechanism mentioned above, the approach that we have followed builds the concepts from networks upwards (Fig. 1).

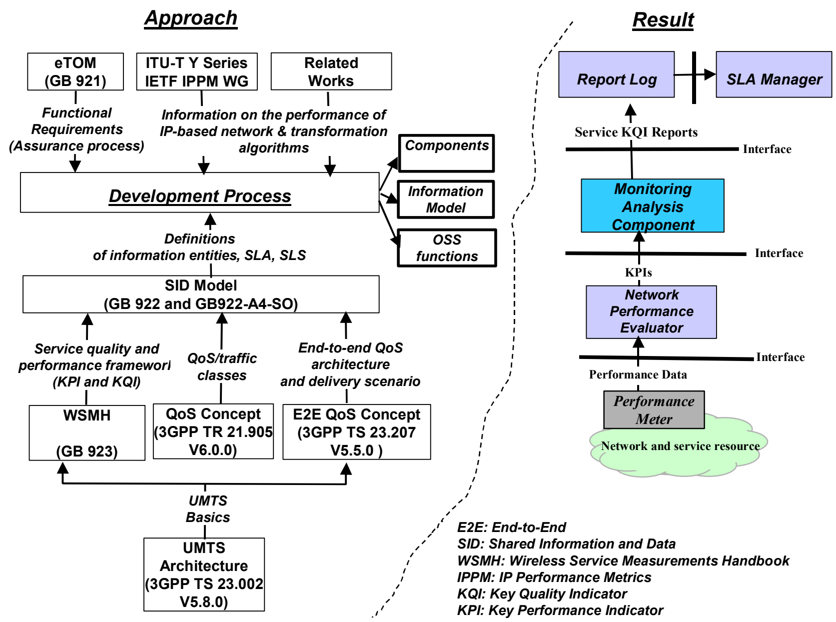

Fig 1: Approach Adopted and Results Achieved

The approach undertakes *Interactive* class of service (e.g., a Web-based service), identifies KPIs and KQIs that are required for the measurement of service quality of this class of service, and proposes a mapping of KPIs onto KQIs. An algorithm that transforms KPIs into KQIs is also proposed. Requirements proposed by TMForum in [4], 3GPP in [1], [2], [3], ITU-T (the Y series of ITU-T recommendation) in [9], and IETF (the IP Performance Metrics Working Group) in [10], [20] are used for the development of the mechanism.

An information model of KPIs and KQIs is based on the WSMH [4] and Shared Information and Data (SID) [21] [22], [23]. The OSS functions are based on the eTOM business process framework [19]. The development of the mechanism also requires an architecture for the delivery of end-to-end QoS and the definition of QoS class (or traffic class) that is applied to web-based 3G services. Both of these aspects are derived from 3GPP technical reports [2] and [3].

Thus far the research work has produced a component architecture that supports QoS measurement and analysis and SLA violation monitoring. A comprehensive information model of end-to-end QoS has also been produced. The OSS components MAC (Monitoring Analysis Component), SLA Manager, and Report Log illustrated in Fig. 1 are the main component of the system and also the subject of this paper. Operations of MAC are supported by two more components NPE (Network Performance Evaluator) and Performance Meter, which have been reused with slight modifications.

4 QoS Classes and End-to-End QoS Architecture

The 3GPP in [2] defines four generic QoS classes (or traffic class) that can be applied to user applications: *conversational*, *streaming*, *interactive* and *background*. Technical specification [3] complements [2] and proposes an architecture and mechanism to manage network performance for these traffic classes. Applications belonging to these classes exhibit certain features. In conversational class, the communication takes place between two or more persons. In cases of streaming and interactive classes, a person communicates with a computer. Lastly, in background class, a computer communicates with another computer. The main difference among these QoS classes lies in the amount of delay that user applications expect from networks. The architecture and scenario presented in this section concern interactive class.

4.1 End-to-end QoS Architecture

The main aim of [2] and [3] is to help OSS developers understand what network components are important and how are they interconnected in delivering end-to-end service quality. In a service delivery chain, it is also important to see how services running on those network components can be managed against agreed service quality objectives. It is important because QoS policies and SLAs are imposed at the boundaries between organizational domains. This paper adds value to the 3GPP architecture by highlighting how SPs can use the architecture to see where one SLA ends and another begins to deliver end-to-end service quality. Fig. 2 shows a specific configuration of the generic architecture described in [2] and operational at Fraunhofer FOKUS. The Gi is the interface between GPRS network and an external packet data network.

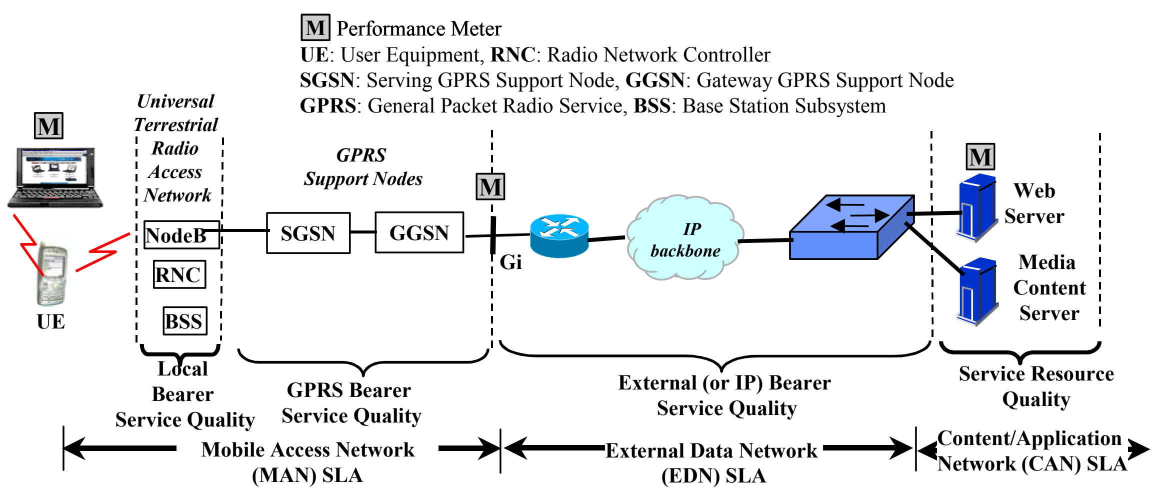

Fig 2: End-to-end QoS Delivery and Measurement Architecture

Measuring end-to-end QoS involves measuring QoS of the three segments shown in Fig. 2. The segment **MAN** combines Local and GPRS/UMTS Bearer Services [5]. GPRS/UMTS Bearer Service carries packets to external network and implements TCP/UDP/IP and GTP (GPRS Tunnelling Protocol). According to WSMH [4], networks providing Local and GPRS Bearer Services constitute MAN.

The second segment is the **EDN**, which corresponds to External Bearer Service. This is the IP backbone over which Web applications, FTP, email (using SMTP) run. In most cases, the External Bearer Service will run the IP protocol and a variety of IETF protocols (RSVP, DiffServ, IntServ) will be used. According to WSMH [4], this is the IP Transport bearer service. Although IP is omnipresent in data networks, the term EDN is more meaningful. EDN also makes a better sense because Gi interface connects UMTS/GPRS core network to an external data network. The **CAN** is the last segment of the end-to-end QoS delivery. According to WSMH, this is the IT Network.

4.2 QoS Delivery and Measurement Scenario

The delivery scenario used in this paper roughly corresponds to the scenario 1 described in [3]. The scenario assumes that the User Equipment (UE) does not provide IP bearer service management, and Gateway GPRS Support Node (GGSN) and IP core network supports the DiffServ. The measurement scenario is built on delivery scenario. Performance Meters (M-boxes in Figure 2) are deployed on the resources deemed important for meeting the requirements of MAC and interactive class of traffic.

The delivery scenario begins when the UE sends a Packet Data Protocol (PDP) Context Activation request and initiates a packet-switched session with the Serving GPRS Support Node (SGSN). During the activation procedure, the UE requests a temporary IP address from the network. It also specifies the Access Point Name (APN), i.e., address of media server shown in Fig 2, with which it wants to communicate. Then an end-to-end session is established and service usage begins. The measurement scenario can begin after service usage has begun. MAC requests NPE to obtain KPI values from network and service resources. NPE communicates with Performance Meters, which in turn obtain performance data. Since the current implementation of Performance Meters cannot be installed on the UE (or a UMTS mobile phone), for end-to-end measuring purposes, the UE is connected to a laptop and then a Web-based service is accessed.

Thus far the paper has provided the rationale for and a background to the research work. The following sections present the results of the research and development work.

5 Information Model

This section presents a comprehensive information model and a transformation algorithm required to analyse end-to-end service quality. Before an information model can be built, it is important to see how the measurement aspects that capture customer perception of service quality can be combined with the resources that provide information for analysis. Fig. 3 proposed a hierarchical relationship between the three main levels or views at which the QoS management information can be organised: **user/customer**, **service** and **network**. Network view corresponds to resource performance, which is captured by KPI values. Service view corresponds to the service quality, which captured by KQI values. User/customer view corresponds to service product quality and SLAs.

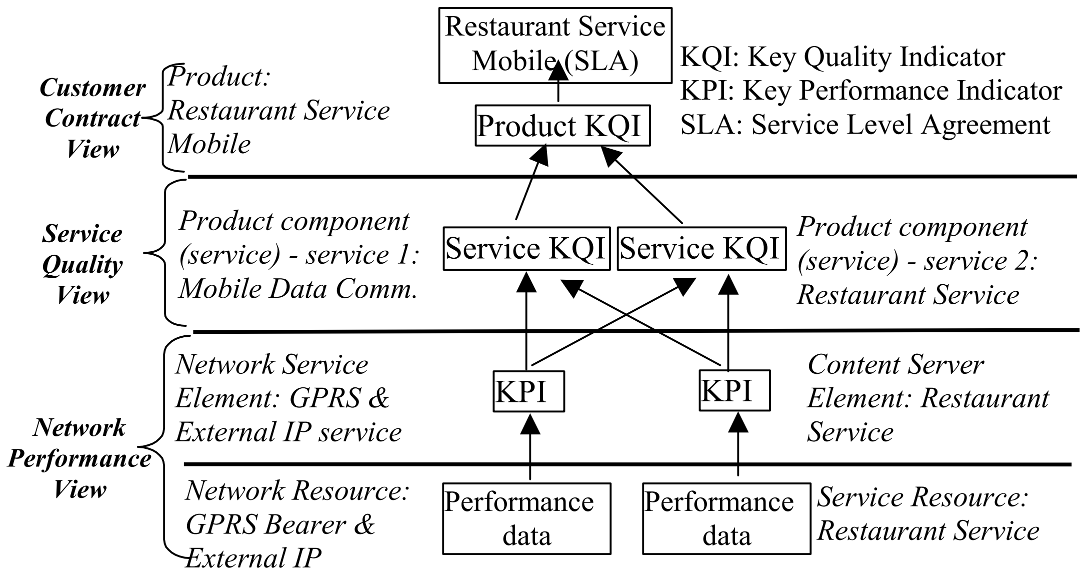

Figure 3: Measurement Views of End-to-End Service Quality For a Service Product

The hierarchical relationship adopts WSMH [4] for the views whereas the position of resources in end-to-end architecture and the KPIs they provide are derived from [2] [3]. Fig. 3 also shows a realistic and practicable model of a 3G service product and its constituents. The service product consists of three product components (or services), namely a Restaurant service, and access to internet (mobile data) via UMTS. This model, which is also reflected in the information model (Fig. 4), shows the resource from where performance data and KPIs can be obtained and the KQIs that they feed into. The Restaurant Service is a web-based third party demonstration service, which can be used with mobile phones, PDAs or PCs. It integrates location awareness based on standardised location interfaces (OSA/Parlay, Parlay X), which is a function that is specific to a 3G environment.

At the highest level of the information model is the SLA, which derives information from product KQIs. Product KQIs derive data from lower level service KQIs, which focus on measuring the performance of individual product components. A service may have one or more service elements. The Restaurant service has only one service element, which is the service itself. Service elements depend on the reliability of network and service resources and KPI values are calculated from the performance data provided by the resources.

Having shaped the relationship, it is important to look into the details of these views and sees what information entities constitute them (Fig. 4) and the KPIs and KQIs required for populating those entities (Tables 1 and 2). Finally a transformation algorithm is proposed.

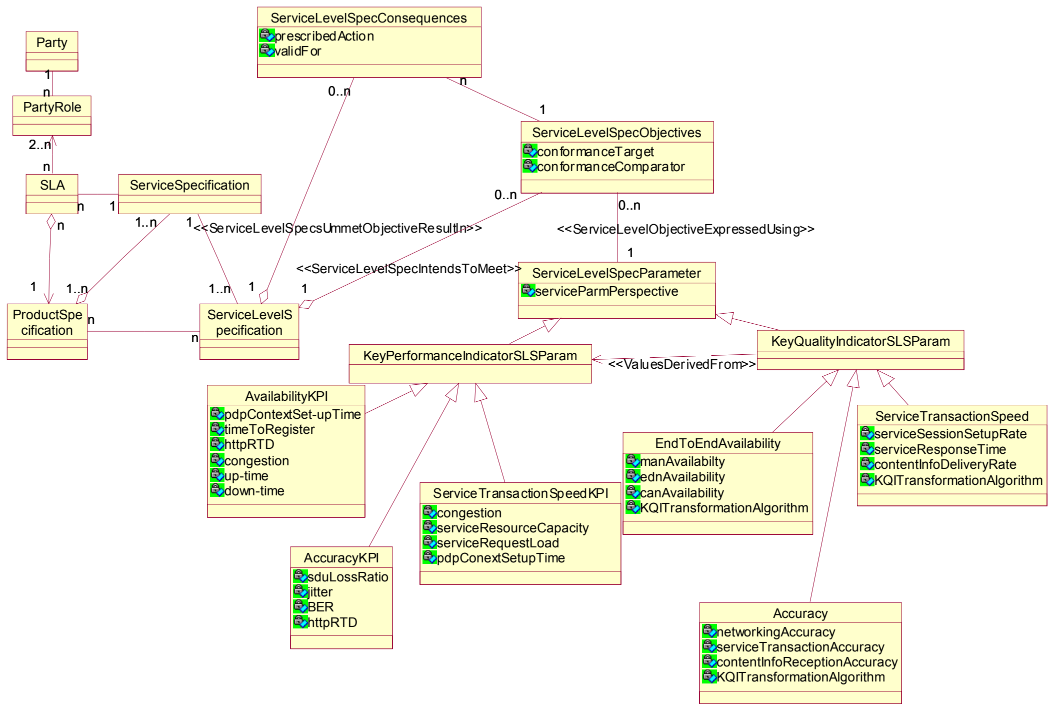

Fig 4: Information Model for QoS Measurement and SLA Management

SID models [21] [22] [23] have provided a set of entities that can be used for the specification of the information entities but they are fairly general. The information model takes those entities and provides details. It also proposes the manner in which aggregated KQIs can be represented in the information model and a relationship between KPI and KQI entities.

3GPP specification [2] does define fundamental characteristics of QoS of interactive class of service but they mainly provide for the QoS of Local Bearer and UMTS Bearer services (or MAN segment shown in Fig. 2). This does not show the complete picture. What is needed is a set of indicators that takes into account the performance of web-based application protocol (HTTP), wireless and fixed networks (UMTS and IP backbone networks) supporting that protocol, and the service resource. The contribution of this paper is that it defines a set of KPIs (Table 1) important for an interactive class of service. This set shows a more complete picture and also leverages from existing end-to-end QoS framework defined by ITU-T [9], and IETF [10] [20]. From these two bodies, performance parameters for IP and application protocol are derived.

It should be noted that the value of a KPI could also be used in the calculation of the values of other KPIs. For example, throughput value can be derived from the congestion and IP packet count values.

Aggregated KQIs are often specified in SLAs. In most cases they will be statistical values derived from KQIs. In must be noted that one or more KPI values can be used in the computation of one or more KQIs value. Furthermore, one KQI value can be used in computation of one or more aggregated KQI values.

Three of the main criteria against which service quality can be assessed are availability, accuracy, and speed [7] [9]. This paper looks at these criteria and proposes the KQI-KPI mapping required for them and a transformation algorithm is developed only for *availability* criterion.

Availability is the likelihood with which the relevant product or service can be accessed for the benefit of the customer at the instant of the request. *Accuracy* deals with the fidelity and completeness in carrying out the communication function with request to a reference. This indicator informs about the degree to which networks and services are dependable and error-free. *Service transaction speed* is the promptness with which network and service complete end-users' transactions.

5.1 Customer Contract View

5.1.1 Party

A party is a legal entity, e.g. an individual person or an organisation that is involved in zero or more business agreements with other parties.

5.1.2 Party Role

PartyRole represents the manner in which a party is involved within a Service Level Agreement. The same party can take part in multiple agreements in different ways, for example as a Service Provider with one of its Customers and as a Customer with one of its suppliers. A certificate is assigned to each party role: this certificate will be used to identify the different parties in every interaction required later on. In particular, this certificate will be used to identify the user/customer when logging on.

5.1.3 ProductSpecification

A product is a marketed collection of services offered to customers.

Product offerings, which are modelled together with the product entity, are targeted at one or more market segments based on the appropriate market strategy. Enterprises create product offerings from product specifications that have additional market-led details applied over a particular period of time. The customers do not need to know anything about the detailed product specification as the enterprise tailors the product offering to their needs. Where a business market strategy dictates, product offerings may be bundled together and re-priced as necessary. In addition, the time period for a bundled product offering will depend upon those for its component offerings.

When an enterprise agrees to deliver a product to a customer, there is a significant amount of information that needs to be recorded to implement that product on the provider's infrastructure or resources and to activate the service. This is modelled in product and associated entities within Figure NN.

The relationship between ProductSpecification and SLA signifies what services are *agreed* with the customer.

5.1.4 ServiceSpecification

This is an abstract base class for defining the ServiceSpecification hierarchy.

All Services are characterised as either being directly visible and usable by a Customer or not. However, each instance of a Service is made up of changeable as well as invariant attributes, methods, relationships and constraints. A ServiceSpecification defines the invariant characteristics of a Service. It can be conceptually thought of as a template that different Service instances can be instantiated from. Each of these Service instances will have the same invariant characteristics. However, the other characteristics of the instantiated Service will be specific to each instance.

This class can be thought of as a template, which represents a generic specification for implementing a particular type of Service. A ServiceSpecification may consist of other ServiceSpecifications supplied together as a collection. Members of the collection may be offered individually or collectively. ServiceSpecifications may also exist within groupings, such as within a Product.

The relationship between ProductSpecification and ServiceSpecification signifies what services are *offered* to the customer.

5.1.5 ServiceLevelAgreement

As described in the *SLA Management Handbook* [24], an SLA is a formal negotiated agreement between two parties regarding provided Service Level. Sometimes called a Service Level Guarantee, it is a contract (or part of one) that exists between two parties that is designed to create a common understanding about services, priorities, responsibilities, etc. It is usually written down explicitly and has legal consequences.

Handbook [24] defines an SLA or Contract as a set of procedures and targets formally or informally agreed between Network Operators/Service Providers (NOs/SPs) or between NOs/SPs and Customers, in order to achieve and maintain a specified QoS in accordance with ITU-T and ITU-R Recommendations. These procedures and targets are related to specific circuit/service availability, error performance, Ready for Service Date (RFSD), Mean Time Between Failures (MTBF), Mean Time to Restore Service (MTRS), Mean Time To Repair (MTTR), etc.

The SLA can involve a number of products, services, resources, and their related specifications. For example, an order from a new customer may specify a number of desired product offerings; a service level agreement may specify the services covered by the agreement. Items in an agreement may also be characterised by a price. The price may be a one-off price or a price associated with a product offering (in the case of an order for a product).

5.1.6 ServiceLevelSpecification

A ServiceLevelSpecification is a group of service level objectives, such as thresholds, metrics, and tolerances, along with consequences that result from not meeting the objectives. Service level specifications are also characterised by exclusions associated with the specification.

As an example, a service can be offered at a “Gold” service level and a “Silver” service level, with different pricing and agreed quality of service.

5.1.7 ServiceLevelSpecObjective

These objectives can be seen as a quality goal for a ServiceLevelSpecification defined in terms of parameters and metrics, thresholds, and tolerances associated with the parameters. The two *required* attributes of this entity are defined below.

conformanceTarget is a value used to determine if ServiceLevelObjective is met.

conformanceComparator is an operator that specifies whether a ServiceLevelObjective is violated above or below the conformanceTarget.

5.1.8 ServiceLevelSpecConsequences

It is an action that takes place in the event that a ServiceLevelObjective is not met. Violations to Service Level Agreements usually result in some consequence for the provider of the service. The consequence could be a general one that applies to any objective violation, or the consequence could be associated with a specific objective. An example of a consequence is a discount applied to a customer's bill if the number of trouble tickets for a product exceeds some number over a specified period of time.

5.2 Service Quality View

5.2.1 KeyQualityIndicatorSLSParam

Service Level Specification parameters can be one of two types: KQIs and KPIs. A KQI provides a measurement of a specific aspect of the performance of a Product (i.e. ProductSpecification, ProductOffering, or Product) or a Service (i.e. ServiceSpecification or Service) [4]. A KQI draws its data from a number of sources, including KPIs. This subsection deals with KQIs and the next subsection with KPIs.

KeyQualityIndicatorSLSParam is the superclass of EndToEndAvailability, Accuracy and ServiceTransactionSpeed.

5.2.2 EndToEndAvailability

EndToEndAvailability KQI, expressed as a percentage, defines the likelihood with which the relevant product or service can be accessed for the benefit of the customer at the instant of the request.

To determine the value of end-to-end availability, availability values of the following three segments are measured: MAN, for example, UMTS core network, EDN, for example, Internet and CAN, for example, SP's intranet. These three values are then combined in a single KQI value. Definitions of the three segments and end-to-end QoS are derived from [2] and [3]. The main attributes of this class are as follows:

manAvailability is the percentage of the time for which the MAN is available when the end user accesses and requires to use it.

ednAvailabilityPercentage is the percentage of time for which the EDN is available for use.

canAvailability is the percentage of the time for which the CAN is available for use.

KQITransformationAlgorithm is a logical step-by-step procedure used to calculate the value of a KQI.

5.2.3 Accuracy

Accuracy defines the fidelity and completeness in carrying out the communication function with request to a reference. This KQI informs about the degree to which networks and services are dependable and error free.

The attributes of this class are as follows:

networkingAccuracy is the degree to which functions of networks (MAN, CAN, EDN) are carried out without fault and error. For example, it can be the probability (expressed as a percentage) that networks will route a user's request to the correct content server.

serviceTransactionAccuracy is the degree to which the function of service resources is carried out without fault and error. For example, it can be the probability that the content server will provide the user with the right content.

ContentInfoReceptionAccuracy is the percentage of uncorrupted content/information exchanged between the UE and the end system (e.g., a content server).

5.2.4 ServiceTransactionSpeed

ServiceTransactionSpeed defines the promptness with which network and service complete end user transactions. The main attributes of this class are as follows:

serviceSessionSetupRate is the time required to complete a service session set-up.

serviceResponseTime is the rate at which a service responds to the user's invocation.

ContentInfoDeliveryRate is the data transfer rate at which content/information is exchanged between the UE and the content server for transactions.

5.3 Network Performance View

5.3.1 KeyPerformanceIndicatorSLSParam

KPIs provide a measurement of a specific aspect of the performance of a service resource (network or non-network) or group of service resources of the same type. A KPI is restricted to a specific resource type [4]. This is the superclass of three classes: AvailabilityKPI, AccuracyKPI and ServiceTransactionSpeedKPI.

5.3.2 AvailabilityKPI

This subclass represents the KPIs whose values are required for the calculation of the EndToEndAvailability KQI value. The main attributes of this subclass are as follows:

pdpContextSet-upTime is the time it takes for the network to establish a PDP context for the UE (User Equipment).

timeToRegister is the time it takes for the UE to register for a PS service.

httpRoundTripDelay is the times for which the source (i.e. a media application) waits to receive a response from the destination (i.e. a content server). Three types of HTTP RTD KPIs can be envisaged: MAN_RTD, EDN_RTD, CAN_RTD.

up-time is the total amount of time during which MAN, EDN or CAN is up and running when the end user requires to use it. Depending on the network segments being monitored, there can be three types of up-time: MAN_Up, EDN_Up or CAN_Up.

down-time is the total amount of time during which MAN, EDN or CAN is down when the end user requires to use it. Depending on the network segments being monitored, there can be three types of down-time: MAN_Dn, EDN_Dn or CAN_Dn.

5.3.3 AccuracyKPI

This subclass represents the KPIs whose values are required for the calculation of the Accuracy KQI value.

The main attributes of this subclass are as follows:

sduLossRatio is the ratio of lost SDUs to total SDUs transmitted. A SDU can get lost when a wrong header is attached to it.

httpRoundTripDelay (See subsection 5.3.2)

BER (Bit Error Rate) is the number of erroneous bits divided by the total number of bits transmitted, received, or processed over some stipulated period. It signifies the degree to which SDUs are transmitted with low error.

Jitter is the variation in delay or response time.

5.3.4 ServiceTransactionSpeed

This subclass represents the KPIs whose values are required for the calculation of the ServiceTransactionSpeed KQI value.

The main attributes of this subclass are as follows:

Congestion is used in the generic sense meaning overburden of activities at the UDP layer in the GSN, at DNS servers operating in MAN, EDN or CAN, or at the service resource due to traffic overload or over-utilisation. The types of Congestion KPIs envisaged are: GSN_Congestion (at the SGSN or GGSN), Gi_Congestion, EDN_DNS_Congestion, CAN_DNS_Congestion.

serviceResourceCapacity is the ability of the content server to process users' requests and expressed as the number of requests processed by the content server in a *specified time interval*.

ServiceRequestLoad is the total number of user requests that the content processes at a specified point in time.

pdpContextSet-upTime (See subsection 5.3.2).

5.4 Transformation Algorithm for End-to-End Service Availability

In the case of "always-on" services, the users don't want to know how long the service was available. Their main concern is that it must be available when *they* try to use it. Customers' perception of QoS is a judgement made by the customer about the overall excellence (or otherwise) of the delivered product. The core of the problem in expressing QoS is that there are many independent variables and aggregated service quality is a variable dependent on those independent variables. Customer-perceived QoS is a function of these variables. End-to-end availability depends on the availability of three segments MAN, EDN, and CAN (Fig. 2).

End-to-End availability = average (MAN_Availability + EDN_Availability + CAN_Availability) **---- (1)**

This equation is based on some assumptions. Firstly, the outage of individual segments is mutually exclusive. Secondly, all traffic between end-systems is carried on a single path. A further assumption is made about the KPI values required for the calculation of the value of Availability KQIs. Although the KPIs listed against MAN_Availability, EDN_Availability and CAN_Availability are important, the following equations are based on only MAN_Up and MAN_Dn KPIs as variables. On the basis of these assumptions, Availability can be expressed by equation 2:

$$\%\ \text{Availability} = \left(Up\_Time / \left(Up\_Time + Down\_Time\right)\ \right) \times 100 \quad \text{------}\ \textbf{\textit{(2)}}$$

Equation 2 is applied to the three segments, which gives the following three equations:

$$\%\ \text{MAN\_Availability} = \left(MAN\_Up / \left(MAN\_Up + MAN\_Dn\right)\ \right) \times 100 \quad \text{------}\ \textbf{\textit{(3)}}$$

$$\%\ \text{EDN\_Availability} = \left(EDN\_Up / \left(EDN\_Up + EDN\_Dn\right)\ \right) \times 100 \quad \text{------}\ \textbf{\textit{(4)}}$$

$$\%\ \text{CAN\_Availability} = \left(CAN\_Up / \left(CAN\_Up + CAN\_Dn\right)\ \right) \times 100 \quad \text{------}\ \textbf{\textit{(5)}}$$

These three equations give us an average of the end-to-end availability:

$$\text{ETE\_Availability} = \left(\%\text{MAN\_Av} + \%\text{EDN\_Av} + \%\text{CAN\_Av}\right) / 3 \quad \text{------}\ \textbf{\textit{(6)}}$$

However, this availability does not indicate whether the end-user was able to use the service when he or she tried to use the service. If any of the three segments is out of order, the service is not available. Therefore, the definition of *availability* given above can be expressed as the rate at which end-user's attempts to access an end-to-end service result in success. Here success means that all the segments (i.e., maximum value of ETE_Availability) are available to use. Therefore the aggregated service KQI (or product KQI) is:

$$Aggregated\_Availability = \left(\left(\sum_{n=1}^{n=m} ETE\_Availability_n\right) / m\right) \times 100 \quad \text{------}\ \textbf{\textit{(7)}}$$

In equation (7): n = n^{th} try, m = total number of tries, ETE_Av_n = availability of end-to-end service at n^{th} try.

5.5 Limitation of the SID Model

SID models are not very clear about combining values of two or more KQIs to form a single combined KQI, as proposed in WSMH [4]. The value of such combined KQIs is useful in determining the extent to which service level objectives are met. In order to provide for a combined KQI, the AlbatrOSS SLA model presented in this section derives subclasses from KeyQualityIndicatorSLSParam of the SID model [22]. A subclass represents the combined KQI whereas attributes of the subclasses represent the KQIs themselves.

6 OSS Functions and Component Architecture

In order to test out the concepts described in this paper, OSS functions and component architecture are being developed and initial results of the implementation are presented in this section. The OSS functions are modelled on the Assurance process of eTOM business process framework. The component architecture also corresponds to that of the Assurance process.

6.1 OSS Functions

Three types of OSS functions are required for the development of a OSS capable of supporting end-to-end QoS measurement and SLA management: *SLA Management*, *Service Quality Management* and *Resource Performance Management*. This section provides only brief descriptions of the OSS functions. More details can be found in AlbatrOSS project deliverable D10 [25].

SLA Management is responsible for ensuring that the performance of a product delivered to customers meets the SLA parameters, or product KQI objectives, as agreed in the contract concluded between the service provider and the customer. Specific OSS functions include specification, monitoring, analysis, and reporting of SLA.

Service Quality Management is responsible for ensuring that the performance of a service delivered to customers meets the service KQI objectives. The specific functions included in the OSS are: (1) specification of service KQIs, (2) collection of KPI values and transformation of those values into KQI values, (3) analysis of KQI values against KQI objectives, and (4) production of overall service quality.

Resource Performance Management is responsible for ensuring that the performance of the resources involved in delivering services to customers meets the resource KPI objectives. Specific functions included in the OSS are: (1) specification of service KPIs, (2) collection and aggregation of performance data and into KPI values, (3) analysis of KPI values against KPI objectives, and (4) reporting of KPI.

6.2 Component Architecture

The components of the architecture (Fig. 5) can be distributed over service and network administrative domains and are connected with one another by means of three interfaces. Interfaces are specified in WSDL (Web Service Description Language).

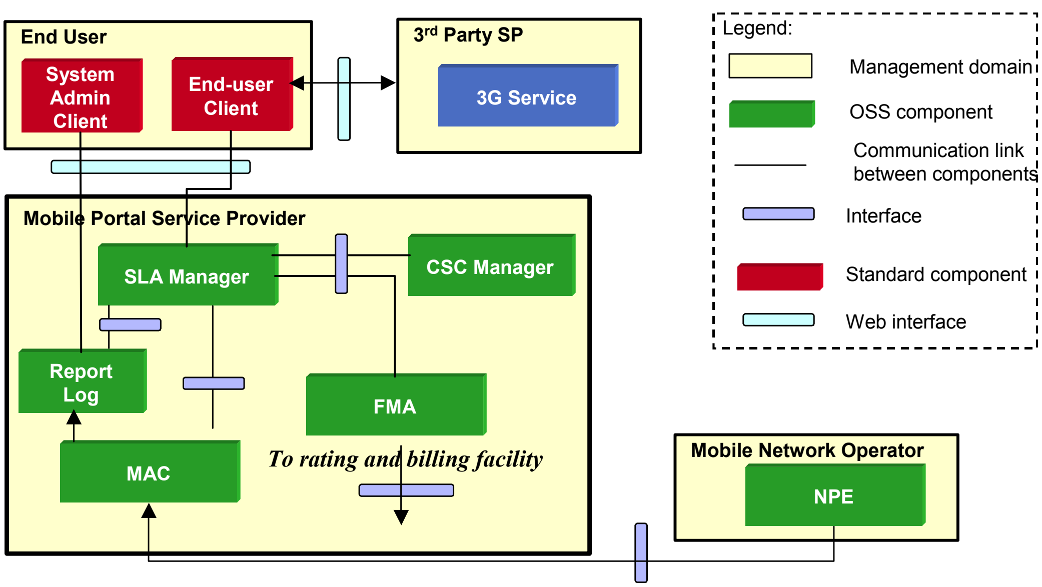

Fig 5: Component Architecture of the End-to-end QoS Measurement and SLA Management

More information on the development of OSS components, such as interfaces specification and technology used can be found in AlbatrOSS deliverable D9 [26]. Deliverable D10 [25] can be consulted for an operational view of the OSS component presented in the section. The operational view provides UML Sequence Diagrams to depict information exchanges between OSS components.

6.2.1 Monitoring Analysis Component

The MAC collects and aggregates values of KPIs from network and service resources and transforms them into KQI values using algorithms, for example, mathematical equations. The MAC uses the value of one or more KQIs and service quality objectives as a basis and issues reports containing warnings or alarms to the SLA Manager. These reports inform the SLA Manager that the service quality is about to deteriorate or has deteriorated below a threshold value.

6.2.2 Network Performance Evaluator

The NPE component accomplishes two functions: measuring performance in KPI values and aggregating usage data. The KPIs values are delivered to the MAC whereas usage data values are delivered to the FMA.

The process of collecting values of KPIs or usage data from network and service resources is automated. The NPE is a rule-based component and operates in conjunction with performance meters, which collect data from network and service resources. Meters can be programmed by means of filter specifications, which are simple rules that instruct meters to monitor data traffic originating from any specific IP address.

6.2.3 Report Log

The Report Log component receives service quality reports from the MAC and stores them in a log. It also allows the service provider to view and browse through the report log via a web browser.

6.2.4 SLA Manager

This component is responsible for maintaining customer SLAs and for ensuring that the delivered QoS meets the QoS specified in customer contracts or other product specifications. It ensures that the provisions in the contract are being met and issues warnings and alarms if this is not the case. This means that it must be able to monitor, analyse and report on the actual QoS achieved compared with the QoS specified in the SLA.

6.2.5 Supporting Components

There are two more components namely FMA and CSC Manager that take part in SLA management therefore they deserve description here.

The CSC Manager provides the database contents in the form of methods including the needed logic by other components. Those methods include the management of parties, services, products and SLA based data such as the service level specification information provided for the SLA Manager.

The FMA component performs usage mediation for services and network resources. It collects and aggregates usage data, encodes the data in IPDR documents, and transfers the documents to a rating component. It also collects SLA violations reports and forwards them to a rating and billing facility.

6.3 Evaluation and Performance Results

The entire work presented in this paper has developed during AlbatrOSS project and the components have also been successfully demonstrated in the project trials, which were executed on the Fraunhofer FOKUS UMTS test-bed. This section summarises the main results of the evaluation. AlbatrOSS deliverable D11 [27] can be consulted for a more details on OSS components evaluation.

The main goals of the evaluation were two-fold: (1) to validate whether the OSS components provided a mechanism for the measurement of end-to-end QoS of interactive service class and SLA violation monitoring (2) to validate whether the components conformed to the work of standard bodies. The results of evaluation showed that components fulfilled these goals.

The results of performance measurement using MAC showed that HTTP RTD varied greatly when a single end-user browsed the web on the UMTS network (minimum 656 and maximum 1415736 in micro seconds). TCP RTD did not vary to that extent (minimum 445 and maximum 587 in microseconds). It is fair to say that this information will be very useful for network operators and SPs who would like use it for many purposes, e.g., creating service profiles.

While MAC was evaluated against the standards used, the usefulness of standards themselves was also assessed. Standards were evaluated against two main criteria: the **benefits** and the **quality** of the standards. From the assessment it can be concluded that standards have greatly helped in the design and development of MAC and configuration of end-to-end architecture on the UMTS test bed. The WSMH provides a useful framework of performance and service quality indicators. However it must also be added that, in comparison to ITU-T and IETF work, 3GPP specifications are more difficult to understand.

7 Conclusions, Discussions and Future Work

The area of QoS and SLA management is clearly vast and a single standards body will not be enough to complete the overall picture, explain the overall problem and provide practicable solutions. The mappings of standards are not specific enough when it comes to workable examples. They provide a generic framework within which to develop an information model. Therefore it rests upon White Papers such as this one to disentangle the whole issue and inform the research community in detail of the standards that are useful and how they complement one another.

By and large, research community is under assumption that UMTS/GPRS core networks have been tested successfully and internet services and applications will soon be running on them. This assumption is far from reality, as network operators are reporting numerous teething problems. This paper adopts an approach whereby measurement concepts are built from the network domain upwards. This approach enables one to see what is available at the network level and how it can be used at the service level, which has proven to a practicable approach for the authors.

This paper tackles the area of end-to-end QoS of service products and UMTS, and addresses two different points of view, namely service management and communications networks. This builds an overall picture of QoS in the terms of customer, network operators, service providers, and content provider. In order to meet the requirement of all these stakeholders, a mapping between service KQIs to the KPIs of the individual networks and services in different domains is proposed. An algorithm that transforms KPI values into KQI values is also proposed. One of the main conclusions is that establishing a clear-cut mapping between KPIs and KQIs for end-to-end service quality and developing a transformation algorithm is a complex area. Nonetheless, algorithms for measuring service performance need to be further investigated in order to ascertain the best ones for measuring the QoS being obtained by end users in a mobile environment.

Another conclusion is that if SPs are to offer commercially viable service contracts to end-users and bill accurately, they will require a practicable regime and meaningful measure of QoS. The measure must provide information not only on the bearer service but also on the QoS of an entire service product. Since the work described in this paper is based on the work of 3GPP and TeleManagement Forum, it is hoped that this paper will benefit a large number of telecom operators and SPs.

The future work will enhance the operations of MAC and SLA Manager by developing more service KQIs and more compound transformation algorithms. Further work on end-to-end QoS measurement of a service product and QoS in a mobile contents delivery scenario will also be carried out.

Glossary

Component: A self-contained entity with well-defined interaction points and behaviour.

Interface: A boundary enabling communication between two entities. The properties of an entity that are being made visible at the interface are provided via an interface specification.

Key Performance Indicator: It provides a measurement of a specific aspect of the performance of a service resource (network or non-network) or group of service resources of the same type. A KPI is restricted to a specific resource type.

Key Quality Indicator: It provides a measurement of a specific aspect of the performance of the product, product components (services) or service elements and draws their data from a number of sources including the KPIs.

OSS: A system providing support for the management of resources, applications and services.

Quality of Service: QoS is a measure that indicates the collective effect of service performance that determines the degree of satisfaction of a user of the service. The measure is derived from the ability of the network and computing resources to provide different levels of services to applications and associated sessions. The measure can be both quantitative and qualitative.

Service: A consistent collection of capabilities (features, functionality, resources, infrastructure, terms and conditions, assets) that is offered to clients. Service can be categorised as follows:

> **End-user service**: A service offered to customers by service providers. Such a service can, for example, provide information, such as text, audio or video, or perform specific operations for the user.
>
> **Network-level service**: A basic service related to the transmission of data through a given infrastructure.
>
> **Middleware service**: A service located above the network-level services supporting the interoperability and cooperation of components in the system.

System: A collection of components organised to provide a specific functionality.

Acronyms

3G: Third Generation

API: Application Programming Interface

ASP: Application Service Provider

CAN: Content/Application Network

CSC Manager: Customer, Service and Contract Manager

DNS: Domain Name Service

EDN: External Data Network

eTOM: enhanced Telecommunication Operations Map

ETSI: European Telecommunications Standards Institute

FMA: Federated Mediation Adapter

GGSN: Gateway GPRS Support Node

GSN: GPRS Support Node

GPRS: General Packet Radio Service

IP: Internet Protocol

ISP: Internet Service Provider

ITU-T: International Telecommunication Union – Telecommunication Standardization Sector

KPI: Key Performance Indicator

KQI: Key Quality Indicator

MAC: Monitoring Analysis Component

MAN: Mobile Access Network

MNO: Mobile Network Operator

MTBF: Mean Time Between Failures

MTRS: Mean Time to Restore Service

MTTR: Mean Time To Repair

NGOSS: New Generation Operations Systems and Software

NPE: Network Performance Evaluator

PDA: Personal Digital Assistant

PDP: Packet data Protocol

OSS: Operations Support System

QoS: Quality of Service

RFSD: Ready for Service Date

SGSN: Serving GPRS Support Node

SID: Shared Information and Data

SLA: Service Level Agreement

SOAP: Simple Object Access Protocol

SP: Service Provider

TM Forum: TeleManagement Forum

UE: User Equipment

UL: User Location

UML: Unified Modelling Language

UMTS: Universal Mobile Telecommunications System

URL: Uniform Resource Locator

WAP: Wireless Application Protocol

WSDL: Web Services Description Language

XML: Extensible Markup Language

References

[1] 3rd Generation Partnership Project, "Technical Specification Group Services and System Aspects: Vocabulary for 3GPP Specifications", 3GPP TR 21.905 V6.0.0 (2002-09) (Release 6).

[2] 3rd Generation Partnership Project, "Technical Specification Group Services and System Aspects: Quality of Service (QoS) concept and architecture", 3GPP TS 23.107 V5.10.0 (2003-09) (Release 5)

[3] 3rd Generation Partnership Project, "Technical Specification Group Services and System Aspects: End-to-End Quality of Service (QoS) concept and architecture", 3GPP TS 23.207 V6.0.0 (2003-09) (Release 6).

[4] "Wireless Service Measurements Handbook", GB 923, Version V2.6, TM Forum, Morristown, New Jersey, December 2003.

[5] "Understanding General Packet Radio Service", App. Note 1377, Agilent Technology, www.agilent.com.

[6] Rajiv Chakravorty, Maurizio D'Arienzo, Jon Crowcroft and Ian Pratt, "Dynamic SLA-based QoS Control for Third Generation Wireless Networks: The CADENUS Extension", Proceedings of *IEEE International Conference on Communications (IEEE ICC 2003)*, 11-15 May 2003, Anchorage, Alaska, US.

[7] A. P. Ooden, K. E. Ward, A. W. Mullee, "*Quality of Service in Telecommunications*", IEE Telecommunication Series 39, Publihsed by IEE UK, 1997, ISBN: 0 85296 919 8.

[8] Tomas Robles, et al, "*QoS Support for an All-IP System Beyond*", IEEE Communication Magazine, August 2001.

[9] ITU-T Y.1541, SERIES Y: GLOBAL INFORMATION INFRASTRUCTURE AND INTERNET PROTOCOL ASPECTS, Internet protocol aspects – Quality of service and network performance, Network Performance Objectives for IP-based services, May 2002.

[10] RFC 3148, *A Framework for Defining Empirical Bulk Transfer Capacity Metrics*, Category: Informational, IETF, July 2001.

[11] Priggouris, et al, "Enhancing the General Packet Radio Service with IP QoS Support", Proceedings of International Workshop QoS-IP 2001, Rome Italy, January 2001.

[12] Staehle, Leibnitz, and Tsipotis, "*QoS of Internet Access with GPRS*", Wireless Networks, Volume 9, No. 3, May 2003.

[13] Aljadhai and Znati "*Predictive Mobility Support for QoS Provisioning in Mobile Wireless Environment*", IEEE Journal On Selected Areas In Communications, Volume 19, No. 10, October 2001.

[14] Zhuang et al, "*Policy-based QoS Architecture in the IP Multimedia Subsystem of UMTS*", IEEE Network, May/June 2003.

[15] Chalmers and Sloman, "*A Survey of Quality of Service in Mobile Computing Environments*", IEEE Communication Survey, Second Quarter 1999.

[16] Chen and Hu, "*Internet Performance Monitoring*", Proceedings of The IEEE, Vol 90, No 9, September 2002.

[17] Yang and Chou, "*Adaptive QoS Parameter Approach to Modeling Internet Performance*", International Journal of Network Management, November 2002.

[18] Keller and Ludwig, "*The WSLA Framework: Specifying and Monitoring Service Level Agreement for Web Services*", Journal of Network and Systems Management, Volume 11, No. 1, March 2003.

[19] *enhanced Telecom Operations Map™ (eTOM). The Business Process Framework*, GB 921, TM Forum Approved Version 3.0, TeleManagement Forum, Morristown, NJ, June 2002

[20] RFC 2681, A Round-trip Delay Metric for IPPM, Category: Standard track, IETF, Sept 1999.

[21] *Shared Information/Data (SID) Model. Concepts, Principles, and Domains*, GB922 Member Evaluation Version 3.1, TeleManagement Forum, Morristown, New Jersey, July 2003

[22] *Shared Information/Data(SID) Model. Addendum 4SP – Service Overview Business Entity Definitions*, GB922 Addendum-4SO, Member Evaluation Version 2.0, TeleManagement Forum, Morristown, New Jersey, July 2003

[23] Shared Information/Data (SID) Model: *Addendum 1A -Common Business Entity Definitions – Agreement (including Service Level Agreement)*, Member Evaluation Version 3.1, TeleManagement FORUM, Morristown, New Jersey, July 2003.

[24] *SLA Management Handbook*, GB917 Public Evaluation/Version 1.5, TeleManagement Forum, Morristown, New Jersey, June 2001.

[25] *AlbatrOSS Final Architecture for a Personalised Mobile Services Environment*, IST-2001-34780 AlbatrOSS Project Deliverable 10, 31 January 2004. www.ist-albatross.org

[26] *System Version 2 ready for Trial 2*, IST-2001-34780 AlbatrOSS Project Deliverable 9, October 2003. www.ist-albatross.org

[27] *Evaluation of Trial System Version 2*, IST-2001-34780 AlbatrOSS Project Deliverable 11, December 2003. www.ist-albatross.org