Scalable Traffic Quality and System Efficiency Indicators Towards Overall Telecommunication System’s QoE Management

Stoyan Poryazov¹(✉), Emiliya Saranova¹², and Ivan Ganchev¹³⁴

¹ Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria
stoyan@math.bas.bg, emiliya@cc.bas.bg
² University of Telecommunications and Post, Sofia, Bulgaria
³ University of Limerick, Limerick, Ireland
ivan.ganchev@ul.ie
⁴ University of Plovdiv “Paisii Hilendarski”, Plovdiv, Bulgaria

Abstract. Conceptual and analytical models of an overall telecommunication system are utilized in this chapter for the definition of scalable indicators towards Quality of Service (QoS) monitoring, prediction, and management. The telecommunication system is considered on different levels – service phase, service stage, network, and overall system. The network itself is presented in seven service stages – A-user, A-terminal, Dialing, Switching, B-terminal Seizure, B-terminal, and B-user, each having its own characteristics and specifics. Traffic quality indicators are proposed on each level. Two network cost/quality ratios are proposed – mean and instantaneous – along with illustrative numerical predictions of the latter, which could be useful for dynamic pricing policy execution, depending on the network load. All defined indicators could be considered as sources for Quality of Experience (QoE) prediction.

Keywords: Overall telecommunication system · Performance model Dynamic quality of service (QoS) · Telecommunication subservices Differentiated QoS subservice indicator · QoS prediction · Human factors of QoS Instantaneous Cost/Quality Ratio · Quality of Experience (QoE)

1 Introduction

Starting from 2010, e.g. [1], a new attitude towards the Quality of Service (QoS) has become dominant, namely to consider QoS and Quality of Experience (QoE) as goods, and the usage of Experience Level Agreement (ELA) [2] has started to be discussed. The importance of the teletraffic models, particularly of the overall QoS indicators, for QoE assessment is emphasized by Fiedler [3]. Until now, however, the usage of performance models of overall telecommunication systems was not very popular. This chapter utilizes the models, elaborated in the Chapter “Conceptual and Analytical Models for Predicting the Quality of Service of Overall Telecommunication Systems” of this book, for the definition of scalable QoS indicators towards overall
telecommunication system’s QoS monitoring, prediction, and management. Some indicators reflect predominantly the users’ experience. All defined indicators depend on human (users’) characteristics and technical characteristics, and may be considered as sources for QoE prediction.

For this, in Sect. 2, traffic characterization of a service in a real device (service phase) is first elaborated. Definitions of served-, carried-, parasitic-, ousted-, and offered carried traffic are proposed, based on the ITU-T definitions, and eight service phase traffic quality indicators are proposed.

In Sect. 3, the service stage concept is developed and corresponding traffic quality indicators are defined.

In Sect. 4, telecommunication system and network efficiency indicators are proposed as follows: eight indicators – on the service stage level, five indicators – on the network level, and three indicators – on the overall system level. The relationship between indicators on the service stage-, network-, and system level are described. A comparison with classical network efficiency indicators is made. The applicability of the approach and results obtained for defining other indicators, as well as for numerical prediction of indicators’ values, is shown.

In Sect. 5, two network cost/quality ratios are proposed – mean and instantaneous – and illustrative numerical predictions of the latter are presented, which may be useful for dynamic pricing policy execution, depending on the network load.

In the Conclusion, possible directions for future research are briefly discussed.

2 Service Phase Concept and Traffic Quality Indicators

The conceptual model utilized in this chapter is described in detail in the Chapter “Conceptual and Analytical Models for Predicting the Quality of Service of Overall Telecommunication Systems” of this book. It consists of five levels: (1) overall telecommunication system and its environment; (2) overall telecommunication network; (3) service stages; (4) service phases; and (5) basic virtual devices. In the following subsections, the concepts of ‘service phase’ and ‘service stage’ are elaborated.

2.1 Service Phase

Based on the ITU-T definition of a service, provided in [4] (Term 2.14), i.e. “A set of functions offered to a user by an organization constitutes a service”, we propose the following definition of a service phase.

Definition 1: The Service Phase is a service presentation containing:

- One of the functions, realizing the service, which is considered indivisible;
- All modeled reasons for ending/finishing this function, i.e. the causal structure of the function;
- Hypothetic characteristics, related to the causal structure of the function (a well-known example of a hypothetic characteristic is the offered traffic concept).
Following the Structural Normalization and Causal Structure approaches described in the Chapter “Conceptual and Analytical Models for Predicting the Quality of Service of Overall Telecommunication Systems” of this book, we may present the service phase in device \( s \) by means of \( k + 1 \) basic virtual causal devices, each representing a different reason for ending this service phase (Fig. 1).

Fig. 1. Traffic characterization of a service phase, represented as device \( S \), by means of \( k + 1 \) basic virtual causal devices.

In Fig. 1, only one causal device represents successful completion of the service in device \( s \) – with carried traffic \((crr.Y_s)\), whereas the remaining causal devices represent \( k \) different reasons for unsuccessful ending of the service – respectively with traffics \( prs.Y_1, prs.Y_2, \ldots, prs.Y_k \).

Generalizing, for more precise traffic characterization in a pool of resources, we propose the following definitions.

**Definition 2:** The Served Traffic in a pool of resources is the traffic, occupying (using) resources in the pool.

In Fig. 1, the served traffic in device \( s \) \((srv.Y_s)\) is the following sum:

\[
srv.Y_s = prs.Y_1 + rs.Y_2 + \ldots + prs.Y_k + crr.Y_s. \tag{1}
\]

**Definition 3:** The Carried Traffic in a pool of resources is the traffic, which was successfully served in the pool (and carried to the next service phase).

In Fig. 1, the carried traffic in device \( s \) is \( crr.Y_s \).

**Definition 4:** The Parasitic Traffic in a pool of resources is the traffic, which was unsuccessfully served in the pool.

---

1 In the expressions, formulas and figures, the sign \( . \) is used only as a separator and NOT as a sign of multiplication. The multiplication operation is indicated by a gap between multiplied variables, e.g. \( X Y \).
In Fig. 1, each of traffics \( prs.Y1, prs.Y2, \ldots, prs.Yk \) is a parasitic one. Parasitic traffic occupies real resources but not for an effective service execution.

In Definitions 2 and 3, the served- and carried traffic are different terms, despite the ITU-T definition of the carried traffic as “The traffic served by a pool of resources” ([5], Term 5.5). We believe that this distinction leads to a better and more detailed traffic- and QoS characterization.

**Definition 5:** The Ousted Traffic is the traffic that would be carried, if there is no unsuccessful service ending in the pool of resources.

In Fig. 1, each parasitic traffic \( prs.Y1, prs.Y2, \ldots, prs.Yk \) ousted a corresponding traffic that would be carried, if there is no unsuccessful service ending of the corresponding type: \( ous.Y1, ous.Y2, \ldots, ous.Yk \). The flow intensity to a parasitic device and the corresponding ousted device is the same by definition, i.e.:

\[
ous.Fi = prs.Fi, \text{ for } i = [1, k],
\]

but the service times are different.

The hypothetic service time for every ousted device (\( ous.Ti \)) equals the carried service time (\( crr.Ts \)):

\[
ous.Ti = crr.Ts, \text{ for } i = [1, k].
\]

The ousted traffic is a hypothetic one with the following intensity:

\[
ous.Yi = prs.Fi \cdot crr.Ts, \text{ for } i = [1, k].
\]

### 2.2 Causal Generalization

In the Chapter “Conceptual and Analytical Models for Predicting the Quality of Service of Overall Telecommunication Systems” of this book, the causal presentation and causal aggregation are discussed. The causal aggregation is understood as an aggregation of all cases in the model, corresponding to different reasons for service ending (referred to as unsuccessful cases further in this chapter).

Here a causal generalization is proposed, as an aggregation of all unsuccessful cases (\( prs.Ys \)). Besides this, an aggregation of all cases of ousted traffic (\( ous.Ys \)) could be used:

\[
prs.Ys = \sum_{i=1}^{k} prs.Yi;
\]

\[
ous.Ys = \sum_{i=1}^{k} ous.Yi.
\]

By Definition 2, the served traffic is a sum of the parasitic and carried traffic (c.f. Fig. 1):
If the system is considered as being in a stationary state, by using the Little’s formula [6] we have: \( \text{prs.Ys} = \text{prs.Fs} \text{prs.Ts} \) and \( \text{crr.Ys} = \text{crr.Fs} \text{crr.Ts} \). Hence:

\[
\text{srv.Ys} = \text{srv.Fs} \text{srv.Ts} = \text{prs.Fs} \text{prs.Ts} + \text{crr.Fs} \text{crr.Ts}.
\] (9)

Formulas (7), (8), and (9) illustrate the advantage of the traffic qualifiers – the notation is invariant to the number of cases considered in a service phase.

### 2.3 Offered Carried Traffic

**Definition 6:** The Offered Carried Traffic \((\text{ofr.crr.Ys})\) in a pool \(s\) of resources is the sum of the carried traffic \((\text{crr.Ys})\) and ousted traffic \((\text{ous.Ys})\) in the pool:

\[
\text{ofr.crr.Ys} = \text{ous.Ys} + \text{crr.Ys}.
\] (10)

From (10), (6), (4), (8), \( \text{prs.Fs} = \sum_{i=1}^{k} \text{prs.Fi} \) and \( \text{crr.Ys} = \text{crr.Fs} \text{crr.Ts} \), the following formula could be obtained:

\[
\text{ofr.crr.Ys} = \text{srv.Fs} \text{crr.Ts}.
\]

Definition 6 is analogous to the ITU-T definition of an Equivalent Offered Traffic [7] but considers the traffic related to the carried call attempts, whereas the ITU-T definition considers the traffic that would be served.

### 2.4 Traffic Quality Indicators

**Indicator 1:** Offered Carried Traffic Efficiency – the ratio of the carried traffic, in a service phase, to the offered carried traffic:

\[
I_1 = \frac{\text{crr.Ys}}{\text{ofr.crr.Ys}} = 1 - \frac{\text{ous.Ys}}{\text{ofr.crr.Ys}}.
\] (11)

**Indicator 2:** Causal Ousted Importance – the ratio of the ousted traffic due to reason \(i\) \((\text{ous.Yi})\) to the offered carried traffic of a service phase \((\text{ofr.crr.Ys})\):

\[
I_2(i) = \frac{\text{ous.Yi}}{\text{ofr.crr.Ys}}.
\] (12)

This indicator allows the estimation of missed benefits due to reason \(i\) and therefore the necessity of countermeasures against this reason.

**Indicator 3:** Ousted Traffic Importance – the sum of all causal ousted importance indicators of a service phase. From Fig. 1, and Formulas (6) and (11), it is:
\[ I_3 = \sum_{i=1}^{k} I_2(i) = \sum_{i=1}^{k} \frac{ous.Yi}{ofr.crr.Ys} = \frac{ous.Ys}{ofr.crr.Ys} = 1 - \frac{crr.Ys}{ofr.crr.Ys}. \]  

**Indicator 4: Service Efficiency** – the ratio of the carried traffic to the served traffic:

\[ I_4 = \frac{crr.Ys}{srv.Ys} = 1 - \frac{prs.Ys}{srv.Ys}. \]  

**Indicator 5: Causal Parasitic Importance** – the ratio of the parasitic traffic due to reason \( i \) (\( prs.Yi \)) to the served traffic of a service phase (\( srv.Ys \)):

\[ I_5(i) = \frac{prs.Yi}{srv.Ys}. \]  

This indicator allows the estimation of an ineffective service due to a reason and therefore the necessity of countermeasures against this reason.

**Indicator 6: Parasitic Traffic Importance** – the sum of all causal parasitic importance indicators of a service phase. From Fig. 1, and Formulas (5) and (14), it is:

\[ I_6 = \sum_{i=1}^{k} I_5(i) = \sum_{i=1}^{k} \frac{prs.Yi}{srv.Ys} = \frac{prs.Ys}{srv.Ys} = 1 - \frac{crr.Ys}{srv.Ys}. \]  

**Indicator 7: Ousted/Parasitic Traffic Ratio** – this is the ratio of the ousted traffic to the parasitic traffic:

\[ I_7 = \frac{ous.Ys}{prs.Ys}. \]  

This indicator estimates the aggregated, by all reasons, ratio of missed benefits to the ineffective service in a service phase.

**Indicator 8: Causal Ousted/Parasitic Traffic Ratio** – this is the ratio of the ousted traffic, due to reason \( i \), to the parasitic traffic due to the same reason. From Definition 5 and Formula (2):

\[ I_8(i) = \frac{ous.Yi}{prs.Yi} = \frac{ous.Ti}{prs.Ti}. \]  

This indicator gives another important estimation of a reason for ineffective service in a service phase.

### 3 Service Stage Concept and Traffic Quality Indicators

**Definition 7:** The Service Stage is a service presentation containing:

- One service phase, realizing one function of the service;
• All auxiliary service phases that directly support this function realization but are not part of the realized function itself.

Examples of auxiliary service phases are the entry, exit, buffer, and queue virtual devices. The performance of the auxiliary devices depends directly on the service phase, realizing a function of the service.

The service stage concept allows the division of the overall telecommunication service into subservices and therefore makes easier the system modeling process.

3.1 Service Stage

For simplicity in this subsection, the simplest possible service stage, consisting of only two service phases, is considered (Fig. 2). For more complex service stages with more phases, please refer to the Chapter “Conceptual and Analytical Models for Predicting the Quality of Service of Overall Telecommunication Systems” of this book.

![Fig. 2. A service stage g, consisting of Entrance and Service phases.](image)

The service stage $g$, in Fig. 2, consists of Entrance and Service phases (represented by corresponding virtual devices). The Entrance device (e) may check the service request (call) attempt for having the relevant admission rights, whereas the Service device (s) checks for service availability or existence of free service resources, etc. Let $ofr.Fg$ is the flow intensity of the service request attempts offered to this stage, $crr.Fg$ – the intensity of the outgoing carried flow, $prs.Fg$ – the flow intensity of the parasitic served requests, and $prs.Fe$ – the intensity of the parasitic call attempts flow. Then from Fig. 2 we have:
\[ \text{prs.Fe} = \text{ofr.Fg \ \text{prs.Pe}}, \]  

(19)

where \( \text{prs.Pe} \) is the probability of directing the service request attempts to the generalized parasitic service in device \( e \). By analogy:

\[ \text{prs.Fs} = \text{ofr.Fg} \ (1 - \text{prs.Pe}) \ \text{prs.Ps}. \]  

(20)

The total parasitic flow in service stage \( g \) is:

\[ \text{prs.Fg} = \text{prs.Fe} + \text{prs.Fs}. \]  

(21)

The carried traffic \((\text{crr.Yg})\) in service stage \( g \) is a sum of the carried traffic in devices \( e \) and \( s \):

\[ \text{crr.Yg} = \text{crr.Ye} + \text{crr.Ys} = \text{crr.Fe} \ \text{crr.Te} + \text{crr.Fs} \ \text{crr.Ts}, \]  

(22)

where:

\[ \text{crr.Fe} = \text{ofr.Fg} \ (1 - \text{prs.Pe}); \]  

(23)

\[ \text{crr.Fs} = \text{ofr.Fg} \ (1 - \text{prs.Pe}) \ (1 - \text{prs.Ps}). \]  

(24)

The total carried traffic in service stage \( g \) is:

\[ \text{crr.Yg} = \text{ofr.Fg} \ (1 - \text{prs.Pe}) \ (\text{crr.Te} + (1 - \text{prs.Ps}) \ \text{crr.Ts}). \]  

(25)

The estimation of the carried traffic in a service stage could be problematic due to the fact that some of the carried service requests attempts in the first device \( e \) are not carried to the next device \( s \), i.e. they become parasitic service requests with probability \( \text{prs.Ps} \) (c.f. Fig. 2).

Based on the ITU-T definition of ‘effective traffic’ [5], i.e. as “The traffic corresponding only to the conversational portion of effective call attempts”, we propose here the Effective Carried Traffic concept.

**Definition 8:** The Effective Carried Traffic in a service stage is the traffic corresponding to the service request attempts leaving the stage with a fully successful (carried) service.

In Fig. 2, the effective carried traffic \((\text{eff.crr.Yg})\) of service stage \( g \) is:

\[ \text{eff.crr.Yg} = \text{eff.crr.Fg} \ \text{eff.crr.Tg}, \]  

(26)

where:

\[ \text{eff.crr.Fg} = \text{crr.Fs} = \text{ofr.Fg} \ (1 - \text{prs.Pe})(1 - \text{prs.Ps}); \]  

(27)

\[ \text{eff.crr.Tg} = \text{crr.Te} + \text{crr.Ts}. \]  

(28)

From (26), (27), and (28), we obtain:

\[ \text{eff.crr.Yg} = \text{ofr.Fg} \ (1 - \text{prs.Pe})(1 - \text{prs.Ps})(\text{crr.Te} + \text{crr.Ts}). \]  

(29)
Note the difference between (25) and (29), i.e. in general, in a service stage, the effective carried traffic is less than the carried traffic.

The **Offered Traffic** is a fundamental teletraffic engineering concept. We use the ITU-T definition of the Equivalent Offered Traffic [7], i.e. “Offered traffic, to a pool of resources, is the sum of carried and blocked traffic of this pool”.

The blocked traffic corresponds to the blocked attempts, as per Definition 2.8 in [5]: “Blocked call attempt: A call attempt that is rejected owing to a lack of resources in the network”. This definition, however, is too narrow to be applied directly to blocked service request attempts as it does not include most of the reasons for rejection, including access control, service unavailability, called terminal unavailability or busyness, and many others. Thus we propose the following extension of it.

**Definition 9:** The Blocked Service Request Attempt is a service request attempt with rejected service, in the intended pool of resources, due to any reason.

Blocked traffic is a service stage concept because it considers blocking of service requests before entering the service phase, or in other words, blocking that occurs in another virtual device before the corresponding service device.

In Fig. 2, blocking occurs in the Entrance device. The blocked traffic \(blc.Ys\) corresponds to the service request attempts offered to service stage \(ofr.Fg\), but not belonging to the served attempts \(srv.Fs\). From Fig. 2 and the Little’s theorem, we obtain:

\[
blc.Ys = blc.Fs \cdot blc.Ts. \tag{30}
\]

The service request attempts that are not carried in phase \(e\), and hence are rejected to the next service phase, are considered parasitic in phase \(e\). For the intensity of the blocked attempts, the following equality holds:

\[
blc.Fs = prs.Fe = ous.Fe = ofr.Fg \cdot prs.Pe. \tag{31}
\]

The offered traffic is a hypothetic one “that would be served” if it is not blocked, and therefore:

\[
blc.Ts = srv.Ts \tag{32}
\]

From (30), (31), and (32), we obtain the following formula:

\[
blc.Ys = ofr.Fg \cdot prs.Pe \cdot srv.Ts, \tag{33}
\]

which is valid for the generalized reason for service request attempts rejection in service phase \(e\) (i.e. in the Entrance device).

From the definition of the equivalent offered traffic, Fig. 2, and Formulas (23), (33), and \(srv.Fs = crr.Fe\), the traffic offered to the service device \(s\) is:

\[
ofr.Ys = blc.Ys + srv.Ys = ofr.Fg \cdot srv.Ts. \tag{34}
\]
3.2 Traffic Quality Indicators

Many of the service-stage traffic quality indicators may be reformulated as service-stage performance indicators as done below.

**Indicator 9: Carried Effectiveness of a Service Stage** – the ratio of the effective carried traffic to the carried traffic:

\[ I_9 = \frac{\text{eff.crr}.Y}{\text{crr}.Y}. \]  

(35)

For instance, from (25) and (29), the Carried Effectiveness of service stage \( g \) in Fig. 2 is:

\[
\frac{\text{eff.crr}.Y_g}{\text{crr}.Y_g} = \frac{(1 - \text{prs}.Ps)(\text{crr}.Te + \text{crr}.Ts)}{\text{crr}.Te + (1 - \text{prs}.Ps) \text{crr}.Ts}.
\]

(36)

4 Telecommunication System and Network Efficiency Indicators

4.1 Telecommunication System QoS Concept

Users are shown in “Fig. 1 – Schematic contributions to end-to-end QoS” in [4] but they are not connected to the network. In Fig. 3, schematic contributions to QoS in an overall telecommunication system, including users, is presented in more detail.

![Fig. 3. Schematic contributions to QoS in an overall telecommunication system, including users.](image)

In Fig. 3, the calling (A) and called (B) users and terminals, as well as the main service stages of the service request (call) attempts in a telecommunication network, are presented. The telecommunication network is usually presented as having five service stages: A-terminal, Dialing, Switching, B-terminal Seizure, and B-terminal, where each
service stage has its own characteristics. However, there are two other stages – A-user and B-user –, with their own specifics.

In Fig. 3, the possible paths of service request (call\(^2\)) attempts are the following:

1. \textit{int.}\(\text{Fa}\): The calling users (A-users) generate intent call attempts, with intensity \(\text{int.}\(\text{Fa}\), represented as a Generate device in the A-User block. Call intent is “The desire to establish a connection to a user”. “This would normally be manifested by a call demand. However, demands may be suppressed or delayed by the calling user’s expectation of poor Quality of Service performance at a particular time” [5].

2. \textit{sup.}\(\text{Fa}\): The intensity of suppressed intent call attempts. Suppressed traffic is “The traffic that is withheld by users who anticipate a poor quality of service (QoS) performance” [5]. “At present, suitable algorithms for estimating suppressed traffic have not been defined” [7].

3. \textit{dem.}\(\text{Fa}\): The intensity of demand call attempts. Call demand is: “A call intent that results in a first call attempt” [5].

4. \textit{rep.}\(\text{Fa}\): The intensity of repeated call attempts. Repeated call attempt is: “Any of the call attempts subsequent to a first call attempt related to a given call demand. NOTE – Repeated call attempts may be manual, i.e. generated by humans, or automatic, i.e. generated by machines” [7].

5. \textit{ofr.}\(\text{Fa}\): The intensity of all call attempts (demand and repeated) trying to occupy A-terminals. A-terminals are considered as the first service stage (c.f. Sect. 3) in the telecommunication network. From Fig. 3:

\[ \text{ofr.}\(\text{Fa}\) = \text{dem.}\(\text{Fa}\) + \text{rep.}\(\text{Fa}\). \] (37)

6. \textit{prs.}\(\text{Fa}\): The intensity of all parasitic (unsuccessfully served, c.f. Sect. 2) call attempts in A-terminals.

We are modeling the system in a stationary state and for each considered service stage the intensity of the offered call attempts equals the sum of the outgoing parasitic and carried flows, e.g. \(\text{ofr.}\(\text{Fa}\) = \text{prs.}\(\text{Fa}\) + \text{crr.}\(\text{Fa}\).

For each service stage, part of the parasitic attempts are terminated by the A-user (c.f. devices of type ‘terminator’ in Fig. 3) and the rest join the repeated attempt’s flow (\(\text{rep.}\(\text{Fa}\)).

7. \textit{ofr.}\(\text{Fd}\) = \text{crr.}\(\text{Fa}\): The intensity of carried (in A-terminals) call attempts (\(\text{crr.}\(\text{Fa}\)) is equal to the intensity of the offered call attempts (\(\text{ofr.}\(\text{Fd}\)) to the Dialing stage in the network.

8. \textit{prs.}\(\text{Fd}\): The intensity of all parasitic (unsuccessfully served, c.f. Sect. 2) call attempts in the Dialing stage.

9. \textit{ofr.}\(\text{Fs}\) = \text{prs.}\(\text{Fs}\) + \text{crr.}\(\text{Fs}\): The intensity of the offered-, parasitic-, and carried flows of call attempts of the Switching stage.

\(^2\) Throughout the rest of this chapter, the term ‘call’ should be interpreted in a broader meaning of a ‘service request’.
10. \[ \text{oifr}.F_z = \text{prsr}.F_z + \text{crrr}.F_z: \text{The intensity of the offered-, parasitic-, and carried flows of call attempts of the 'B-terminal seizure' stage. The intended B-terminal may be busy or unavailable and this will cause blocking of call attempts.} \]

11. \[ \text{oifr}.F_b = \text{prsr}.F_b + \text{crrr}.F_b: \text{The intensity of the offered-, parasitic-, and carried flows of call attempts of the B-terminal stage.} \]

12. \[ \text{oifr}.F_{bu} = \text{prsr}.F_{bu} + \text{crrr}.F_{bu}: \text{The intensity of the offered-, parasitic-, and carried flows of call attempts of the B-user stage. The B-user may be absent, busy, tired, etc.} \]

### 4.2 Efficiency Indicators

The efficiency indicators, proposed in this chapter, are considered on five levels: (1) service phase; (2) service stage; (3) part of network; (4) overall network; and (5) overall telecommunication system.

#### 4.2.1 Proposed Efficiency Indicators on Service Stage Level

In each service stage, a basic performance indicator is the ratio between intensities of the carried flow and offered flow of call attempts. An exception is the A-User stage because there are two sub-stages in it – Ai (considering the intent call attempts) and Ad (considering the demand call attempts).

**Indicator 10: Efficiency indicator \( Q_{ai} \) on the Ai sub-stage:**

\[
I_{10} = Q_{ai} = \frac{\text{dem}.F_a}{\text{int}.F_a} \tag{38}
\]

**Indicator 11: Efficiency indicator \( Q_{ad} \) on the Ad sub-stage.**

Let \( Pr \) is the aggregated probability of repetition of the offered (to the A-terminals) call attempts:

\[
Pr = \frac{\text{rep}.F_a}{\text{oifr}.F_a}. \tag{39}
\]

From (37) and (39), the following formula could be obtained for the efficiency indicator \( Q_{ad} \):

\[
I_{11} = (1 - Pr) = \frac{\text{dem}.F_a}{\text{oifr}.F_a} = \frac{\text{dem}.F_a}{\text{dem}.F_a + \text{rep}.F_a} = \frac{1}{\beta} = Q_{ad}, \tag{40}
\]

where \( \beta \) is defined in [7] as:

\[
\beta = \frac{\text{All call attempts}}{\text{First call attempts}}. \tag{41}
\]

In (40), \( Q_{ad} \) is de-facto the probability corresponding to the ratio of the primary (demand) call attempts’ intensity to the offered attempts’ intensity. It may be considered as an aggregated overall network performance indicator (as per the initial attempt in [8]).
Indicator 12: Efficiency indicator $Q_a$ on the A-terminal stage:

$$I_{12} = Q_a = \frac{crr.Fa}{ofr.Fa}.$$  (42)

Indicator 13: Efficiency indicator $Q_d$ on the Dialing stage:

$$I_{13} = Q_d = \frac{crr.Fd}{ofr.Fd}.$$  (43)

Indicator 14: Efficiency indicator $Q_s$ on the Switching stage:

$$I_{14} = Q_s = \frac{crr.Fs}{ofr.Fs}.$$  (44)

Indicator 15: Efficiency indicator $Q_z$ on the ‘B-terminal Seizure’ stage:

$$I_{15} = Q_z = \frac{crr.Fz}{ofr.Fz}.$$  (45)

Indicator 16: Efficiency indicator $Q_b$ on the B-terminal stage:

$$I_{16} = Q_b = \frac{crr.Fb}{ofr.Fb}.$$  (46)

Indicator 17: Efficiency indicator $Q_{bu}$ on the B-user stage:

$$I_{17} = Q_{bu} = \frac{crr.Fbu}{ofr.Fbu}.$$  (47)

4.2.2 Proposed Efficiency Indicators on Network Level

Network efficiency indicators estimate QoS characteristics of portions of the network comprising more than one service stages, or the overall network. In this subsection, as usually, the indicated network portion begins with the starting points of the network and ends in another network point of interest. All network efficiency indicators are fractions with denominators offered to the A-terminals’ flow intensity $ofr.Fa$.

The classic network efficiency indicators are the following three, e.g. as defined in [9]:

1. **Answer Seizure Ratio (ASR)** = (number of seizures that result in an answer signal)/(the total number of seizures)” … “Measurement of ASR may be made on a route or on a destination code basis” … “A destination can be a mobile network, a country, a city, a service, etc.” [9].

2. **Answer Bid Ratio (ABR)** = (number of bids that result in an answer signal)/(total number of bids); ABR is similar to ASR except that it includes bids that do not result in a seizure” [9].
3. “Network Effectiveness Ratio (NER):” NER is designed to express the ability of networks to deliver calls to the far-end terminal. NER expresses the relationship between the number of seizures and the sum of the number of seizures resulting in either an answer message, or a user busy, or a ring no answer, or in the case of ISDN a terminal rejection/unavailability. Unlike ASR, NER excludes the effects of customer behavior and terminal behavior” [9].

These classic efficiency indicators reflect network providers’ attitude but don’t consider the possibilities for initiated but unsuccessful communication as well as the influence of repeated attempts.

Below we propose new network efficiency indicators, all having as an index the first letter of the last service stage considered.

**Indicator 18:** Network efficiency indicator $E_a$ on the A-terminal stage, c.f. also (42):

$$ I_{18} = E_a = \frac{cr.r.Fa}{ofr.Fa} = Qa = I_{12}. $$

**Indicator 19:** Network efficiency indicator $E_d$ on the Dialing stage, c.f. also (43):

$$ I_{19} = E_d = \frac{cr.r.Fd}{ofr.Fa} = Qa Qd = I_{12} I_{13}. $$

**Indicator 20:** By taking into account that $cr.r.Fd = ofr.Fs$, c.f. Fig. 3 and (44), the network efficiency indicator $E_s$ on the Switching stage is:

$$ I_{20} = E_s = \frac{cr.r.Fs}{ofr.Fa} = Qa Qd Qs = I_{14} I_{19}. $$

**Indicator 21:** By taking into account that $cr.r.Fs = ofr.Fz$, c.f. Fig. 3 and (45), the network efficiency indicator $E_z$ on the ‘B-terminal seizure’ stage is:

$$ I_{21} = E_z = \frac{cr.r.Fz}{ofr.Fa} = Qa Qd Qs Qz = I_{15} I_{20}. $$

**Indicator 22:** By taking into account that $cr.r.Fz = ofr.Fb$, c.f. Fig. 3 and (46), the network efficiency indicator $E_b$ on the B-terminal stage is:

$$ I_{22} = E_b = \frac{cr.r.Fb}{ofr.Fa} = Qa Qd Qs Qz Qb = I_{16} I_{21}. $$

This indicator corresponds to the cases of B-user answers, but does not consider the successfulness of the communication.

### 4.2.3 Proposed Efficiency Indicators on Overall System Level

**Indicator 23:** By taking into account that $cr.r.Fb = ofr.Fbu$, c.f. Fig. 3 and (47), the system efficiency indicator $E_{bu}$ on the B-user stage is:
\[ I_{23} = Ebu = \frac{crr.Fbu}{oFr Fa} = Qa Qd Qs Qz Qb Qbu = I_{17} I_{22}. \] (53)

This indicator corresponds to the cases of fully successful communication, from the users’ point of view, regarding all call attempts offered to the network.

**Indicator 24: System efficiency indicator** \( E_u \) on the Ad sub-stage, c.f. also (40):

\[ I_{24} = Eu = Qad Ebu = Qad Qa Qd Qs Qz Qb Qbu = I_{11} I_{23}. \] (54)

This indicator corresponds to the cases of fully successful communication, from the A-users’ point of view, regarding demand call attempts. It shows what part of the first (demand) attempts is fully successful. It may be called ‘Demand Efficiency’. It is a user-oriented indicator, compounding explicitly repeated attempts, connection and communication parameters.

**Indicator 25: System efficiency indicator** \( E_i \) on the Ai sub-stage, c.f. also (38):

\[ I_{25} = Ei = Qai Eu = Qai Qad Qa Qd Qs Qz Qb Qbu = I_{10} I_{24}. \] (55)

This indicator corresponds to the cases of fully successful communication, from the A-users’ point of view, regarding intent call attempts. It shows what part of the intent attempts is fully successful. It is very difficult to measure \( E_i \) directly because suppressed attempts (forming the demands w.r.t. point 2 in Subsect. 4.1) can’t reach the network and therefore can’t be measured there.

### 4.3 Approach Applicability and Results

Most of the proposed indicators are flow-oriented as they take into account the flow intensities. Flow-oriented indicators are in the core of time- and traffic-oriented indicators. In this subsection, numerical results for some of the proposed flow indicators and other time- and traffic-oriented indicators, built on their basis, are presented. An analytical model of the overall telecommunication system, corresponding to Fig. 3, is used. Methods of building such models are described in the Chapter “Conceptual and Analytical Models for Predicting the Quality of Service of Overall Telecommunication Systems” of this book.

The numerical results are presented for the entire theoretical network-traffic-load interval, i.e. the terminal traffic of all A- and B-terminals \( Y_{ab} \) is ranging from 0% to 100% of the number \( N_{ab} \) of all active terminals in the network. The input parameters are the same, excluding the capacity of the network (the number of the equivalent connection lines), given as a percentage of all terminals in the system. Differences in the network capacity cause different blocking probabilities due to resource insufficiency. Three cases have been considered:

- **Case 1**: Without repeated service request attempts and without blocking;
- **Case 2**: With repeated service request attempts but without blocking;
- **Case 3**: With repeated service request attempts and with blocking.
Figures 4, 5, 6 present numerical results obtained for some of the proposed efficiency indicators, whereas Figs. 7, 8, 9 present numerical results obtained for some time- and traffic-oriented indicators.

**Fig. 4.** Efficiency indicators $Q_{ad}$, $E_{bu}$ and $E_u$ for Case 1 ($Q_{ad} = 1$ and $E_{bu} = E_u$ because there are no repeated service request attempts in the system).

**Fig. 5.** Efficiency indicators $Q_{ad}$, $E_{bu}$ and $E_u$ for Case 2 (the network performance is degraded considerably due to repeated service request attempts).
Fig. 6. Efficiency indicators $Qad$, $Ebu$ and $Eu$ for Case 3 (the network performance degrades sharply due to blocking).

Fig. 7. Time and traffic AB-efficiency for Case 1 ($(Ebu \ Tcc) / Tab = (Eu \ Tcc) / Tab$, because there are no repeated service request attempts).
Fig. 8. Time and traffic AB-efficiency for Case 2 \((\text{Eu Tcc}) / \text{Tab}\) is sensitive to repeated service request attempts in contrast to \((\text{Ebu Tcc}) / \text{Tab}\), which is not).

Fig. 9. Time and traffic AB-efficiency for Case 3 (the indicator \(\text{paid Traffic} / \text{Yab}\) is not sensible enough in the network-load interval without blocking).

5 Network Cost/Quality Ratios

We consider the overall telecommunication system model, presented in Fig. 3, with the following assumptions:
Assumption 1: The observation time interval $\Delta t$ is limited;
Assumption 2: The full system costs ($SC$) in the time interval $\Delta t$ are known;
Assumption 3: The cost/quality ratio depends on the paid volume of traffic ($paid.V$) in this time interval and the QoS indicator ($Q$);
Assumption 4: The full system costs ($SC$) don’t depend considerably on the served traffic volume in the time interval $\Delta t$;
Assumption 5: The QoS indicator ($Q$) is dimensionless with values from the interval $(0,1]$ and is proportional to the quality ($Q = 1$ means ‘the best quality’).

5.1 Mean Cost/Quality Ratio

Based on these assumptions and the definition of the traffic volume, i.e. “The traffic volume in a given time interval is the time integral of the traffic intensity over this time interval” [5], the ‘Cost per Unit’ quantity is:

$$\text{Cost per Unit} = \frac{\text{Full System's Costs [Euro]}}{\text{Paid Traffic Volume [Erlang} \times \Delta t]}.$$  \hspace{1cm} (56)

By dividing this to the QoS indicator ($Q$), we obtain the following:

$$\frac{\text{Cost per Unit}}{\text{Quality}} = \frac{\text{Full System's Costs [Euro]}}{Q \ paid.V [\text{Erlang} \times \Delta t]} = \frac{SC}{Q \ paid.V}.$$  \hspace{1cm} (57)

The definition of the paid traffic may depend on the telecommunication service provider. The estimation of the paid traffic volume is a routine operation (c.f. ITU-T Recommendations Series D: General Tariff Principles).

The definition of the QoS indicator ($Q$) may differ from users’ perspective (i.e. as a generalized QoE parameter) to the telecommunication service provider’s perspective. In general, the best is to include the $Q$ definition in the Service Level Agreement (SLA). In any case, the value of the QoS indicator ($Q$) in (57) is the mean value in the time interval considered.

The mean cost/quality ratio (57) is suitable for relatively long intervals – days, months, years.

5.2 Instantaneous Cost/Quality Ratio

We consider the traffic intensity ($Y$) as per the ITU-T definition, i.e. “The instantaneous traffic in a pool of resources is the number of busy resources at a given instant of time” [5]. From assumptions made and (57), the following formula could be obtained:

$$\frac{\text{Cost per Unit}}{\text{Quality}} = \frac{\text{Full System's Costs [Euro]}}{\Delta t [\text{Time}] \ paid.Y [\text{Erlang}]} = \frac{SC}{\Delta t \ Q \ paid.Y}.$$  \hspace{1cm} (58)

The paid traffic intensity ($paid.Y$) is an instantaneous quantity but the ratio ‘Cost per Unit/Quality’ (58) depends on the time interval duration. We define the ‘System’s Costs
Intensity’ (SCI) parameter, independent of the time interval duration (but dependent of the interval position in the service provider’s life time), as per the following formula:

\[
\text{System’s Costs Intensity (SCI)} = \frac{\text{Full System’s Costs [Euro]}}{\Delta t \ [\text{Time}]} = \frac{SC}{\Delta t}. \quad (59)
\]

The System’s Costs Intensity (SCI) parameter allows defining a new useful parameter – the Normalized Cost/Quality Ratio (NCQR):

\[
\text{Normalized Cost/Quality Ratio (NCQR)} = \frac{1}{Q \ \text{paid}. Y \ [\text{Erlang}]). \quad (60)
\]

The Normalized Cost/Quality Ratio (NCQR) is independent of the absolute system’s costs amount. It is normalized, because it is the cost/quality ratio per 1 Euro cost. From (57), (58), and (59), we obtain:

\[
\frac{\text{Cost per Unit}}{\text{Quality}} = \frac{\text{Full System’s Costs [Euro]}}{\Delta t \ [\text{Time}] \times \frac{1}{Q \ \text{paid}. Y \ [\text{Erlang}]} = SCI \ NCQR \quad (61)
\]

The proposed quantities SCI and NCQR allow the estimation of the cost/quality ratio for every suitable (paid) time interval with a relatively short duration, e.g. seconds, minutes, hours.

The paid traffic intensity depends on the network traffic load. In any case, the instantaneous values of the QoS indicator (Q) depend on many factors, including the network load.

The expressions (57) and (58) are similar (the mean value of the instantaneous indicator, in \(\Delta t\), gives the value of the Mean Cost/Quality Ratio indicator in \(\Delta t\)), but the methods for their estimation and usage are different.

The Instantaneous Cost/Quality Ratio may be useful for dynamic pricing policies, depending on the network load. Related works on this subject were not found in the literature.

### 5.3 Prediction of Instantaneous Cost/Quality Ratio

An advantage of the Normalized Cost/Quality Ratio (NCQR) is its independence of the absolute system’s costs amount. This allows separation of the estimations for NCQR and System’ Cost Intensity (SCI). In this subsection, we estimate NCQR using the telecommunication system model described in the Chapter “Conceptual and Analytical Models for Predicting the Quality of Service of Overall Telecommunication Systems” of this book.

We consider Q as an overall telecommunication system’s QoS indicator. Each of the described indicators on the overall system level (c.f. Subsect. 4.2.3) may be used. Numerical examples below are for the indicator \(Q = Ebu\) c.f. (53). This corresponds to cases of fully successful communication, from the users’ point of view, regarding all call attempts, offered to the network. Sometimes it is called “Network Call Efficiency”.

As a paid traffic, the successful communication (carried) traffic is used:

\[ NCQR = \frac{1}{Ebu \text{ paid}.Y}. \]  

(62)

The values of input parameters of human behavior and technical system, to the model, in the presented output numerical results are typical for voice-oriented networks. Figures 10 and 11 present numerical results for the entire theoretical network traffic load interval, i.e. the terminal traffic of all A- and B-terminals \((Y_{ab})\) is within the range of 0% to 100% of the number \(N_{ab}\) of all active terminals in the system. The input parameters are the same, excluding the capacity of the network (the number of equivalent connection lines), given as a percentage of all terminals in the system. Differences in the network capacity cause different blocking probabilities due to resource insufficiency. Two cases have been considered:

- **Case 1**: The network capacity equals 10% of all terminals presented in the system;
- **Case 2**: The network capacity equals 25% of all terminals presented in the system.

![Fig. 10. Numerical prediction of the Normalized Cost/Quality Ratio (NCQR), Network Call Efficiency (Ebu), and Paid Traffic Intensity in an overall telecommunication system with QoS guarantees (Case 1: Network capacity = 10% of terminals).](image-url)
The results show considerable sensitivity of the Normalized Cost/Quality Ratio (NCQR) from the network capacity and traffic load.

6 Conclusion

In this chapter, a more detailed and precise approach to the traffic characterization has been taken, which allows the definition of new efficiency indicators on different levels, starting with the service phases and stages, continuing with (part of) the network, and finishing with the overall telecommunication system. The proposed Instantaneous Cost/Quality Ratio may be used for the establishment and utilization of dynamic pricing policies, depending on the network load. The use of the instantaneous cost/quality ratios as a source for QoE prediction is a very perspective direction of research. Using similar approaches and specific QoS indicators for other types of networks, e.g. multimedia- and multiservice networks, seems also very topical.

Acknowledgments. This work was coordinated under the EU COST Action IC1304 “Autonomous Control for a Reliable Internet of Services (ACROSS)”. The work was partially funded by the Bulgarian NSF Projects DCOST 01/9 (the work of S. Poryazov) and DCOST 01/20 (the work of E. Saranova).
References

1. Reichl, P.: From charging for quality of service to charging for quality of experience. Ann. Telecommun. 65(3–4), 189–199 (2010)
2. Varela, M., Zwickl, P., Reichl, P., Xie, M., Schulzrinne, H.: From service level agreements (SLA) to experience level agreements (ELA): the challenges of selling QoE to the user. In: Proceedings of IEEE ICC QoE-FI, London, June 2015. ISSN: 2164-7038, https://doi.org/10.1109/iccw.2015.7247432
3. Fiedler, M.: Teletraffic models for quality of experience assessment. Tutorial at 23rd International Teletraffic Congress (ITC 23), San Francisco, CA, September 2011. http://iteletrafic.org/_Resources/Persistent/9269df1e3dca0bf58ee715c3b9af8bce71d4fb26/fiedler11.pdf. Accessed 20 July 2017
4. ITU-T Recommendation E.800 (09/08): Definitions of terms related to quality of service
5. ITU-T Recommendation E.600 (03/93): Terms and definitions of traffic engineering
6. Little, J.D.C.: A Proof of the Queueing Formula L = λW. Oper. Res. 9, 383–387 (1961)
7. ITU-T Recommendation E.501(05/97): Estimation of Traffic Offered in The Network
8. Poryazov, S., Saranova, E.: User-oriented, overall traffic and time efficiency indicators in telecommunications. In: TELFOR 2016 International IEEE Conference #39555, Belgrade, Serbia, 22–23 November 2016, IEEE Catalog Number: CFP1698P-CDR. IEEE (2016). ISBN: 978-1-5090-4086-5/16, INSPEC Accession Number: 16603129, https://doi.org/10.1109/telfor.2016.7818729
9. ITU-T Rec. E.425 (03/2002): Network management – Internal automatic observations

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.