Service Level Agreement as an issue of teletraffic

Manfred Schneps-Schneppe[a] and Villy B. Iversen[b]

[a]Svetets Ltd, 2nd Baumanskaya st #9/23, 105005 Moscow, Russia

[b]COM, Building 345v, Technical University of Denmark, DK-2800 Lyngby, Denmark

After discussion of SLA issues the mathematical methodology – how to design the global SLA index on the basis of Linear Discriminant Functions – is developed. The experience of Verizon New York company is used as a best practice. Service quality monitoring is a useful base for the teletraffic community to join around for future packet switching network studies.

1.INTRODUCTION

1.1. Emergence of SLA studies

The problem of how to deal with multi-dimensionality of QoS measures has a long history. We will go through the problem by the few introductory references (without the mind of completeness).

Refer, say, to the British engineer E P G Wright's paper from ITC-4 in 1964 [1]. He pointed out that "the Quality of Service specification (Figure 1) strives to establish an appropriate balance over the whole range of parameters to which the subscriber is exposed." Until now, it is not clear how to deal with these many measures.

min		Max
.01	Grade of service	.02
1 sec	Speed	2 sec
.00001	False Clear Down	.00005
.001	Wrong number	.002
1 sec	Post dialing delay	5 sec
.005	No tone	.01
.00001	Wrong charging	.00005
.0001	Double connection	.0002

Figure 1. Quality of Service specification

Newertheless, the year-long experiences of Oftel in the field of QoS regulation is widely acceptance in Europe, and in 1997 ETSI published a standard [2], recommending the following nine QoS measures: Fault report for access line per year, Unsuccessful call ratio, Call set up time, Supply time for initial network connection, Percentage of orders completed on or before the date confirmed or contracted with the customer, Response time for operator service, Availability of card or coin operated public pay phones, Fault repair time, and Service restoration.

The deregulation process that started during the last decade in telecommunications is the cause of meaningful changes. Nowadays service/network providers try to improve the service quality in order to differentiate their products from those of their competitors. Service Level

Agreements (SLAs) are a useful tool to formalize the inter-relationships between entities. It is a formal agreement between two or more entities with the scope of assessing service characteristics, responsibilities and priorities of every part. And what is of the special interest, an SLA may include compensations for an unreached level of quality as an economic issue of the contract. The recent ITU-T Recommendation E.860 "Framework of a service level agreement" (2002) [15] reflects the very initial international efforts to look for solution of this extremely acute interconnection problem.

The Internet expansion is an additional key factor for SLA studies. These studies are carried on in the framework of several IETF Working Groups. The IP Performance Metrics WG, for example, is developing a set of standard metrics that can be applied to the quality, performance, and reliability of Internet data delivery services [3]. The metrics are: connectivity, one-way delay and loss, round-trip delay and loss, delay variation, bulk transport capacity and several others. Earlier SLAs were focused on network infrastructure performance, for example, that packets should take no longer than 200 ms to travel from router A to router B. Today SLAs are focusing on application performance, and application service providers (ASP) may assume some risk by providing service guarantees. Such a responsibility can charge higher rates, say 3 %, in penalties of the revenues received under the contract.

Besides, there are several international projects on new technologies (as Intserv, Diffserv, MPLS) in the pursuit of end-to-end QoS, in particular the projects AQUILA [4] and TEQUILA [5]. Many SLA monitoring tools are available on the market, naming a few: HP OpenView Firehunter, CiscoWorks2000 Service Management Solution [6], and Lucent's CyberService provider [7]. For example, the Lucent's SLA managers offer three QoS levels: gold, silver and bronze. The "gold–silver–bronze" approach is going now in the wide use, e.g. ETSI TIPHON project offers the following quality grades for IP telephony voice packet loss: < .5% for class 1 = gold, .5% to 1% for class 2 = silver, 1% to 2% for class 3 = bronze.

1.2. Outline

The paper deals with three issues:

1) Designing a global SLA index for interconnection regulation having the form of linear discriminant function $Q = w_1x_1 + w_2x_2 + \ldots + w_nx_n$ where w_i are the measured SLA metrics and x_i are their weighting coefficients, calculated on the basis of multidimensional statistical analysis. Any service/network provider can be correlated to one of three classes according to the following resolution scheme: gold level (expensive) if $Q < Q1$, silver level (not so expensive) if $Q1 < Q < Q2$, and bronze level (inexpensive) if $Q2 < Q$.

2) Developing the approach to choose a subset of informative quality indicators to assure the robustness of the interconnection conflict resolution scheme.

3) Developing a compensation and penalty scheme that would reduce the company's rate of return and provide funds to correct problem areas and address customer complaints.

The paper, written in the form of tutorial, includes some results from our paper [8] presented at the ITC Specialist Seminar held in Girona, Spain (Apr 2001). Best practice − New York Telephone (now Verizon NY) regulation − is from ITC15 paper [9] co-authored by the U.S. FCC staff member J.Kraushaar. How to built covariance matrix for SLA feature analysis, was discussed in an ITC16-paper [10]. The mathematical methodology and software has been presented in a Softstat'93 paper [11].

2. MATHEMATICAL METHODOLOGY

2.1. How to built the global SLA index

Classification of two Gaussian groups. We consider an n-dimensional space of SLA indicators. According to an agreement between the incumbent operating company and the entrant (say, ASP), some negotiated n SLA measures $q_1, q_2, ..., q_n$ are arrived at. During a negotiated period of operation (say, a year or a quarter), a new set of SLA values $x_1, x_2, ..., x_n$ could be measured. Therefore, we have two points, q and x, in an n –dimensional space representing two service level "states". Naturally, the difference between these points may require corrective action by the incumbent or entrant carrier (or both) and some financial consequences (penalties or rebates).

How to solve the problem? We reduce it to the classification of two Gaussian groups by means of multidimensional statistical analysis. Let us have two groups D_1 and D_2 described by their SLA measures. These could be the groups of telecommunication objects, say, exchanges, operators, ISPs, and ASPs. For example, group D_1 contains companies operating during the previous time periods (years, months, etc.) in accordance with SLA, another group D_2 was in conflict with service level standards. If we now have one new current object x, the problem is to classify it as either belonging to D_1 or D_2. Similar classification methodology is widely used in many fields, particularly in medical diagnostics. We may consider these two groups as patients, and talk about measurements and data as medical symptoms. Then we have two diagnosis: D_1 (healthy patients) and D_2 (some illness, say tuberculosis).

As a mathematical model for these two groups we choose two $n-$dimensional normal (Gaussian) distributions $N_n(\mu_1, \Sigma)$ for "good" cases and $N_n(\mu_2, \Sigma)$ for "bad" cases, respectively. That is, we have the following objects: 1) two mean value vectors $\mu_1 = (\mu_{11}\mu_{12}...\mu_{1n})^T$ and $\mu_2 = (\mu_{21}\mu_{22}...\mu_{2n})^T$, where T is a symbol of transposition (in this case from row-vector to column-vector), and 2) covariance matrix $\Sigma = (\sigma_{ij})$ the same for both distributions. Our current object is described by n measures $x_1, x_2, \ldots, x_n$. This vector x presents a point of n –dimensional space and the problem of classification is reduced to the choise of one of two decisions: D_1 or D_2. We use a linear discriminant function LDF, in other words, a scalar product of vectors w and x:

$$z = w_1 x_1 + w_2 x_2 + \ldots + w_n x_n \tag{1}$$

where $w_1, \ldots, w_n$ are unknown constants, and we choose some threshold value a so that the decision rule is as follows:

$$\begin{aligned} &x \subset D_1 \_if \_ z \le a \\ &x \subset D_2 \_if \_ z > a \end{aligned} \tag{2}$$

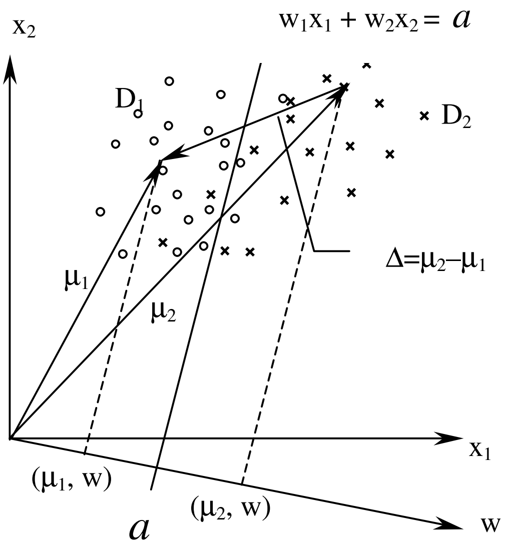

Figure 2. Geometrical interpretation of classify-cation for a two-dimensional case.

Thus we reduce the classification issue to two problems:

1) to determine n unknown coefficients $w_1,\ldots,w_n$ so that the distance between projections of mean vectors μ_1 and μ_2 on vector w is maximal and

2) to choose point a between these projections on vector w minimising the probability of false classification.

Figure 2 illustrates the classify-cation process for two-dimensional case: two clouds D_1 and D_2 divided by line $w_1x_1 + w_2x_2 = a$

Mahalanobis distance. If our point x belongs to group D_1 then variable z defined by (1) has one-dimensional normal distribution with mean and variance:

$$z_1 = \sum_{i=1}^{n} w_i \mu_{1i} = w^T \mu_1 \qquad \sigma_z^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{ij} = w^T \Sigma w \tag{3}$$

In a similar way if x belongs to D_2, then z has normal distribution with mean

$$z_2 = \sum_{i=1}^{n} w_i \mu_{2i} = w^T \mu_2 \tag{4}$$

and the same variance.

Problem 1 is to choose constants $w_1,\ldots,w_n$ maximising the so called Mahalanobis distance

$$M^2 = \frac{(z_2 - z_1)^2}{\sigma_z^2}$$

This measure written in initial notations (3) - (4) is as follows:

$$M^2 = (\mu_2 - \mu_1)^T \sum{}^{-1} (\mu_2 - \mu_1) \tag{5}$$

How to determine the vector w ? Let us use the technique of Lagrange multipliers to minimise the mis-classification probability. After rather long algebraic transformations, we get the solution:

$$w = \Sigma^{-1}(\mu_2 - \mu_1) \tag{6}$$

Therefore, we offer the global QoS index in form:

$$Q = w_1 x_1 + w_2 x_2 + \ldots + w_n x_n \tag{7}$$

Optimal discriminating threshold. How to choose threshold value a to minimize the mis-classification? Let us denote by α the probability of choosing decision D_2 when the true one is decision D_1, i.e. z belongs to D_1, and β otherwise (Figure 3):

$$\alpha = \mathrm{P}\{z > a / D_1\} = 1 - \Phi\left(\frac{a - z_1}{\sigma_z}\right) \qquad \beta = \mathrm{P}\{z \le a / D_2\} = 1 - \Phi\left(\frac{z_2 - a}{\sigma_z}\right) \tag{8}$$

where $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp\{-u^2/2\}$

In case of equal apriori probabilities of belonging to the classes D_1, D_2 and equal penalties to classify the new case as D_1 if D_2 is true and vice versa, the constant a may be chosen it he following form:

$$a = \frac{z_1 + z_2}{2} \tag{9}$$

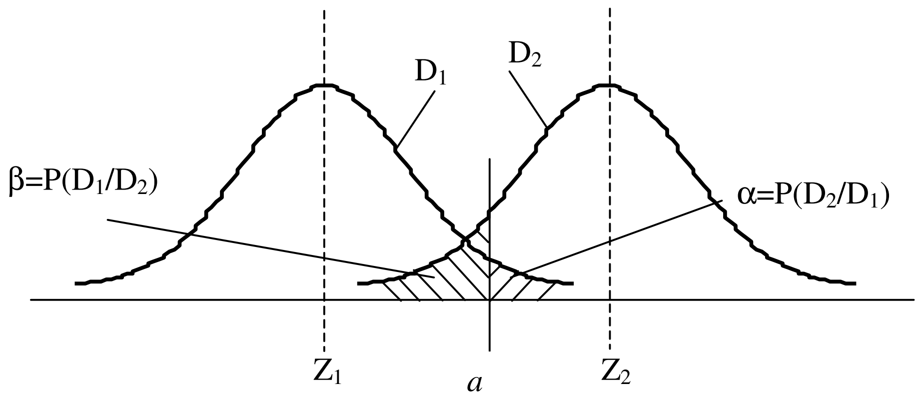

Figure 3. The geometrical interpretatıon of the mis-classification probabilities α and β

Then from (5), (8) and (9) the probability of misclassification equals

$$\alpha = 1 - \Phi\left(\frac{z_2 - z_1}{2\sigma_z}\right) = 1 - \Phi\left(\frac{M}{2}\right) = \Phi\left(-\frac{M}{2}\right)$$

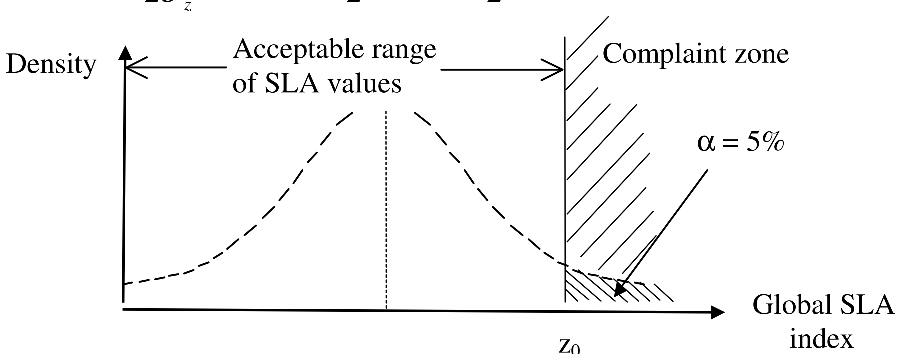

Figure 4. Scheme for SLA conflict resolution

For classification purpose we choose some large enough threshold z_0, e.g. such that Prob{Q>z_0} = 0.05. If the current SLA index value Q is greater than z (Figure 4), then we reach the quality surveillance level and apply for penalties.

2.2. SLA indicator selection

As above, we have two normal populations with known parameters: n –dimensional mean vectors μ_1, μ_2 and common covariance matrix Σ. The problem is to arrange n, given SLA indicators according their efficiency. As a measure of efficiency for the k-th indicator among n given indicators we choose the increment of the Mahalanobis distance (6):

$$M_k^2 = M^2 - M_{-k}^2$$

where the second item M_{-k}^2 is the Mahalanobis distance for $n-1$ indicators (without k th indicator). The less informative indicator is that one which has minimal increment of Mahalanobis distance M_k^2. This is the first step of an SLA indicator selection procedure. By repeating the same algorithm for remaining $n-1$ indicators we obtain the second less informative indicator and so on. After denumeration we obtain the indicators $x_1, x_2, \ldots, x_n$ arranged in their decrease of information. As a basic set for practical use, we may choose the subset of k first indicators that have high enough Mahalanobis distance and respectively low enough misclassification probability.

2.3. SLA optimisation scheme and penalties

The optimal level of quality could be described as a level where the benefit to customers of a small improvement in quality is worth the costs of achieving such an improvement: the aim of efforts concerns in moving to ideal SLA state (say, point "0", see Figure 5).

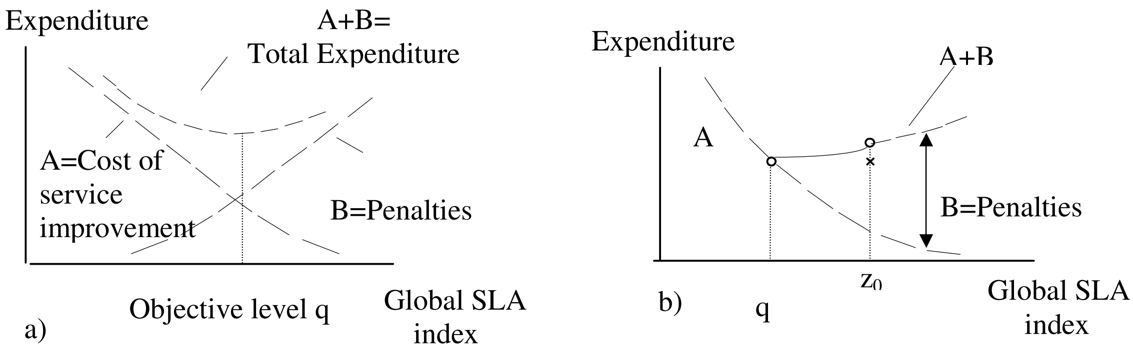

Figure 5. Quality optimisation scheme: a) basic idea, b) penalty scheme

Curve A corresponds to the cost of service level improvement: the higher the quality, i.e. the closer to "0" point, the more costly it becomes. Savings in improvement turn into benefits for telephone operator, but into some economic loss for user. Therefore, we introduce penalty function – curve B: the more savings on service improvement, the higher value of penalties. Objective level of SLA index corresponds to the minimum of total expenditures A+B (at point q). In reality, the penalty function could be defined for SLA values only exceeding the surveillance level z_0 as shown in Fig. 5b, but should be high enough to stimulate the requested SLA improvement and the move to objective level q.

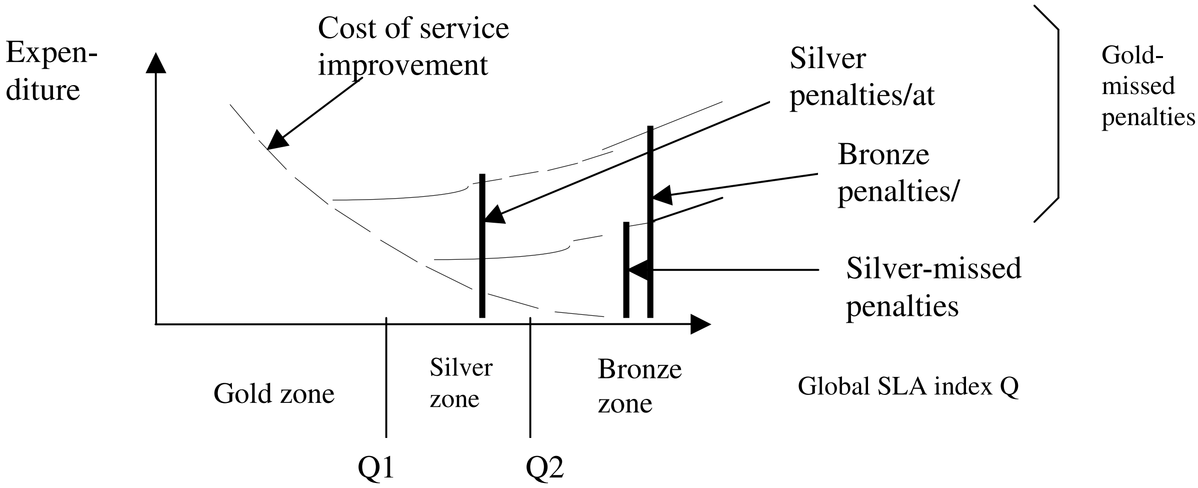

Figure 6. “Gold-silver-bronze” penalty

Figure 6 displays a simple penalty scheme pictured on the basis of Figure 5b. Three types of penalties could be fixed into an SLA document: at gold-agreed service level there are two type penalties: (1) Silver penalties at gold-agreed service and (2) Bronze penalties at gold-agreed service, and (3) bronze level penalties at silver-agreed service.

Comment on interconnection rates and penalties. How to implement “gold-silver-bronze” penalty scheme? How to determine the very penalties, named into Figure 6? One possible solution of this crucial problem follows from the LRAIC methodology. It is widely accepted now that interconnection rates in the future are going to be based on Long Run Average Incremental Costs (LRAIC). The purposes of the LRAIC approach:

1. To make the interconnection charges reflect the actual production costs of an interconnection product and not the costs de facto charged by operators, which are based on the historical development of the networks. This will enable new entrant operators to use existing network facilities without having to pay for inefficiency, misinvestments, etc.
2. In addition, new entrant operators will be stimulated to invest in alternative networks as soon as their business can support it.
3. Finally, the aim is to create consensus on the cost level among telecom operators by involving them in the process.

We propose to consider the SLA as the common target for LRAIC analysis, and consider the SLA features as the border point between bottom-up and top-down LRAIC estimates. Therefore, if the penalty level should be in the strong correlation with the bottom-up cost estimate offered by incumbent company then there is, we hope, much less reasons to over-estimate the LRAIC value.

3. BEST PRACTICE: "VERIZON NEW YORK" SERVICE STANDARDS

The Telephone Service Standards of New York Telephone Company (now Verizon NY) were adopted by the New York State Public Service Commission in 1973 and revised in 1989 and 1991. The Service Standards appear as Part 603 of 16 NYCRR (New York Company Rules and Regulation) and require measurement of service quality in four separate categories:

1) Maintenance Service, 2) Dial-Line Service, 3) Answer Time Performance, and 4) Installation Service.

These measurements are categorised into three levels:

1) *Objective levels* are specifically defined by the standards and described as the level of service that represents good quality service to consumers.

2) On the low end of service scale, the standards incorporate *weakspot levels* to denote a level of service below which immediate analysis and corrective action may be required.

3) Three or more of five consecutive months of weak-spot results are usually considered as *a surveillance level failure*, requiring the filing of a Service Inquiry Report by the serving company and a report to the Commission by the staff.

Figure 7 gives an illustration for data relating to customer trouble report rate divided into three ranges: fully acceptable (Objective level), tolerable (Weakspot level) and unusable (Surveillance level failure).

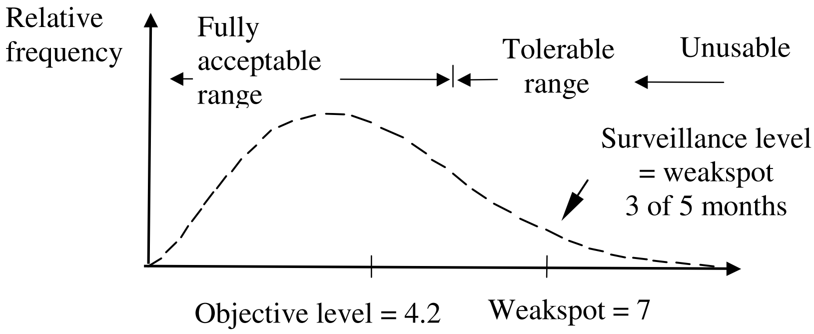

Figure 7. Illustration of customer trouble report rate CTRR

The emergence of measure reduction. Initially - during the period of 1991-1993 service standards contained 20 different measures. The list was reduced to 15 measures for period 1995-1996, then to 11 measures (partly shown in Table 1) and now the list contains only four measures plus Complaints to the New York State Public Service Commission [12].

Service element	Objective level	Weakspot level
Maintenance service		
CTRR per 100 access lines	0.0 - 4.2	Over 7
Missed repair appointments (%)	0.0 - 10.0	Over 15
Out-of-service over 24 hours	0.0 - 20.0	Over 30
Installation performance		
Installations within 5 days (%)	85 - 100	Below 70
Installation appointments	0.0 - 3.0	Over 10
Answering time performance (%)		
Business office - within 20 sec	90.0 - 100.0	Below 85
" - " - all positions busy	0.0 - 10.0	Over 15
Repair service - within 20 sec	90.0 100.0	Below 85
" - " - all positions busy	12.0 - 16.0	Over 27

Table 1. NYT service standards.

Let's consider *Customer Trouble Report Rate (CTRR)* weak-spot level penalties. The Plan includes monthly penalties for CTRR; if any central office exceeds 8.2 reports per 100 lines

(RPHL) during any three-of-five-months period in 1995, a 20% rebate of monthly access fees would be given to all customers in that central office. The target would be 7.6 RPHL in 1996, and 7.0 RPHL in all subsequent years. Besides, there are rebates for all customers in the region (Table 2). For instance, if the company achieves the annual target objective level, say, for 1995 equal to 79%, no rebates would apply. Failure to achieve the target would result in rebates assessed according to the schedule given in Table 2.

	Range of offices without penalties, %			
Target level	79%	81%	83%	85%
Rebate (Million)	1995	1997	1999	2001
$5.0	78%	80	82	84%
$6.0	77	79	81	83
$7.0	76	78	80	82
$8.0	75	77	79	81
$10.0	74	76	78	80
$12.0	73	75	77	79
$15.0	72	74	76	78
$25.0	<72	<74	<76	<78

Table 2. Rebates to all Manhattan customers relating CTRR

4. FUTURE WORK

4.1. Teletraffic Engineering Handbook

Accuracy of QoS measures as a teletraffic theory issue. It is clear from the above how to built the covariance matrix for multivariate analysis on the basis of statistical data. The same could be done on the basis of traditional teletraffic theory – for well known measures: call failure rate, waiting probability, average queue length and others. To develop a multidimensional QoS model we need covariance matrix, i.e. covariance between pairs of measures. This is a new attractive teletraffic theory issue worth of being included in Teletraffic Engineering Handbook [13]. Another topic for the Handbook relates to (1) methodology for "gold-silver-bronze" classification on the basis of LDF approach, (2) SLA contract verification procedure, (3) penalty schemes for missed SLA contracts and (4) the insurance schemes for Force Majeure conditions.

4.2. Gathering around 3GPP

Telecom world is going now to packet switching. The 3rd Generation Partnership Project (3GPP) [14] is today the most promising international telecommunication project and is of special interest for teletraffic community: the above discussed QoS and SLA issues are important for Universal Mobile Telecommunication System (UMTS) and could serve as a framework for future ITCs. In particular, SLA between Radio Access Bearer Service and Core Network Bearer Service is specified by more than 10 parameters including: Maximum bitrate (kbps), Guaranteed bitrate (kbps), Maximum Service Data Unit (SDU) size (octets), etc. And all these many parameters should be collected for each of four different QoS classes: conversational class; streaming class; interactive class; and background class. Therefore we have more than 40 parameters for each service and for each of the interconnecting operators. Is this not too much for SLA conflict resolution between radio access and core network providers?

REFERENCES

1. E. P. G. Wright, ITC-4, London, 1964.
2. ETSI: ETR 138, Dec 1997.
3. V. Paxson et al.: "Framework for IP Performance Metrics". *IETF RFC 2330*, May 1998.
4. AQUILA Project: QoS Architecture. http://www-st.inf.tu-dresden.de/aquila/
5. D. Goderis et al, “Service Level Specification”, <draft-tequila-sls-02.txt>, August 2002.
6. "Service Level Management: Best Practices. White Paper", http://www.cisco.com/warp/
7. http://www.lucent.com/software/docs/SLAManagement.html.
8. M. Schneps-Schneppe and I. Enyukov: Internet access service level agreement – a challenge to teletraffic engineering, pp. 265-275 in proceedings. ITC Specialist Seminar, Girona, 2001.
9. M. Schneps-Schneppe and J. Kraushaar: Global QoS index coordination for interconnection.ITC-15, pp 7-14, ITC Workshop, Washington, D.C., 1997.
10. M. Schneps-Schneppe: Interconnection charges, service quality and penalties from voice telephony to IN and ATM. 8 pp. ITC-16, Workshop, Edinburgh, 1999.
11. I. Enyukov. *Softstat'93*. Ed. by F. Faulbaum. G. Fischer Verlag, Stuttgart, 1994.
12. "*Performance regulation plan for NYC*". New York State PSC, № 95-17, Aug 16, 1995.
13. V. B. Iversen “Teletraffic Engineering Handbook”, ITU-D, Nov 2002. 323 pp.
14. 3GPP TS 23.107 V5.1.0 “3rd Generation Partnership Project; QoS Concept” (Rel5), 2001.
15. ITU-T Recommendation E.860 (06/2002): Framework of a Service Level Agreement.