






Conceptual and Analytical Models for Predicting the Quality of Service of Overall Telecommunication Systems

Stoyan Poryazov¹, Emiliya Saranova^{1,2}, and Ivan Ganchev^{1,3,4}

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. This chapter presents scalable conceptual and analytical performance models of overall telecommunication systems, allowing the prediction of multiple Quality of Service (QoS) indicators as functions of the user- and network behavior. Two structures of the conceptual presentation are considered and an analytical method for converting the presentations, along with corresponding additive and multiplicative metrics, is proposed. A corresponding analytical model is elaborated, which allows the prediction of flow-, time-, and traffic characteristics of terminals and users, as well as the overall network performance. In accordance with recommendations of the International Telecommunications Union’s Telecommunication Standardization Sector (ITU-T), analytical expressions are proposed for predicting four QoS indicators. Differentiated QoS indicators for each subservice, as well as analytical expressions for their prediction, are proposed. Overall pie characteristics and their causal aggregations are proposed as causal-oriented QoS indicators. The results demonstrate the ability of the model to facilitate a more precise dynamic QoS management as well as to serve as a source for predicting some Quality of Experience (QoE) indicators.

Keywords: Overall telecommunication system · Performance model
Overall causal QoS indicator · Dynamic QoS management
Telecommunication subservices · Differentiated QoS subservice indicator
QoS prediction · Human factors of QoS

1 Introduction

The telecommunication service is the basis for the Information Service Networks. From the very beginning the Internet began its existence as a packet-based communication system without guarantees for the quality of the services, which are provided on a best-effort basis. At the same time, with the evolution of hardware technologies, and services and applications becoming more and more complex, the quality of service (QoS) has

become a hot topic and the term “Internet QoS” has widely spread. The question of providing QoS guarantees in the Internet is still open (c.f., for instance, the history of the Internet Engineering Task Force (IETF) standards for Integrated Services (IntServ) and Differentiated Services (DiffServ) as well as the Third Generation Partnership Project (3GPP)/European Telecommunications Standards Institute (ETSI) Internet Protocol (IP) Multimedia Subsystem (IMS) initiative).

The QoS has many aspects – QoS offered by the provider, QoS delivered (QoSD), QoS achieved by the provider, QoS experienced by the user/customer (QoSE or QoS_P – QoS perceived) and others. “The understanding of QoSE is of basic importance for the optimization of the income and the resources of the service provider” [1]. A new attitude towards the QoS has become dominant – QoS and Quality of Experience (QoE) are considered as goods. The agreement is made according to the perceived quality – Experience Level Agreement (ELA) [2]. This approach considerably increased interest in the perceived quality among researchers, providers, and users of telecommunication services.

As a result of the intensive research, the definition of QoE evolved and at the moment the QoE is perceived as a degree of satisfaction or irritation of the users of some application or service which is a result of the fulfillment of their expectations about the utility or/and the satisfaction from the application or service in the context of the user’s personality and the current state [3, 4]. The QoS perceived by users depends not only on the quality offered by the provider but also on the context of the services, including the techno-socio-economic environment, user’s context, and other factors. The importance of the teletraffic models, particularly of the overall QoS indicators, for QoE assessment is emphasized by Fiedler [5].

From among the many services provided by a telecommunication system, this chapter deals with flow-, time-, and traffic characteristics of the connection and communication services. The other QoS characteristics of information transmission service are reflected partially and indirectly as a probability of the call attempt abandoning by users.

The main objective of the authors of this chapter is the development of scalable performance models of overall telecommunication systems, as a part of Information Service Networks, including many of the observable system-dependent factors determining the values of QoS indicators.

These models may be used for multiple purposes but the aim of this chapter is to develop prediction models for some key QoS indicators’ values, as functions of the user behavior and technical characteristics of the overall telecommunication system. Such values may be useful for the network design, for the management of telecommunication systems’ QoS, and as a source for predicting some QoE indicators.

The work presented in this chapter continues the development of the approach for the conceptual and analytical modeling of overall telecommunication systems (with QoS guarantees), presented in [6].

Firstly, in Sect. 2, a scalable conceptual model of an overall telecommunication system with QoS guarantees is presented. Two structures of the conceptual presentation are compared – the normalized structure and the pie structure. An analytical method for converting the presentations, along with corresponding additive and multiplicative metrics, is proposed. A qualitative extension of the conceptual model, in comparison with [7], is proposed. This includes two new service branches corresponding to the cases

of ‘called party being busy with another call’ and ‘mailing a message’. This allows analyzing telecommunication systems’ QoS indicators as a composition of QoS indicators of consecutive and parallel subservices.

The developed model is based on: a Bernoulli–Poisson–Pascal (BPP) input flow; repeated calls; limited number of homogeneous terminals; 11 cases of losses of call attempts (due to abandoning, interrupting, blocking, and unavailable service); and three successful cases (normal interactive communication, communication after call holding, and mailing). The calling (A) and called (B) terminals (and users) are considered separately, but in interaction to each other. This allows formulation of QoS indicators separately for A-, B-, and AB-terminals.

In Sect. 3, on the basis of the developed conceptual model, a corresponding analytical model is elaborated. User behavior parameters and technical characteristics of the telecommunication network serve as an input for the model. The model itself is intended for systems remaining in a stationary state. It is insensitive to the distributions of random variables and provides results in the form of mean values of the output parameters. The model is verified for the entire theoretical interval of network load. It allows the prediction of flow-, time-, and traffic characteristics of A-, B-, and AB-terminals (and users), as well as of the overall network performance.

In accordance with recommendations of the International Telecommunications Union’s Telecommunication Standardization Sector (ITU-T), analytical expressions for the prediction of three QoS indicators are proposed:

- **Carried Switching Efficiency**, for finding B-terminal (Subservice 1);
- **B-Terminal Connection Efficiency**, for connection to the B-terminal, which aggregates the Carried Switching Efficiency;
- **Overall Call Attempt Efficiency**, for call attempts finishing with fully successful communication, which aggregates the Carried Switching Efficiency, B-Terminal Connection Efficiency, Finding B-User Subservice, and Communication Subservice.

Four differentiated QoS indicators for each subservice are proposed along with analytical expressions for their prediction:

- **Carried Switching Efficiency** (Ecs), for finding B-terminal (Subservice 1) as per the ITU-T recommendations;
- **QoS specific indicator** (Qb) of Connection to the B-terminal (Subservice 2);
- **QoS specific indicator** (Qu) of Finding B-user (Subservice 3);
- **QoS specific indicator** (Qc) of Communication (Subservice 4).

The four QoS specific indicators are independent. They are components (in multiplicative metrics) of the ITU-T concordant Overall Call Attempt Efficiency indicator (Ec):

$$Ec = Ecs Qb Qu Qc.$$

The usage of the proposed QoS indicators of telecommunication subservices allows conducting a more specific QoS analysis and more adequate QoS management.

In Sect. 4, in accordance with the ITU-T recommendations, analytical expressions for the prediction of the Overall Traffic Efficiency Indicator and other overall pie

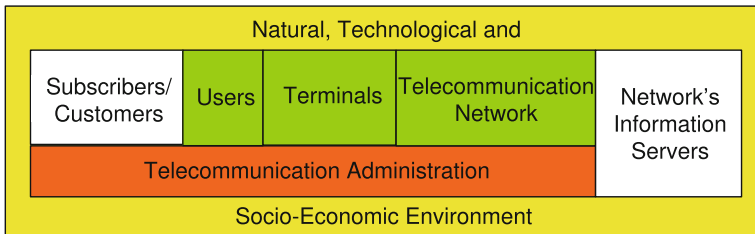
parameters and their causal aggregations are proposed and illustrated numerically. The overall pie characteristics and their causal aggregations could be considered as causal-oriented QoS indicators. The results allow a more precise estimation of the dynamic importance of each reason of call attempts finishing and thus a more precise dynamic effort targeting of the QoS management.

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

2 Conceptual Model

2.1 Background

At the telecommunication system level, Ericson has proposed a reference model consisting of five parts – terminals, access-, transport-, network management-, and network intelligence part [8]. We have extended this reference model by making difference between the telecommunication system and the telecommunication network, and by applying the present ITU-T terminology (Fig. 1). It contains seven parts (subsystems): (1) Network Environment (natural-, technological-, and socio-economic environment); (2) Users; (3) Subscribers/Customers¹; (4) Terminals; (5) Telecommunication Network; (6) Network’s Information Servers (network intelligence); and (7) Telecommunication Administration (network service provider). The interaction between subsystems (if any) is presented by a common border between their representing rectangles in Fig. 1. Each subsystem is part of the environment (context) of the other subsystems.



Legend:

- Detailed interactive behavior description
- Indirect (equipment characteristics & interruption probabilities)
- Target values & restrictions
- Not considered in the low-level models

Fig. 1. A reference model of an overall telecommunication system and its environment (an extension of [9]).

¹ According to [1], the user is “A person or entity external to the network, which utilizes connections through the network for communication”, whereas the customer is “A user who is responsible for payment for the services”.

For designing and managing telecommunication systems one needs scalable models in all aspects of the term ‘scalability’: “*scale down*: make smaller in proportion; reduce in size”; “*scale up*: make larger in proportion; increase in size”; “*to scale*: with a uniform reduction or enlargement” [10]. Models’ scalability includes: temporal-, spatial-, structural-, parametric-, conceptual-, functional-, and etc. scalabilities.

Basic Virtual Devices: At the bottom of the structural model presentation, we consider ‘basic virtual devices’ that do not contain any other virtual devices. A basic virtual device has the graphic representation as shown in Fig. 2.

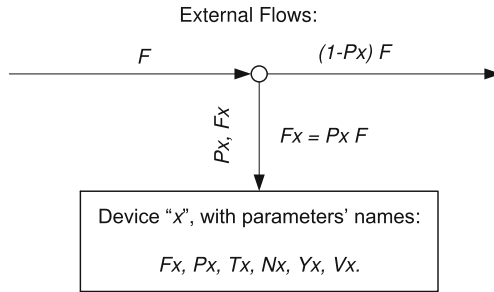


Fig. 2. A graphical representation of a basic virtual device x .

Parameters of the basic virtual device x are the following (c.f. [11] for terms definition):

- F_x – Intensity or incoming rate (frequency) of the flow of requests (i.e. the number of requests per time unit) to device x ;
- P_x – Probability of directing the requests towards device x ;
- T_x – Service time (duration of servicing of a request) in device x ;
- Y_x – Traffic intensity [Erlang];
- V_x – Traffic volume [Erlang - time unit];
- N_x – Number of lines (service resources, positions, capacity) of device x .

Functional Normalization: In our models, we consider monofunctional idealized basic virtual devices of the following types (Fig. 3):

- *Generator* – this device generates calls (service requests, transactions);
- *Terminator* – this block eliminates every request entered (so it leaves the model without any traces);
- *Modifier* – this device changes the intensity of the incoming flow, creating or nullifying requests. It is used to model the input flow, in conformance with the system status (c.f. Fig. 7);
- *Copier* – this block creates copies of the requests received and directs them to a route different from the original one;
- *Director* – this device unconditionally points to the next device, which the request shall enter, but without transferring or delaying it;

- *Enter Switch* – this block checks if there is a free resource/place in the next block for a request to be accommodated in: if yes, the request is passed to it without delay; if not – the request is re-directed to another device;
- *Server* – this device models the delay (service time, holding time) of requests in the corresponding device without their generation or elimination. It models also traffic and time characteristics of the requests processing (c.f. Fig. 2);
- *Transition* – this device selects one of its possible exits for each request entered, thus determining the next device where this request shall go to;
- *Graphic Connector* – this is used to simplify the graphical representation of the conceptual model structure. It has no modeling functions.

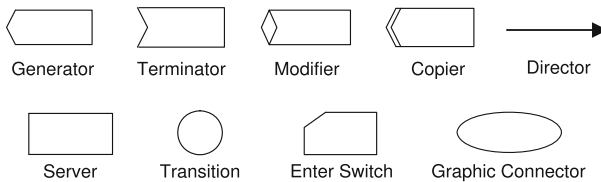


Fig. 3. A graphical block representation of the main basic virtual mono-functional devices used.

Structural Normalization: Following the theorem of Böhm and Jacopini [12], we use basic virtual devices mainly with one entrance and one exit. Exceptions are: the transition device, which in our structural normalization has one entrance and two exits (for splitting the requests' flows) or two entrances and one exit (for merging the requests' flows); and the copier with its one entrance and two exits.

Causal Structure Presentation: Any service may end due to many reasons. In a telecommunication network, all reasons are classified into four types: network failures, user failures (ineffective calls associated with the callers and callees), network service provider failures, and successful ending (completed seizures) [13, 14]. The 'cause value' field in [14] contains 99 items. In [13], there are 127 'cause value' numbers. Cisco lists 131 'call termination cause codes' and 44 'Cisco-specific call termination cause codes' [15].

Complex Virtual Devices: Each reason for service ending has its own probability to occur and mean service time (duration). In our conceptual model, the service execution goes through different stages (e.g. dialing, switching, ringing, etc.), each consisting of different phases. Each stage of a modeled service corresponds to one (or more) complex virtual device and contains 'service branches' (service phases). Typically, a service phase includes a service device and all necessary auxiliary devices such as queues, entry and exit devices, as well as virtual devices reflecting the user behavior, associated with this phase, e.g. the waiting time before initiating a repeated call attempt. Each service branch corresponds to a different reason of service ending. The service branches form the 'causal structure' of the modeled service. The causal structure of a complex virtual

device x (with input requests' flow frequency F_x , mean service time T_x , and traffic intensity Y_x) could be presented in two ways – by using a normalized structure or a pie structure (Fig. 4).

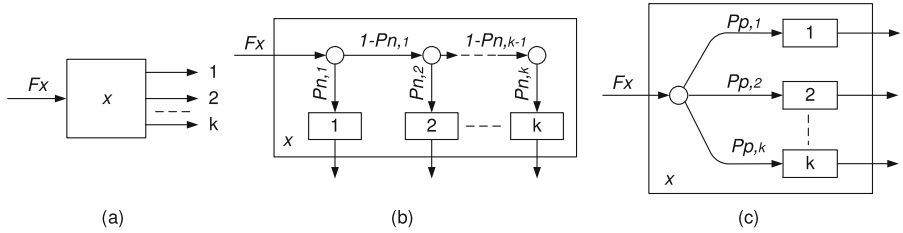


Fig. 4. (a) A complex virtual device x , representing a service with k reasons for ending; (b) the normalized causal structure of device x ; (c) the pie causal structure of device x .

Both structures include k virtual ‘causal devices’, each with its own mean input requests’ flow frequency F_i , mean service time T_i , and traffic intensity Y_i . Obviously:

$$Y_x = \sum_{i=1}^k Y_i; \quad F_x = \sum_{i=1}^k F_i. \tag{1}$$

The difference between the two presentations is in the internal flow structures only. In the pie causal structure (Fig. 4c), all causal service branches have common beginning. The probability $P_{p,i}$ shows what part (pie) of the service incoming flow is directed to the causal device i . All probabilities $P_{p,i}$ are dependent:

$$\sum_{i=1}^k P_{p,i} = 1. \tag{2}$$

In the normalized structure (Fig. 4b), all service branches are ordered consecutively as derivations of one ‘successful completed service branch’. The probability $P_{n,i}$ shows what part of the flow, already passed through the previous causal branches, is derived to the considered service case (causal device) i . The probabilities $P_{n,i}$ are independent (orthogonal, normal). The order of causal branches does not matter (has no mathematical meaning) but usually the branch of successful completion of the service ($P_{n,k}$) is the last one.

Both structures lead to different presentations of the same QoS indicators. For example, the probability (resp. efficiency E_c) for successful completion of the service in the normalized (3) and pie presentation (4) is respectively:

$$E_c = \prod_{i=1}^{k-1} (1 - P_{n,i}). \tag{3}$$

$$E_c = 1 - \sum_{i=1}^{k-1} P_{p,i}. \quad (4)$$

The normalized- and pie structures are used by many authors but usually without these associated names, and without discussions about the nature of parameters and how one structure could be converted to the other. For example, in [16] expressions like (2), (3) and (4) are classified as ‘aggregation functions’, whereas (2) is additive, (3) is multiplicative, and (4) is not specified.

The conversion between the values of the normalized and pie probabilities (and vice versa) could be done by means of the following system of k equations with k variables ($P_{n,i}$ or $P_{p,j}$, $j = 1, 2, 3, \dots, k$):

$$\begin{cases} P_{p,j} = P_{n,j}, & \text{if } j = 1 \\ P_{p,j} = P_{n,j} \prod_{i=1}^{j-1} (1 - P_{n,i}), & \text{if } j = 2, \dots, k \end{cases} \quad (5)$$

Each structure has advantages over the other. The normalized structure allows clearer conceptual presentation and simpler inference of the analytical models, but normalized probabilities depend on the causal branch positions. The pie structure is more natural and impressive in business presentations (pie charts, pie graphs). Each structure is a mathematical equivalent of the other. Both allow for model scalability.

2.2 Conceptual Model

We consider a virtual overall telecommunication system including users, terminals and possibly several telecommunication networks, operated by different operators. We consider VNET carrying Class 0 traffic (real-time, jitter-sensitive, with high interaction (Voice over IP (VoIP), video teleconference) [17]. The VNET utilizes virtual channel switching principles, following the main method for traffic QoS guarantees – resource reservation [18]: “Bandwidth reservation is recommended and is critical to the stable and efficient performance of Traffic Engineering methods in a network, and to ensure the proper operation of multiservice bandwidth allocation, protection, and priority treatment.”

In our approach, the overall network QoS parameters are aggregation of all end-to-end QoS parameters of all terminals and connections in the network, within the considered time interval (Fig. 5).

The VNET in Fig. 5 includes also users, not just the terminals, and generalizes call intensity, time- and traffic parameters of the calling (A), called (B) and all active (AB) terminals, as well as of the overall network equivalent switching lines, reflecting resources of all comprised telecommunication networks.

In this chapter, we propose a considerable extension of the conceptual and analytical performance models of the overall telecommunication system with QoS guarantees, described in [6]. This includes two new service branches corresponding to the cases of ‘called party being busy with another call’ and ‘mailing a message’.

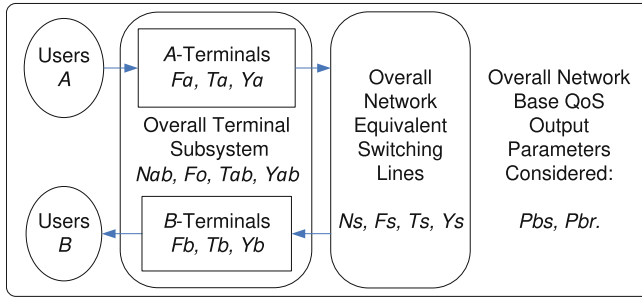


Fig. 5. A generalized VNET, including users and terminals, with overall QoS guaranties (a modification of [19]).

Basic Virtual Devices’ Name Notation. In the normalized conceptual model, each virtual device has a unique name, depending on its position and the role it plays in the model (Figs. 6, 7, 8, 9 and 10).

Virtual Device Name = <BRANCH EXIT><BRANCH><STAGE>		
BRANCH EXIT:	BRANCH:	STAGES:
r = repeated;	e = enter	d ialling;
t = terminated (not considered usually).	a = abandoned;	s witching;
	b = blocked;	r inging;
	i = interrupted;	h olding;
	n = not available;	c ommunication;
	c = carried.	m ailing (voice).

Fig. 6. The basic virtual devices’ name notation.

The model is partitioned into service stages (**d**ialling, **s**witching, **r**inging, **h**olding, communication, and **m**ailing).

Each service stage has different branches (**e**ntered, **a**bandoned, **b**locked, **i**nterrupted, **n**ot available, **c**arried), corresponding to the modeled possible cases of ending the service.

Each branch has two exits (**r**epeated, **t**erminated) that show what happens with the service request after it enters the telecommunication system. Users may make a new bid (repeated service request) or may stop attempting (terminated service request).

In the virtual devices’ name notation, the corresponding first letters of the name of the branch exit, the branch, and the service stage are used (in this order) to form the name of the virtual device:

$$\text{Virtual Device Name} = \langle \text{BRANCH EXIT} \rangle \langle \text{BRANCH} \rangle \langle \text{STAGE} \rangle$$

Complex Virtual Devices’ Names. We use the following complex virtual devices (i.e. devices, consisting of several basic virtual devices):

a – a virtual device that comprises all A-terminals (i.e. the calling terminals) in the system. The *a* device is represented as a ‘dotted line’ box, named *a0* in Fig. 7, *a 1* in Fig. 8, *a 2* in Fig. 9, and *a 3* in Fig. 10;

b – a virtual device that comprises all B-terminals (i.e. the called terminals) in the system. The *b* device is represented as a ‘dashed line’ box, corresponding to the B-terminal load, in Figs. 8, 9, and 10;

ab – this device comprises all the active (i.e. calling and called) terminals in the system;

s – a virtual device corresponding to the equivalent connection lines in the switching system. It is represented as a ‘dotted and dashed line’ box, named *s*, inside the *a0* box in Fig. 7, and other *a* boxes (*a 1* in Fig. 8, *a 2* in Fig. 9, and *a 3* in Fig. 10).

The network environment includes also basic virtual devices outside the *a* and *b* complex devices. Service requests in the environment do not occupy network devices, but rather form incoming flows out of demand and repeated call attempts.

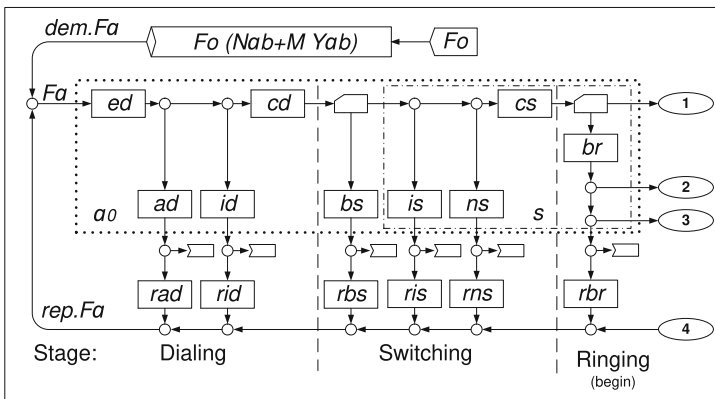


Fig. 7. Service stages ‘Dialing’, ‘Switching’, and the beginning of stage ‘Ringing’.

In Fig. 7, *Fo* is the intent intensity of calls², with a Poisson distribution, generated by a terminal; *dem.Fa*³ is the intensity of demand (first, primary calls), generated by all A-terminals, according the BPP-traffic model (c.f. the modifier block in Fig. 7); *M* is a constant. In our approach, every value of *M* within the interval $[-1, +1]$ is allowed. If $M = -1$, the intensity of the demand flow corresponds to the Bernoulli (Engset) distribution; if $M = 0$ – to the Poisson (Erlang) distribution; and if $M = +1$ – to the Pascal (Negative Binomial) distribution.

² In this chapter, the term ‘call’ means ‘service request’, ‘call attempt’ or ‘bid’ according to the terminology in [11].

³ 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*.

rep.Fa stands for repeated attempts, generated by A-users and A-terminals, in the case of unsuccessful call attempts; *Fa* is the flow generated by and occupying the A-terminals (it is a sum of the intensities of primary (demand) call attempts (*dem.Fa*) and repeated attempts *rep.Fa*).

Devices ‘entered dialing’ (*ed*), ‘carried dialing’ (*cd*), and ‘carried switching’ (*cs*), belong to the successful service branch.

Devices ‘abandoned dialing’ (*ad*), ‘interrupted dialing’ (*id*), ‘blocked switching’ (*bs*), ‘interrupted switching’ (*is*), ‘not available switching (service, number)’ (*ns*), and ‘blocked ringing’ (*br*) belong to the unsuccessful (due to different reasons) service branches. They reflect durations of the correspondent signaling, e.g. the ‘busy tone’ duration.

Devices ‘repeated abandoned dialing’ (*rad*), ‘repeated interrupted dialing’ (*rid*), ‘repeated blocked switching’ (*rbs*), ‘repeated interrupted switching’ (*ris*), ‘repeated not available switching (service, number)’ (*rns*), and ‘repeated blocked ringing’ (*rbr*) correspond to the duration of users’ requests waiting, outside the network equipment, before the next repeated call attempt.

The device of type ‘Enter Switch’ (just before the ‘blocked switching’ (*bs*) device) in Fig. 7) deflects calls if there is no free line in the switching system, with probability of blocked switching (*Pbs*). The second ‘Enter Switch’ device (after the block ‘carried switching’ (*cs*) in Fig. 7) deflects calls, with probability of blocked ringing (*Pbr*), if the called B-terminal is busy.

Note that there is no B-terminal traffic in the part of the conceptual model, presented in Fig. 7.

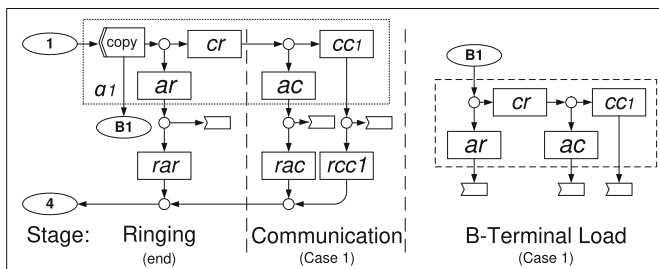


Fig. 8. Service stages ‘Ringing’ (end), ‘Communication’, and ‘B-terminal Load’ (Case 1).

Figure 8 presents the call flows, in Case 1, when the B-terminal is found free (c.f. Connector 1 in Figs. 7 and 8). In this case, the flow intensity, occupying the B-terminal, is generated by the Copy device (c.f. Connector B1), because at the beginning of the ringing stage, the B-terminal becomes busy. In Case 1, the traffic load on the A-terminal equals the traffic load on the B-terminal.

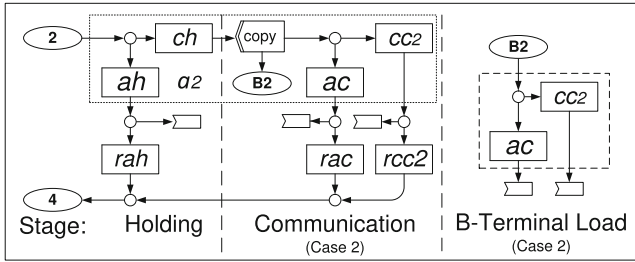


Fig. 9. Service stages ‘Holding’, ‘Communication’, and ‘B-terminal Load’ (Case 2).

Figure 9 presents the call flows, in Case 2, when the B-terminal is found busy (c.f. Connector 2 in Figs. 7 and 9). This is the case of call holding – the A-user is put to wait (virtual devices ‘carried holding’ (*ch*) and ‘abandoned holding’ (*ah*)). In pure voice communication systems, in this case, a pre-recorded music/message is usually played to the caller while waiting. The connection is not terminated but no verbal communication is possible. At the same time the B-user is notified (by a sound and/or light indication on his/her terminal/phone) that another call is trying to reach him/her, with the options of answering (virtual devices ‘carried holding’ (*ch*)) or not answering it (virtual device ‘abandoned holding’ (*ah*)). During the hold time, the B-user is able to continue with or answer another call, retrieve a waiting call, etc. Note that in this case, traffic loads on the A- and B-terminals are considerably different.

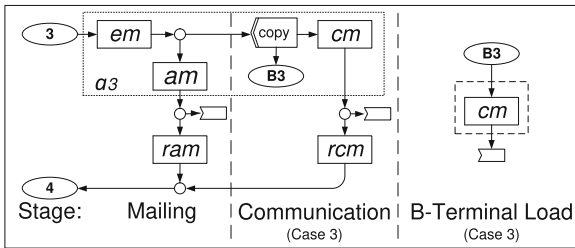


Fig. 10. Service stages ‘Mailing’, ‘Communication’, and ‘B-terminal Load’ (Case 3).

Figure 10 presents the call flows, in Case 3, when the B-terminal is found busy (c.f. Connector 3 in Figs. 7 and 10). This is the case when the A-user is redirected to a mail service to leave an audio message. In some systems, there is also a possibility to leave a video message, e.g. a visual voicemail. The A-user receives an invitation to leave a mail message (virtual device ‘enter mailing’ (*em*)) and may decide to use this service (virtual device ‘carried mailing’ (*cm*)) or to abandon the service (virtual device ‘abandoned mailing’ (*am*)). The message is retrieved (later) by the B-user either as audio directly from his/her terminal/phone or from another device via a web link supplied by an email message, or as a text by utilizing a voicemail-to-text functionality. This message retrieval is reflected by the case of using the B-terminal by the B-user in our conceptual model.

Parameters' Qualification. In Fig. 7, one may see notations '*Fa*', '*dem.Fa*', and '*rep.Fa*', using qualifiers *dem* and *rep*. Traffic qualification is necessary and it is used in [11], but without any attempt for including the qualifiers in the parameters' names. The problem is more complex: (1) one would like to have the same, or very similar, parameters' names in the conceptual-, analytic-, and computer models; (2) one would like to meet the Name Design Criteria: "Names with which human beings deal directly should be user-friendly. A user-friendly name is one that takes the human user's point of view, not the computer's. It is one that is easy for people to deduce, remember and understand, rather than one that is easy for computers to interpret." [20], Annex J: "Name Design Criteria".

Since 2006 [6] we use up to two qualifiers as a part of the parameter's name. The first is for the parameter value's origin, e.g. *emp* for 'empirical', *dsn* for 'designed', *trg* for 'target', etc. The second qualifier characterizes the traffic. Most of the traffic qualifiers are described in [11]. In this paper we use *dem* for 'demand', *rep* for 'repeated', *ofr* for 'offered', and *crr* for 'carried'. We expand the meaning of the traffic qualifiers to the other parameters determining the traffic, e.g. in our notations, *ofr.Ys = ofr.Fssrv.Ts* means: 'the offered traffic intensity to the switching system is a product of the offered requests' frequency (rate) and the service time in the switching system.

The definition of the offered traffic needs more explanations. There are two offered traffic definitions in the ITU-T recommendations: (1) Equivalent Traffic Offered [21]; and (2) Traffic Offered [11]. In the other standardization documents, there is only one offered traffic definition, close to the Equivalent Traffic Offered [21]. In the overall network performance models, both definitions give considerably different values [22]. In this chapter, we use only the definition of the Equivalent Traffic Offered [21].

2.3 QoS Prediction Task Formulation

We consider the conceptual model presented in Figs. 6, 7, 8, 9, 10 and described in Sect. 2.2. In this chapter, we consider that the overall telecommunication system provides four services: (1) finding B-terminal; (2) connection to B-terminal; (3) finding B-user (with sound, vibration, message, etc.); and (4) transmission and/or record of messages. The quality of this services depends on many subsystems (c.f. Fig. 1), including the user- and network behavior.

Types of Parameters. There are two types of parameters – static and dynamic. The 10 basic dynamic parameters (with values dependent of the system state) are: *Fo*, *Yab*, *Fa*, *dem.Fa*, *rep.Fa*, *Pbs*, *Pbr*, *ofr.Fs*, *Ts*, and *ofr.Ys*. All others dynamic parameters can be obtained from these.

Note that the traffic *Yab* from all terminals is accepted as a system macro-state parameter.

Input Parameters. These are mostly static, i.e. related to the network technical characteristics or the user behavior. We choose one dynamic parameter - *Fo* (the intent intensity of calls of one idle terminal) as an independent input variable. The proposed analytical model allows to find all dynamic values, if *Fo* and all static parameters are known.

The probability of finding the B-user is considered static (i.e. independent of the system state).

The basic *QoS output parameters* are:

- Quality of finding the B-terminal service, represented by the probability of call blocking due to unavailable network equipment (equivalent network switching lines) – blocked switching (*Pbs*);
- Quality of connection to the B-terminal, represented by the probability of call blocking due to busy B-terminal – blocked ringing (*Pbr*).

These two parameters allow determination of many other QoS indicators, related to traffic-, time-, and flow characteristics of users and terminals.

The goal of this section is to find analytically all unknown basic dynamic parameters, including the basic QoS output parameters.

2.4 Main Assumptions

For a clear analytical modeling of a telecommunication system with QoS guarantees, the following assumptions were made:

Assumption 1. A closed service system, presented in Figs. 6, 7, 8, 9 and 10, is considered;

Assumption 2 (Capacity of Devices). The switching system (*s*) has capacity of *Ns* connections (every virtual internal switching line may carry only one call attempt). Complex devices have limited capacity: the capacity of the *ab* device is $Nab \in [2, \infty)$ terminals; the capacity of every terminal is engaging in one call (incoming or outgoing); all basic virtual devices have unlimited capacity;

Assumption 3 (Occupation of A-terminals). Every incoming call attempt (*Fa*), from the environment, falls only on a free A-terminal. This terminal becomes a busy one;

Assumption 4 (Steady State). Every device is in a stationary state. Hence the Little's theorem [23] is applicable to each device: $Y = F T$;

Assumption 5 (Capacity of Call Attempts). Every call attempt may occupy no more than one place, if any, in each basic virtual device;

Assumption 6 (Network Environment). The calls and devices in the environment (outside blocks *a* and *b* in Figs. 7, 8, 9 and 10) form the intent- and repeated calls flows). They don't create telecommunication network's load;

Assumption 7 (Device Independence). Excluding the dependences described in the mathematical model, all parameters of a virtual device are independent from the parameters' values of any other virtual device in the model;

Assumption 8 (Randomness of the Processes). All variables in the analytical model are considered random with a fixed distribution; the Little's theorem is used for working with their mean values.

Assumption 9 (A and B Simultaneity). If a call attempt is served in corresponding virtual devices belonging to A- and B-terminal's load (e.g. *ar*, *cr*, *ac*, *c1*, *cc2*, *cm* in Figs. 7, 8, 9 and 10), it seizes and releases them simultaneously, with the same service load and duration.

Assumption 10 (Virtual Channel Switching). Every call attempt occupies simultaneously places in all the basic virtual devices of the complex device *a* or *b* it is passing through, including the basic device where it is at the moment of observation. Every call attempt releases all occupied places at the very moment it leaves the complex device *a* or *b*.

Assumption 11 (Homogeneity⁴). All terminals and users are homogeneous.

Assumption 12 (Self-Excluding). Every A-terminal directs, with uniform distribution, all its call attempts to other terminals, not to itself;

Assumption 13 (B-flow). The flow of call attempts, occupying B-terminals (*Fb*), is ordinary. (The case when two or more call attempts reach simultaneously a free B-terminal is not considered, due to its statistical unimportance);

Assumption 14 (B-terminal Busy Probability). The stationary probability of a call to find the intended B-terminal busy ('blocked ringing' (*Pbr*)) during the first (primary, demand) attempt and all subsequent (repeated) attempts is one and the same.

3 Analytical Model

3.1 Overall Input Flow Intensity

The input (incoming) flow to the telecommunication network, with intensity *Fa*, is the flow generated by (and occupying) A-terminals. From the ITU E.600 definitions and Fig. 7 it is obvious that the intensity of incoming flow is a sum of the intensities of primary (demand) call attempts (*dem.Fa*) and repeated attempts (*rep.Fa*):

$$Fa = dem.Fa + rep.Fa. \quad (6)$$

From the definition of the BBP-flow and Fig. 7 we have:

$$dem.Fa = Fo(Nab + M Yab) \quad (7)$$

⁴ *Homogeneity* means that all relevant characteristics and their considered mean values are the same.

3.2 QoS Indicator 1: Carried Switching Efficiency

According to Definition 2.11 in [11]: “fully routed call attempt; successful call attempt” is “A successful call attempt that receives an answer signal”. We define the Carried Switching Efficiency of the ‘Finding B-Terminal’ service as a ratio of the flow intensity of the calls reaching the intended B-terminal (Fcs) and receiving an answer signal ‘busy tone’ or ‘ringing tone’, to the incoming call attempts intensity (Fa).

The Carried Switching Efficiency corresponds to the concept of “answer bid ratio (ABR)” in [11]: “On a route or a destination code basis and during a specified time interval, the ratio of the number of bids that result in an answer signal, to the total number of bids.”

In the conceptual model considered (c.f. Fig. 7), the calls served in the device ‘carried switching’ (Fcs) are those, reaching the B-terminals. The intensity Fcs may be calculated by taking into account Fa and losses on the way to the cs device (c.f. Fig. 7). This, expressed in two ways – by using the lost call flows and probabilities of successful moving of requests along the successful branch, results in the following: a

$$Fcs = Fa(1 - Pad)(1 - Pid)(1 - Pbs)(1 - Pis)(1 - Pns). \quad (8)$$

So, the Carried Switching Efficiency (Ecs) of the ‘Finding B-Terminal’ service is:

$$Ecs = \frac{Fcs}{Fa} = (1 - Pad)(1 - Pid)(1 - Pbs)(1 - Pis)(1 - Pns). \quad (9)$$

3.3 Repeated Calls Flow

Based on the repeated calls definition [21] and the proposed conceptual model (Figs. 7, 8, 9 and 10), the intensity of the repeated attempts ($rep.Fa$) is:

$$rep.Fa = Frad + Frid + Frbs + Fris + Frns + Frbr + Fr1 + Fr2 + Fr3, \quad (10)$$

where $Fr1 = Frar + Frac + Frcc1$ is the intensity of repeated attempts in Case 1, directed to Connector 4 (c.f. Fig. 8); $Fr2 = Frah + Frac + Frcc2$ is the intensity of repeated attempts in Case 2, directed to Connector 4 (c.f. Fig. 9); $Fr3 = Fram + Frcm$ is the intensity of repeated attempts in Case 3, directed to Connector 4 (c.f. Fig. 10).

Proposition 1. The intensity of the repeated attempts $rep.Fa$ may be obtained as:

$$\begin{aligned} rep.Fa = & Fa(Pad Prad + (1 - Pad)(Pid Prid + (1 - Pid)(Pbs Prbs \\ & + (1 - Pbs)(Pis Pris + (1 - Pis)(Pns Prns + (1 - Pns)Pbr (Ph Pr2 \\ & + (1 - Ph)(Pm Pr3 + (1 - Pm) Prbr)) + (1 - Pbr) Pr1))))), \end{aligned} \quad (11)$$

where Ph (‘holding’) is the probability of calls going to Case 2 (c.f. Connector 2 in Fig. 7), Pm (‘mailing’) is the probability of calls going to Case 3 (c.f. Connector 3 in Fig. 7), and:

$$Pr1 = Ecs (Pah Prah + (1 - Pah) (Pac Prac + (1 - Pac) Prcc1)); \tag{12}$$

$$Pr2 = Ecs Pbr Ph (1 - Par) (Pah Prah + (1 - Pah)(Pac Prac + (1 - Pac)Prcc2)); \tag{13}$$

$$Pr3 = Ecs Pbr (1 - Ph) Pm (Pam Pram + (1 - Pam)Prcm). \tag{14}$$

Proof: As can be seen from Figs. 7, 8, 9 and 10, Assumption 1, and (10), *rep.Fa* is a sum of intensities of repeated attempt flows, in all branches. The intensities of repeated attempt flows, in all branches, may be easily expressed as functions of *Fa*, following the conceptual model structure depicted in Figs. 7, 8, 9 and 10:

$$Frad = Fa Pad Prad; \tag{15}$$

$$Frid = Fa (1 - Pad) Pid Prid; \tag{16}$$

$$Frbs = Fa (1 - Pad)(1 - Pid) Pbs Prbs; \tag{17}$$

$$Fris = Fa (1 - Pad)(1 - Pid)(1 - Pbs)Pis Pris; \tag{18}$$

$$Frns = Fa (1 - Pad)(1 - Pid)(1 - Pbs)(1 - Pis)Pns Prns; \tag{19}$$

$$Frbr = Fa (1 - Pad)(1 - Pid)(1 - Pbs)(1 - Pis)(1 - Pns)Pbr Prbr; \tag{20}$$

$$Fr1 = Fa (1 - Pad)(1 - Pid)(1 - Pbs)(1 - Pis)(1 - Pns)(1 - Pbr) (Pah Prah + (1 - Pah)(Pac Prac + (1 - Pac)Prcc1)) = Fa Pr1; \tag{21}$$

$$Fr2 = Fa (1 - Pad)(1 - Pid)(1 - Pbs)(1 - Pis)(1 - Pns)Pbr Ph(1 - Par) (Pah Prah + (1 - Pah)(Pac Prac + (1 - Pac)Prcc2)) = Fa Pr2; \tag{22}$$

$$Fr3 = Fa (1 - Pad)(1 - Pid)(1 - Pbs)(1 - Pis)(1 - Pns)Pbr (1 - Ph)Pm (Pam Pram + (1 - Pam)Prcm) = Fa Pr3. \tag{23}$$

By adding Eqs. (15) to (23) and taking into account (10), we obtain (11).

Proposition 2. By distinguishing static and dynamic parameters in (11), and after some algebraic operations, we obtain *rep.Fa* as a simple function of *Fa*, *Pbr*, and *Pbs*:

$$rep.Fa = Fa (R1 + R2 Pbr (1 - Pbs) + R3 Pbs), \tag{24}$$

where:

$$R1 = Pad Prad + (1 - Pad)(Pid Prid + (1 - Pid) Pis Pris + (1 - Pis)(Pns Prns + (1 - Pns) Pr1)); \tag{25}$$

$$R2 = (1 - Pad)(1 - Pid)(1 - Pbs)(1 - Pns)(Ph Pr2 + (1 - Ph)(Pm Pr3 + (1 - Pm)Prbr) - Pr1); \tag{26}$$

$$R3 = (1 - Pad)(1 - Pid)(Prbs - (Pis Pris + (1 - Pis)(Pns Prns + (1 - Pns)Pr1))). \tag{27}$$

3.4 QoS Indicator 2: B-Terminal Connection Efficiency

Definition 2.10. in [11] describes “completed call attempt; effective call attempt” as “A call attempt that receives intelligible information about the state of the called user”.

Based on this, we define the B-Terminal Connection Efficiency as a ratio of the flow intensity of the calls occupying the intended B-terminal (Fb) to the incoming call attempts’ intensity (Fa).

In the considered conceptual model, the calls occupying the B-terminal receive information about the state of the called B-user such as signals ‘ringing tone’ (Case 1 in Fig. 8), ‘holding signal’ (Case 2 in Fig. 9), or ‘invitation to mailing’ signal (Case 3 in Fig. 10). The A-user may accept (devices *cr*, *ch*, *cm*) or reject (devices *ar*, *ah*, *am*) the offers.

3.5 B-Terminals’ Characteristics

The intensity of the input flow occupying all B-terminals (Fb) is a sum of the following intensities of input flows (to B-terminals): $Fb1$, in Case 1 - Ringing stage (generated in the copy device in Fig. 8); $Fb2$, in Case 2 - Communication stage (generated in the copy device in Fig. 9); and $Fb3$, in Case 3 - Communication stage (generated in the copy device in Fig. 10), or:

$$Fb = Fb1 + Fb2 + Fb3. \quad (28)$$

The flow intensities $Fb1$, $Fb2$ and $Fb3$ can be calculated by considering the intensity of the carried switching flow Fcs . From Figs. 7, 8, 9 and 10, we obtain directly:

$$Fb1 = Fcs (1 - Pbr) \quad (29)$$

$$Fb2 = Fcs Pbr Ph (1 - Pah) \quad (30)$$

$$Fb3 = Fcs Pbr (1 - Ph) Pm (1 - Pam) \quad (31)$$

After summation, we obtain Fb as:

$$Fb = Ecs Fa ((1 - Pbr) + Pbr (Ph (1 - Pah) + (1 - Ph) Pm (1 - Pam))) = Eb Fa \quad (32)$$

where Fb is the B-Terminal Connection Efficiency, or shortly ‘B-Efficiency’. B-Efficiency (Eb) is expressed as a ratio of flow intensity, occupying B-terminals (Fb), to the intensity of the incoming flow (Fa). It is considerably different from the Carried Switching Efficiency (Ecs):

$$Eb = \frac{Fb}{Fa} = Ecs((1 - Pbr) + Pbr(Ph(1 - Pah) + (1 - Ph) Pm (1 - Pam))). \quad (33)$$

Flow of Call Attempts, Occupying all B-Terminals

Traffic intensity to B-terminals (Yb) is a sum of traffic intensities (to them) in cases 1, 2, and 3. From Figs. 8, 9 and 10 and the Little’s theorem, we can obtain directly the following:

$$Yb = Yb1 + Yb2 + Yb3, \tag{34}$$

where

$$Yb1 = Yar + Ycr + Yac + Ycc1 = Fb1 Tb1. \tag{35}$$

$$Yb2 = Yac + Ycc2 = Fb2 Tb2. \tag{36}$$

$$Yb3 = Ycm = Fb3 Tb3. \tag{37}$$

and

$$Tb1 = Par Tar + (1 - Par)(Tcr + Pac Tac + (1 - Pac) Tcc1) \tag{38}$$

$$Tb2 = Pac Tac + (1 - Pac) Tcc2 \tag{39}$$

$$Tb3 = Tcm \tag{40}$$

Proposition 3. Traffic intensity to B-terminals (Yb) may be calculated from the equation:

$$Yb = Ecs Fa ((1 - Pbr) Tb1 + Pbr (Ph (1 - Pah) Tb2 + (1 - Ph) Pm (1 - Pam) Tb3)), \tag{41}$$

where Tb is the mean holding time of calls in B-terminals and Fb is the intensity of call attempts that occupy B-terminals.

Proof: After summation of $Yb1$, $Yb2$ and $Yb3$, and taking into account expressions (2.8)–(2.10), we obtain:

$$Yb = Fcs ((1 - Pbr) Tb1 + Pbr (Ph (1 - Pah) Tb2 + (1 - Ph) Pm Tb3))$$

and after replacing Fcs with $Ecs Fa$ from (9) we get (41).

Proposition 4. The mean holding time of all B-terminals (Tb), in accordance with cases 1, 2, 3, is:

$$Tb = \frac{Yb}{Fb} = \frac{(1 - Pbr) Tb1 + Pbr (Ph (1 - Pah) Tb2 + (1 - Ph) Pm (1 - Pam) Tb3)}{(1 - Pbr) + Pbr (Ph (1 - Pah) + (1 - Ph) Pm (1 - Pam))}. \tag{42}$$

Proof: This follows directly from the formulas for Yb and Fb , and by directly applying the Little’s theorem.

Consequence: Traffic intensity of B-terminals (Yb) is:

$$Yb = Fb Tb = Fa Eb Tb. \tag{43}$$

3.6 A-Terminals' Characteristics

In this subsection, analytical expressions characterizing all A-terminals (traffic intensity (Ya), intensity of occupation flow (Fa), holding time (Ta)) are obtained, as functions of known variables.

Proposition 5. A-terminals' traffic intensity (Ya) is:

$$Ya = Ya0 + Ya1 + Ya2 + Ya3 = Fa Ta, \tag{44}$$

where

$$Ya0 = Yed + Yad + Yid + Ycd + Ybs + Yis + Yns + Ycs \tag{45}$$

$$Ya1 = Yb1, \tag{46}$$

$$Ya2 = Yah + Ych + Yb2, \tag{47}$$

$$Ya3 = Yem + Yam + Yb3. \tag{48}$$

Proof: Based on the proposed conceptual model and Figs. 7, 8, 9 and 10, and by applying the Little's theorem, we can obtain the traffic intensity for each virtual device, in $a0$ ($Ya0$), $a1$ ($Ya1$), $a2$ ($Ya2$), and $a3$ ($Ya3$) blocks, of stages Dialing (Yed, Yad, Yid, Ycd), Switching (Ybs, Yis, Yns, Ycs), Holding (Yah, Ych), and Mailing (Yem, Yam), and by using the found traffic intensities of B-terminals ($Yb1, Yb2, Yb3$).

After summation, we obtain the following:

$$\begin{aligned}
 Ya = Fa Ta = Fa (Ted + Pad Tad + (1 - Pad)(Pid Tid + (1 - Pid)(Tcd + Pbs Tbs \\
 + (1 - Pbs)(Pis Tis + (1 - Pis)(Pns Tns + (1 - Pns)(Tcs + (1 - Pbr) Tb1 + Pbr (Tbr \\
 + Ph (Pah Tah + (1 - Pah)(Tch + Tb2) \\
 + (1 - Ph) Pm (Tem + Pam Tam + (1 - Pam) Tb3)))))))). \tag{49}
 \end{aligned}$$

Proposition 6. By distinguishing static and dynamic parameters, the mean holding time Ta of A-terminals is:

$$\begin{aligned}
 Ta = S1 - S2(1 - Pbs)Pbr - S3 Pbs - Ecs (Tb1 + Pbr (-Tb1 \\
 + Ph (1 - Pah) Tb2 + (1 - Ph) Pm (1 - Pam) Tb3)), \tag{50}
 \end{aligned}$$

where $S1, S2$, and $S3$ are generalized static parameters:

$$\begin{aligned}
 S1 = Ted + Pad Tad + (1 - Pad)(Pid Tid + (1 - Pid)(Tcd + Pis Tis \\
 + (1 - Pis)(Pns Tns + (1 - Pns)(Tcs + 2Tb1))). \tag{51}
 \end{aligned}$$

$$\begin{aligned}
 S2 = (1 - Pad)(1 - Pid)(1 - Pis)(1 - Pns)(2Tb1 - Tbr - Ph (Pah Tah \\
 + (1 - Pah)(Tch + 2Tb2)) - (1 - Ph) Pm (Tem + Pam Tam + (1 - Pam) 2Tb3)) \tag{52}
 \end{aligned}$$

$$\begin{aligned}
 S3 = (1 - Pad)(1 - Pid)(Pis Tis - Tbs + (1 - Pis)(Pns Tns + (1 - Pns)(Tcs + 2Tb1))) \tag{53}
 \end{aligned}$$

Proof: Based on (49) in Proposition 5, obviously:

$$\begin{aligned}
Ta = & Ted + Pad Tad + (1 - Pad)(Pid Tid + (1 - Pid)(Tcd + Pbs Tbs + (1 - Pbs)(Pis Tis \\
& + (1 - Pis)(Pns Tns + (1 - Pns)(Tcs + (1 - Pbr) Tb1 + Pbr (Tbr + Ph (Pah Tah \\
& + (1 - Pah)(Tch + Tb2) + (1 - Ph)Pm (Tem + Pam Tam + (1 - Pam) Tb3)))))))).
\end{aligned} \tag{54}$$

After simple mathematical transformations we obtain (50).

3.7 QoS Indicator 3: Overall Call Attempt Efficiency

Definition 2.12 in [11] describes “successful call” as “A call that has reached the wanted number and allows the conversation to proceed”. Note that ‘call’ is “A generic term related to the establishment, utilization and release of a connection. Normally a qualifier is necessary to make clear the aspect being considered, e.g. call attempt.” [11]. A ‘call attempt’ is “An attempt to achieve a connection to one or more devices attached to a telecommunications network.” Therefore, a call may content several call attempts.

Based on this, we define the Overall Call Attempt Efficiency (Ec), of a communication service, as a ratio of the flow intensity (Fc) of the calls attempts with a fully and successfully finished communication, to the incoming call attempts’ intensity (Fa).

In the considered conceptual model, Fc is a sum of flow intensities of virtual devices $cc1$ (Case 1 in Fig. 8), $cc2$ (Case 2 in Fig. 9), and cm (Case 3 in Fig. 10):

$$Fc = Fcc1 + Fcc2 + Fcm. \tag{55}$$

Then the Overall Call Attempt Efficiency (Ec) is:

$$\begin{aligned}
Ec = \frac{Fc}{Fa} = & Ecs((1 - Pbr)(1 - Par)(1 - Pac) + Pbr(Ph(1 - Pah)(1 - Pac) \\
& + (1 - Ph)Pm(1 - Pam)))
\end{aligned} \tag{56}$$

3.8 Network Generalized Subservice Indicators

The Overall Call Attempt Efficiency (Ec) obviously includes the described indicators Carried Switching Efficiency (Ecs) and B-Terminal Connection Efficiency (Eb). From users’ and service providers’ point of view, it is important to distinguish the efficiency of the subservices of the telecommunication system. Such subservices include: switching (finding B-terminal), connection to B-terminal, finding B-user, transmission of messages (communication). Here we introduce specific QoS indicators for each of these subservices, as parts of the Overall Call Attempt Efficiency (Ec).

As a QoS-specific indicator of the switching subservice (finding B-terminal), the Carried Switching Efficiency (Ecs), proposed in (9), could be used, i.e. as the ratio of the flow intensity of the calls reaching the intended B-terminal (Fcs) and receiving either a ‘busy tone’ or a ‘ringing tone’ signal, to the incoming call attempt intensity (Fa):

$$Ecs = \frac{Fcs}{Fa}. \tag{57}$$

Definition 1. A QoS-specific indicator (Qb) of the subservice ‘Connection to B-terminal’ is the ratio of the intensity of the flow seizing B-terminals (Fb) to the flow intensity of all calls reaching the intended B-terminal (Fcs):

$$Qb = \frac{Fb}{Fcs}. \quad (58)$$

Definition 2. A QoS-specific indicator (Qu) of the subservice ‘Finding B-user’ is the ratio of the intensity of the flow seizing B-users (Fu) to the intensity of the flow seizing B-terminals (Fb):

$$Qu = \frac{Fu}{Fb}. \quad (59)$$

The intensity of the flow seizing B-users (Fu) is a sum of intensities of the flows: after ringing $Fb1 - (Far + Fcr)$ in Case 1 (c.f. Fig. 8); after holding $Fb2$ in Case 2 (c.f. Fig. 9); and of the carried mailing $Fb3$:

$$Fu = Fb1 - Far - Fcr + Fb2 + Fb3. \quad (60)$$

Definition 3. A QoS-specific indicator (Qc) of the communication subservice is the ratio of the flow intensity of call attempts with fully successfully finished communication (Fc) to the intensity of the flow seizing B-users (Fu):

$$Qc = \frac{Fc}{Fu}. \quad (61)$$

The proposed specific QoS indicators of telecommunication subservices are aggregated because: they aggregate many call attempts from many users and terminals (they are stochastic); some of them comprise several parallel services, e.g. Qc includes three successful cases – normal interactive communication, communication after call holding, and mailing.

Considering the Overall Call Attempt Efficiency (Ec) as a composition of the four considered subservices, one may find that the quality metric is multiplicative:

$$Ec = \frac{Fc}{Fa} = \frac{Fcs}{Fa} \frac{Fb}{Fcs} \frac{Fu}{Fb} \frac{Fc}{Fu} = Ecs Qb Qu Qc. \quad (62)$$

This result allows more specific QoS analysis and more adequate QoS management.

3.9 AB-Terminals’ Characteristics

In this subsection, analytical expressions of characteristics of AB-terminals (all occupied calling terminals (A) and called terminals (B)) – i.e. traffic intensity (Yab), intensity of occupation flow (Fab), and holding time (Tab) – are obtained as functions of known variables.

From the assumptions made and the conceptual model proposed in Subject. 2.2, it is clear that the intensity of the call flows occupying all terminals (F_{ab}) is a sum of intensities of the call flows occupying A-terminals (F_a) and the call flows occupying B-terminals (F_b):

$$F_{ab} = F_a + F_b. \quad (63)$$

The traffic intensity of all terminals (Y_{ab}) is a sum of traffic intensity of the A- (Y_a) and B-terminals (Y_b):

$$Y_{ab} = Y_a + Y_b. \quad (64)$$

Proposition 7. The call flows intensity occupying all terminals (F_{ab}) can be obtained by the following equation:

$$F_{ab} = F_a (1 + Ecs((1 - Pbr) + Pbr(Ph(1 - Pah) + (1 - Ph)Pm(1 - Pam)))) = F_a(1 + Eb), \quad (65)$$

where Ecs is the Carried Switching Efficiency (9) and Eb is the B-efficiency (33).

Proof: It can be easily seen that (65) follows directly from (33) and (63).

Proposition 8. The traffic intensity of all terminals (Y_{ab}) can be presented by the following expression:

$$Y_{ab} = F_a (T_a + E_b T_b). \quad (66)$$

Proof: (66) follows directly from (43), i.e.:

$$Y_{ab} = Y_a + Y_b = F_a T_a + F_b T_b = F_a T_a + F_a E_b T_b = F_a (T_a + E_b T_b). \quad (67)$$

Terminal Traffic Limitations. Since the number of terminals is limited to N_{ab} (Assumption 2), and there is no negative occupancy, the following terminal traffic limitations obviously apply in the studied system:

$$0 \leq Y_{ab} \leq N_{ab}. \quad (68)$$

Proposition 9. Traffic of all simultaneously busy terminals (Y_{ab}), after separation of static parameters from dynamic parameters, may be expressed from Eqs. (50) and (66) as:

$$Y_{ab} = F_a (S_1 - S_2 (1 - Pbs) Pbr - S_3 Pbs), \quad (69)$$

where S_1 , S_2 , and S_3 are generalized static parameters as per (51), (52), and (53).

Proof: Based on (64), (49) and (41) we obtain:

$$\begin{aligned}
 Yab = & Fa (Ted + Pad Tad + (1 - Pad)(Pid Tid + (1 - Pid)(Tcd + Pbs Tbs \\
 & + (1 - Pbs)(Pis Tis + (1 - Pis)(Pns Tns + (1 - Pns)(Tcs + (1 - Pbr) 2 Tb1 + Pbr (Tbr \\
 & + Ph (Pah Tah + (1 - Pah)(Tch + 2 Tb2) \\
 & + (1 - Ph) Pm (Pam Tam + (1 - Pam) 2 Tb3)))))))).
 \end{aligned} \tag{70}$$

After algebraic transformation and taking into account (51), (52), and (53), we obtain (69).

Proposition 10. The mean occupation time (Tab) of all simultaneously busy terminals can be obtained from (70) as a function of Ta , Tb , and Eb .

Proof: From the obvious formula $Yab = Fab Tab$, after replacing Yab with (66), Fab with (63), and Fb with (32), we have:

$$Tab = \frac{Yab}{Fab} = \frac{Ya + Yb}{Fa + Fb} = \frac{Fa Ta + Fb Tb}{Fa + Fb} = \frac{Fa Ta + Fa Eb Tb}{Fa + Fa Eb} = \frac{Ta + Eb Tb}{1 + Eb}.$$

Proposition 11. The mean occupation time (Tab) of all simultaneously busy terminals can be obtained from (71) as a function of $S1$, $S2$, $S3$, Pbr , Pbs , and Eb .

Proof: From the formula $Yab = Fab Tab$, and after replacing Yab with (69) and Fab with (65), we obtain:

$$Tab = \frac{Yab}{Fab} = \frac{S1 - S2 (1 - Pbs) Pbr - S3 Pbs}{1 + Eb}. \tag{71}$$

3.10 Offered Traffic to the Switching System

Following the definition of equivalent traffic offered to the switching system, traffic ($ofr.Ys$) depends on the offered flow intensity ($ofr.Fs$) and the occupation (service) time Ts of an equivalent switching line:

$$ofr.Ys = ofr.Fs Ts. \tag{72}$$

The offered flow to the switching system is the flow offered to the first Enter Switch device in Fig. 7. This device deflects calls, if there is no free line in the switching system, with probability of blocked switching (Pbs) to the Blocked Switching (bs) device, or with probability $(1 - Pbs)$ of calls seizing free equivalent switching lines. So the offered flow intensity $ofr.Fs$ is:

$$ofr.Fs = Fa(1 - Pad)(1 - Pid). \tag{73}$$

The occupation (service) time of an equivalent switching line (Ts) is determined by the engaged devices of the switching system (c.f. Subsect. 2.2), namely the s device, represented by a box with a dotted dashed line inside the $a0$ box in Fig. 7, and three other a -boxes ($a1$ in Fig. 8, $a2$ in Fig. 9, and $a3$ in Fig. 10). So consequently:

$$Ts = S1z - S2z Pbr, \quad (74)$$

where:

$$S1z = PisTis + (1 - Pis)(Pns Tns + (1 - Pns)(Tcs + Tb1)); \quad (75)$$

$$S2z = Tb1 - Tbr - Ph(Pah Tah + (1 - Pah)(Tch + Tb2)) - (1 - Ph)Pm(Tem + Pam Tam + (1 - Pam) Tb3) \quad (76)$$

Probability of Blocked Switching

Proposition 12. The probability of blocked switching (Pbs) could be obtained from (72) as:

$$Pbs = Erl_b(Ns, ofr.Ys). \quad (77)$$

Proof: (77) simply expresses the usage of the Erlang-B formula for determination of the blocking probability in the switching system, on the basis of the number of equivalent internal switching lines (Ns) and the offered traffic $ofr.Ys$.

Probability of Blocked Ringing (B-terminal Busy). Under Assumptions 4, 12, 14, the following expressions, presenting the probability of blocked ringing (Pbr) as a function of the network state Yab (traffic of all A- and B-terminals) and the number Nab of all active terminals in the system, could be obtained:

$$\left| \begin{array}{l} Pbr = \frac{Yab - 1}{Nab - 1} \text{ if } 1 \leq Yab \leq Nab, \\ Pbr = 0 \text{ if } 0 \leq Yab < 1. \end{array} \right. \quad (78)$$

(78) was first proposed as part of the simple overall network teletraffic model, described in [24], and its proof was given in [25].

4 Results

4.1 QoS Indicator 4: Overall Traffic Efficiency Indicator

Based on the “effective traffic” definition [11] as “The traffic corresponding only to the conversational portion of effective call attempts”, we define the Overall Traffic Efficiency Indicator (Ey) as a ratio of the effective traffic of A-terminals (Ycc) to the overall traffic of the A-terminals (Ya):

$$Ey = \frac{Ycc}{Ya}, \quad (79)$$

where

$$Ycc = Ycs + Ycc1 + Ycc2 + Ycm. \quad (80)$$

The Overall Traffic Efficiency Indicator is used for simpler models in [7]. Some authors use the name “Overall Traffic Efficiency” in other meaning, and without any definition, e.g. [26].

4.2 Numerical Results

The input data considered is typical for voice communications in the Global System for Mobile communication (GSM). For simplicity we set M (defined in the explanations of Fig. 7) to 0.

Figure 11 presents results (as functions of the state of the network load – the traffic of all AB-terminals Y_{ab} , in the theoretical interval $[0, 100]$) for a network with blocking probability due to insufficient resources. The number of all terminals (N_{ab}) in the system is 1000 and the number of equivalent switching lines is $N_s = 200$ (i.e. 20% of N_{ab}).

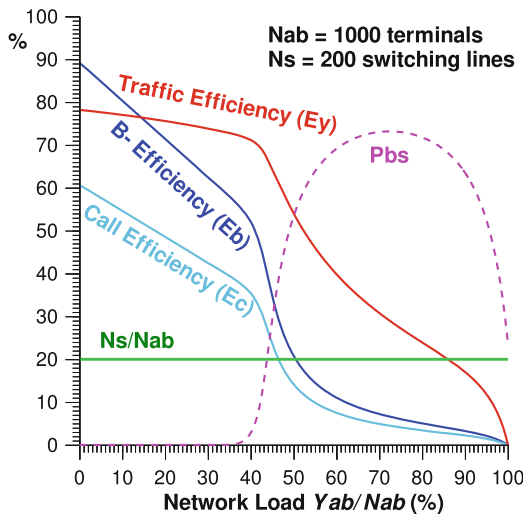


Fig. 11. The values of the main output parameters of the model of an overall network with limited capacity.

The probability of finding B-terminal busy (P_{br}), not shown in Fig. 11, increases linearly with the network traffic load, c.f. (78), to almost 1. The numerical results demonstrate the existence of a local maximum for the probability of blocked switching P_{bs} . This is because the overall blocking probability in the network, including P_{br} and P_{bs} , has an absolute maximum of 1, c.f. Fig. 12.

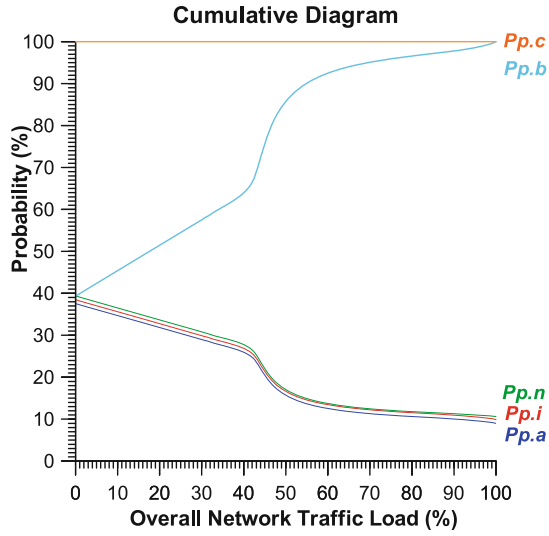


Fig. 12. The causal aggregated overall pie probabilities

4.3 Overall Pie Parameters

In the model considered (c.f. Figs. 7, 8, 9 and 10), there are five reasons for call attempt ending: abandoning (6 branches), interruption (1 branch), blocking (2 branches), unavailable service (1 branch), and successful communication (3 branches). By describing the effect caused by each reason, one can construct a 'causal branch' for it. The causal branch comprises all basic virtual devices involved in the call attempt ending due the considered reason, which form the corresponding causal complex virtual device with its flow-, time-, and traffic characteristics. Overall, in the model, there are 13 causal branches considered.

The three branches of successful communication have the following service times $T_{p.cc1}$, $T_{p.cc2}$, and $T_{p.cm}$:

$$T_{p.cc1} = T_{ed} + T_{cd} + T_{cs} + T_{cr} + T_{cc1}; \quad (81)$$

$$T_{p.cc2} = T_{ed} + T_{cd} + T_{cs} + T_{ch} + T_{cc2}; \quad (82)$$

$$T_{p.cm} = T_{ed} + T_{cd} + T_{cs} + T_{em} + T_{cm}. \quad (83)$$

The pie flow intensities, of the three subcases of successful communication, coincide with the flow intensity of the last virtual device in the causal branch, respectively: $F_{p.cc1} = F_{cc1}$; $F_{p.cc2} = F_{cc2}$ and $F_{p.cm} = F_{cm}$.

The pie flow probabilities of the three branches of successful communication respectively are:

$$Pp.cc1 = \frac{Fp.cc1}{Fa}; Pp.cc2 = \frac{Fp.cc2}{Fa}; Pp.cm = \frac{Fp.cm}{Fa}. \quad (84)$$

The pie traffic intensities of the three branches of successful communication are:

$$Yp.cc1 = \frac{Fp.cc1 Tp.cc1}{Ya}; Yp.cc2 = \frac{Fp.cc2 Tp.cc2}{Ya}; Yp.cm = \frac{Fp.cm Tp.cm}{Ya}. \quad (85)$$

By analogy, one may easily obtain all other overall pie probabilities, pie flows, and pie traffic intensities in the model, by using the normalized parameters found in Sect. 3.

4.4 Causal Aggregated Overall Pie Parameters

The overall causal branches may be aggregated as might be needed for telecommunication system monitoring, design, or management. A usable aggregation is the causal aggregation of all the branches corresponding to one type of call attempts ending.

For instance, for the case of successful communication, one can express the aggregated parameters of the branches of the Aggregated Overall Successful Carried Communication Branch, considered as a complex virtual device *p.c*. The metrics are additive because this is a pie presentation of the model.

The causal aggregated overall pie probability of a call attempt ending with successful communication (*Pp.c*) is:

$$Pp.c = Pp.cc1 + Pp.cc2 + Pp.cm. \quad (86)$$

By taking into account (56), the overall causal pie flow intensity of successful communication (*Fp.c*) respectively is:

$$Fp.c = Fp.cc1 + Fp.cc2 + Fp.cm = \frac{Fcc1 + Fcc2 + Fcm}{Fa} = \frac{Fc}{Fa} = Ec. \quad (87)$$

The overall causal pie traffic intensity of successful communication (*Yp.c*) is:

$$Yp.c = Yp.cc1 + Yp.cc2 + Yp.cm. \quad (88)$$

Similarly, one may find all other causal aggregated overall pie parameters of the model.

4.5 Numerical Results for Pie Characteristics

Figures 12 and 13 present numerical results for the causal overall pie probabilities and traffic intensities for each of the five reasons for call attempt ending (i.e. abandoning *p.a*, interrupting *p.i*, blocking *p.b*, service not available *p.n*, and successful communication *p.c*) as functions of the network traffic load.

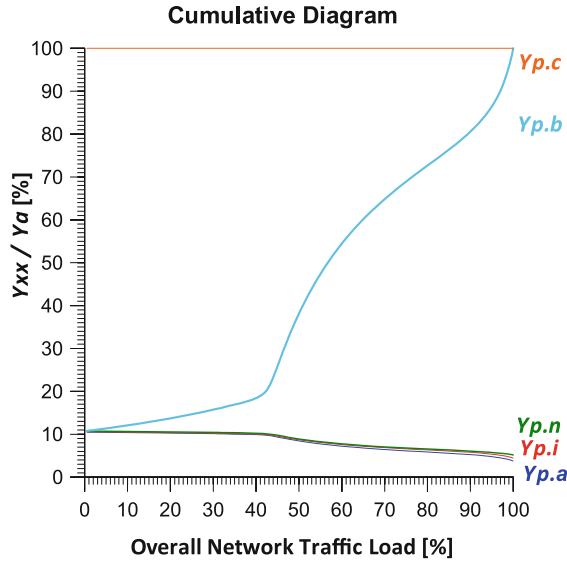


Fig. 13. The causal aggregated overall pie traffic intensities

The overall pie characteristics and their causal aggregations may be considered and used as causal-oriented QoS indicators. They allow more precise estimation of the dynamic importance of each reason for call attempt ending and thus a more precise dynamic effort targeting of the QoS management.

5 Conclusion

The presented modeling approach and corresponding numerical results demonstrate the big potential and importance of the overall teletraffic models of telecommunication systems with QoS guarantees.

Such models allow prediction of many overall QoS indicators as regards the flow-, time-, and traffic characteristics of the A-, B-, and AB-terminals and users, as well as of the overall network performance.

The approach makes easy the separation of an overall telecommunication service into different subservices with specific QoS indicators for each of them.

In this chapter, the newly proposed indicators are network-oriented or terminal-oriented. The model, however, is suitable for the development of user-oriented indicators as well. This will be a task for future research.

Applying pie characteristics and their causal aggregations to the subservices results in causal-oriented QoS indicators. This allows a more precise estimation of the dynamic importance of each reason, in every subservice, of call attempt ending, and thus a more precise dynamic effort targeting of the QoS management. Applying a similar approach (with specific QoS indicators) for multimedia and multiservice networks seems very attractive and promising.

Another important goal could be the development of methods for using specific QoS indicators as sources for predicting QoE indicators.

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. ITU-T Recommendation E.800: Definitions of terms related to quality of service, September 2008
2. Martin, V., 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. <https://doi.org/10.1109/iccw.2015.7247432>. ISSN 2164-7038
3. Möller, S., Raake, A.: Quality of Experience: Advanced Concepts, Applications and Methods. Springer, Cham (2014). <https://doi.org/10.1007/978-3-319-02681-7>
4. Reichl, P.: Quality of experience in convergent communication ecosystems. In: Lugmayr, A., Dal Zotto, C. (eds.) The Media Convergence Handbook, vol. 2, pp. 225–244. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-642-54487-3_12
5. Fiedler, M.: Teletraffic models for Quality of Experience assessment. In: Tutorial at 23rd International Teletraffic Congress (ITC 23), San Francisco, CA, September 2011. http://i-teletraffic.org/_Resources/Persistent/9269df1c3dca0bf58ee715c3b9afabb71d4fb26/fiedler11.pdf. Accessed 20 July 2017
6. Poryazov, S.A., Saranova, E.T.: Some general terminal and network teletraffic equations in virtual circuit switching systems. In: Nejat Ince, A., Topuz, E. (eds.) Modeling and Simulation Tools for Emerging Telecommunications Networks, Chap. 24, pp. 471–505. Springer, Boston (2006). https://doi.org/10.1007/0-387-34167-6_24. ISBN-10 0-387-32921-8 (HB)
7. Poryazov, S., Saranova, E.: Models of Telecommunication Networks with Virtual Channel Switching and Applications, p. 238. Prof. Marin Drinov, Academic Publishing House, Sofia (2012). ISBN 978-954-322-540-8
8. Ericsson, T.: Understanding Telecommunications, Book 2, 673 pages. Chartwell-Bratt, Bromley (1998). ISBN 91-44-00214-9
9. Poryazov, S., Saranova, E.: Overall QoS referencing in telecommunication systems – some current concepts and open issues. *Int. J. Inf. Technol. Knowl.* **5**(4), 424–452 (2011). ITH EA. ISSN 1313-0455
10. Soanes, C., Stevenson, A. (eds.): Concise Oxford English Dictionary, 11th edn. Oxford University Press, Oxford (1911)
11. ITU-T Recommendation E.600: Terms and definitions of traffic engineering, March 1993
12. Böhm, C., Jacopini, G.: Flow diagrams, Turing machines and languages with only two formation rules. *Comm. ACM* **9**, 366–371 (1966)
13. ITU-T Recommendation Q.850: Usage of cause and location in the Digital Subscriber Signalling System No. 1 and the Signalling System No. 7 ISDN user part, May 1998
14. ITU-T Recommendation E.425: Network management – Internal automatic observations, March 2002
15. Cisco Unified Communications Manager Call Detail Records Administration Guide, Release 8.5(1). Text Part Number: OL-22521-01, p. 172. Cisco Systems (2010)

16. Henggeler, A.C., Craveirinha, J., Clímaco, J., Barrico, C.: A multiple objective routing algorithm for integrated communication networks. In: Key, P., Smith, D. (eds.) Proceedings of the ITC-16, vol. 3B, p. 14. Elsevier Science B. V. (1999). http://di.ubi.pt/~cbarrico/Investigacao/Downloads/16th_International_Teletraffic_Congress_99.pdf. Accessed 17 July 2017
17. ITU-T Recommendation Y.1541: Network performance objectives for IP-based services, February 2006
18. ITU-T Recommendation E.360.1: Framework for QoS routing and related traffic engineering methods for IP-, ATM-, and TDM-based multiservice networks, May 2002
19. Poryazov, S., Saranova, E.: On the minimal state tuple of VNET with overall QoS guarantees. In: IEEE 19th Telecommunications Forum TELFOR 2011, Belgrade, Serbia, 22–24 November 2011. IEEE Catalog Number CFP1198P-CDR. ISBN 978-1-4577-1498-6
20. ITU-T Recommendation X.501: International standard ISO/IEC 9594-2, ITU-T Recommendation X.501. Information technology – Open Systems Interconnection – The Directory: Models, August 2005
21. ITU-T Recommendation E.501: Estimation of Traffic Offered in The Network, May 1997
22. Poryazov, S.A.: What is offered traffic in a real telecommunication network? In: Liang, X.J., Xin, Z.H., Iversen, V.B., Kuo, G.S. (eds.) 19th International Teletraffic Congress, Beijing, China, August 29–September 2 2005, vol. 6a, pp. 707–718. Beijing University of Posts and Telecommunications Press. https://itc-conference.org/_Resources/Persistent/321454b378f49b23d9a04e23dd5a5fe58da7ad60/poryazov05.pdf. Accessed Mar 2017
23. Little, J.D.C.: A proof of the queueing formula $L = \lambda W$. *Oper. Res.* **9**(1961), 383–387 (1961)
24. Poryazov, S.: Determination of the probability of finding B-subscriber busy in telephone systems with decentralized control. *Comptes Rendus de l'Academie Bulgare des Sciences – Sofia* **44**(3), 37–39 (1991)
25. Poryazov, S.: The B-terminal busy probability prediction. *IJ Inf. Theor. Appl.* **11**(4), 409–415 (2004)
26. de Souza, A.M., Brennard, C.A.R.L., Yokoyama, R.S., Donato, E.A., Madeira, E.R.M., Villas, L.A.: Traffic management systems: a classification, review, challenges, and future perspectives. *Int. J. Distrib. Sens. Netw.* **13**(4), 14 (2017). <https://doi.org/10.1177/1550147716683612>

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.

