Proceedings of 2020 IEEE 10th International Conference on Intelligent Systems

Traffic Quality Aggregations of a Queuing System

Stoyan Poryazov
Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Sofia, Bulgaria
stoyan@math.bas.bg

Velin Andonov
Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Sofia, Bulgaria
velin_andonov@math.bas.bg

Emiliya Saranova
Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
emiliya@math.bas.bg

***Abstract*—The problem of the traffic quality of a queuing system (QS), presented as an aggregation of the QS components' parameters, is studied. The QS is considered as a portion of an informational service network. Developments of the conceptual modeling approach, used by the authors, are proposed. The notation of the conceptual virtual devices and their parameters is extended. The concept of partial service time is considered. This allows easier and clearer graphical and analytical model presentations and defining QS causal Quality of Service (QoS) indicators. The results obtained are useful for QoS measurements, prediction and management in real informational service networks.**

Keywords-service networks; causal structure; conceptual models; traffic quality indicators; quality indicators composition.

I. INTRODUCTION

The importance of Quality of Service (QoS) indicators grows with the usage of the informational service networks and it became a commodity in 2015 [1]. The QoS and Quality of Experience (QoE) are defined in different ways, but we will follow the definition in standardization documents such as the ITU-T [2]. The prediction of the overall network quality, as a function of qualities of composed services, is a foremost question in service networks design and maintenance. There are two main approaches of QoS aggregation – analytical (e.g. [3]) and simulational (e.g. [4]). In this paper, conceptual and corresponding graphical approaches for QoS presentation are used. The approach is a considerable development of the results in [5, 6].

In Section II, causal and fictive functional types of base virtual devices are added to the conceptual modeling.

In Section III, partialized and normalized conceptual models of an internal structure of a virtual service device are considered. This furnishes the development of the notation system of devices in causal device structure, described in the Section IV.

In Section V, a causal structure of a virtual device with limited capacity is considered, as a preparation to the formulation of the causal traffic quality indicators, in Section VI.

In Section VII, a new conceptual model of a queuing system (QS), comprising a buffer with limited capacity, is proposed. This makes the derivation of analytical dependences among parameters of the queuing system's causal sub-devices, in Section VIII, easier.

In Section IX, the effectiveness of the proposed approach is demonstrated through deducing a causal traffic quality indicator of the QS and a presentation of a causal flow quality indicator of the QS as a function of the quality indicators of the buffer and the server.

II. BASE VIRTUAL DEVICES' PARAMETERS AND FUNCTIONAL TYPES

In the conceptual models of service networks, we use base virtual devices [6]. A general representation of a base virtual device parameters., is shown in Fig. 1.

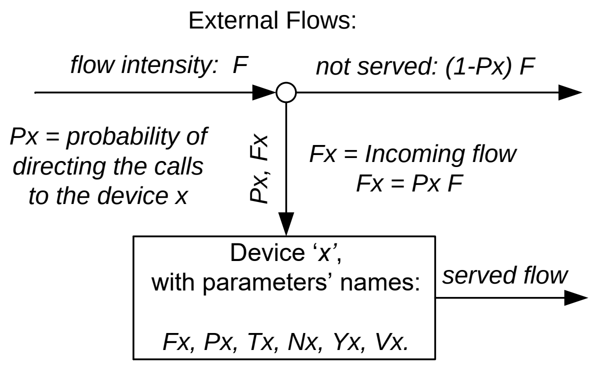

Figure 1. Graphical representation of a base virtual device *x*

Every such device named x has the following parameters:

The work of V. Andonov was supported by the Bulgarian NSF under grant DM 12/2 – "New Models of Overall Telecommunication Networks with Quality of Service Guarantees". The work of E. Saranova was supported by the Task 1.2.5 of the Contract MES DOI-205/23.11.2018. The work of S. Poryazov was supported by the Joint Research Project (No 18 / 02.04.2019) of the IMI-BAS and Joint Institute for Nuclear Research (JINR), Dubna, Russia.

Fx – Intensity or incoming rate (frequency) of the flow of requests (i.e., the number of requests per time unit) to device *x*;

Px – Probability of directing the requests towards device *x*;

Tx – Service time (duration of servicing of a request) in device *x*;

Yx – Traffic intensity [Erlang];

Vx – Traffic volume [Erlang – time unit];

Nx – Number of lines (service resources, positions, capacity) of device *x*.

In the conceptual models of telecommunication systems, the following functional types of base virtual devices are used: Director, Terminator, Server, Switch, Causal Device, and Fictive Device. Their graphical representations are shown in Fig. 2.

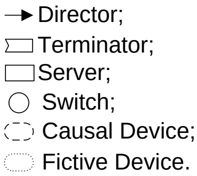

Figure 2. Graphical functional types representation of the base virtual devices

Each type of the base virtual devices has specific function:

- Director – this device unconditionally points to the next device, which the request shall enter, but without transferring, changing or delaying it;
- Terminator – this device eliminates every request entered (so it leaves the model without any traces);
- 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 processsing (cf. Fig. 2);
- Switch (Transition) – this device selects one) – this device selects one of its possible exits for each request entered, thus determining the next device where this request shall go to;
- Causal device – virtual device defined for presentation of causes of service ending, e.g. successful (carried) or not (interrupted, abandoned, etc.);
- Fictive device – device presenting fictive traffic which is necessary for engineering. For example, not carried traffic is fictive, but is used for calculating the equivalent offered traffic [11], which is necessary for device dimensioning.

The base virtual devices do not contain other devices. In In our models of telecommunication networks, we also use comprise virtual devices which include base virtual devices.

III. Partialized and Normalized Conceptual Models

Any service may end due to many reasons. Cisco lists 131 ‘call termination cause codes’ and 44 ‘Cisco-specific call termination cause codes’ [8]. Each reason for service ending has its own probability to occur and mean service time (duration). The service branches, corresponding to the reasons of service ending, form the ‘causal structure’ of the modeled service (see Fig. 3). In [7] two possible internal causal structures of the device *x* are presented: normalized (Fig. 4) and partialized (Fig. 5).

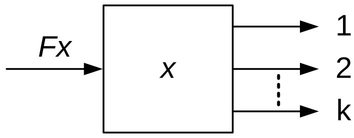

Figure 3. Virtual device ***x***, representing a service with *k* reasons for ending

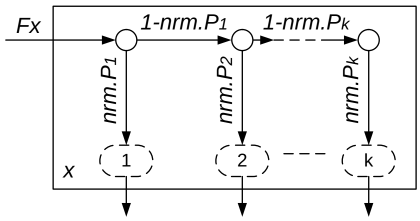

Figure 4. Normalized causal internal structure of device ***x***

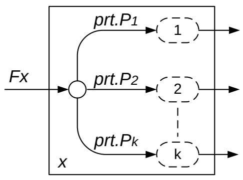

Figure 5. The partializedcausal structure of device ***x***

The three presentations (in Fig. 3, 4 and 5) of the virtual device *x* have the same input requests’ flow frequency F_x, mean service time T_x traffic intensity Y_x and k output flows with the same frequencies F_i ($i \in [1, k]$).

The difference among them is in the internal structures presentations only. Both structures include *k* virtual ‘causal devices’, each with its different mean input requests’ flow frequency F_i, mean service time T_i and traffic intensity Y_i, but the internal structures of device i in normal and in partial model may be completely different. Each of the k causal branches in Fig. 4 and Fig. 5 may consist of many sub-servicing devices (usually different for normalized and partial models), comprised in different virtual devices name t, $i \in [1, k]$ in the both figures.

Obviously, in the both presentations:

$$Y_x = \sum_{i=1}^{k} Y_i; \; Y_x = \sum_{i=1}^{k} Y_i. \qquad (1)$$

The internal probabilities are different. In the normalized structure (Fig. 4) 'normalized probabilities $nrm.P_i$ are independent and their sum is not equal to 1.

In the partial causal structure (Fig. 5), all causal service branches have common beginning. The partial probability $prt.P_i$ shows what part of the service incoming flow is directed to the causal device i:

$$prt.P_i = \frac{F_i}{F_x}. \qquad (2)$$

All probabilities $prt.P_i$ are mutually dependent:

$$\sum_{i=1}^{k} prt.P_i = 1. \qquad (3)$$

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 partial 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 and we will use them in the indicators definition.

IV. Causal Structure and Device Notation

For the generalized comprising and/or causal devices we use qualifiers. The flow intensity in the service branch i, in a partial-model, is called 'partial-flow intensity' ($prt.Fi$). From (1):

$$F_x = \sum_{i=1}^{k} prt.F_i = \sum_{i=1}^{k} prt.P_i \, F_x. \qquad (4)$$

The traffic intensity in branch i is called 'partial traffic intensity' and notated by $prt.Y_i$ (in Fig. 5, $prt.Yi=Y_i$, see (1)). Following (4) and the Theorem of Little [9] we have:

$$prt.Y_i = prt.F_i \, T_i = F_x \, prt.P_i \, T_i \qquad (5)$$

where T_i is the service time duration in device i.

We define 'partial service time' in device i ($prt.Ti$) by the expression:

$$prt.T_i = prt.P_i \, T_i. \qquad (6)$$

The partial service time in device i ($prt.T_i$) represents what part of the service time ($srv.T_x$) of the comprising device x takes the service in device i (see Fig. 5).

Proposition 1: $srv.T_x = \sum_{i=1}^{k} prt.T_i$.

Proof: From the Theorem of Little, Fig. 5, (1), (5) and (6) follows:

$$srv.T_x = \frac{\sum_{i=1}^{k} Y_i}{\sum_{i=1}^{k} F_i} = \frac{\sum_{i=1}^{k} F_i T_i}{F_x} = \sum_{i=1}^{k} prt.T_i = \\ = \sum_{i=1}^{k} prt.P_i T_i = \sum_{i=1}^{k} prt.T_i. \qquad (7)$$

Remark: Obviously,

$$prt.srv.Tx = srv.Tx, \; prt.srv.Yx = srv.Yx, \; prt.srv.Fx = srv.F_x,$$

because there is not a device comprising x (see Fig. 5).

We group causes of service ending, and corresponding causal devices, in generalized comprising causal devices. If device x has unlimited capacity (Fig. 6), three causal generalizations are enough: 'parasitic', 'carried' and 'served' [6].

The Parasitic Traffic in a pool of resources is the traffic, which was unsuccessfully served in the pool. Parasitic traffic occupies real resources but not for a useful service execution.

The Carried Traffic in a pool of resources is the traffic, which was successfully served in the pool (and carried to the next service device). We distinguish two types of carried traffic:

- 'zero service' e.g. zero waiting in a buffer if the buffer is empty and there is free requested place for the service in the following device. The requests are receiving zero service in the causal 'zero service device' and may be served without delay;
- 'genuine service' – successfully and real served requests in the pool. The service time is noticeable.

The Served Traffic in a pool of resources is any traffic, occupying (using) resources in the pool. The Served Traffic is a sum of carried and parasitic traffic.

Every Causal Device Parameter's Name is a concatenation:

$$\text{Causal Name} = \{ \langle \text{ qualifier } \rangle \langle \text{ qualifier } \rangle \} \\ \bullet \langle \text{ Parameter's Symbol} \rangle \bullet \langle \text{ Device Name } \rangle$$

In this paper, used qualifiers are:

crr. = carried;
gen. = genuine;
nrm. = normalized
nsr. = not served;
ofr. = offered;
prs. = parasitic;
prt. = partial;
srv. = served;
zer. = zero.

'**Parameter's symbol**' is one of letters P, F, T, Y, V, N, as they are described in Section II.

Qualifiers are used to characterize the parameters of the devices [6]. Used qualifiers may be two, one or none.

If parameter's symbol is omitted, the causal name is a name of a device (Fig. 6).

Device name may be in small or subscript letters.

For instance, *crr.Fx* is the intensity of the carried flow of requests of the device *x*, *prs.Fx* is the intensity of the parasitic flow of requests of the device *x* (see Fig. 6).

In the figures, only the names of causal devices may be presented. The names of device parameters are implicit, according to Sections II. In Fig. 6 there are three generalized conceptual devices: *x*, *prs.x*, and *crr.x*.

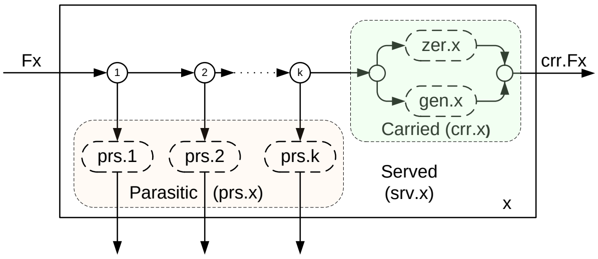

Figure 6. Generalization of the causal structurex of a virtual device *x* with an unlimited capacity

For the causal presentation shown in Fig. 6 the following obvious equalities hold:

$$prs.Yx = prsY_1 + prsY_2 + \ldots + prsY_k; \tag{8}$$

$$crr.Yx = zer.Yx + gen.Yx; \tag{9}$$

$$srv.Yx = prs.Yx + crr.Yx; \tag{10}$$

$$crr.Fx = zer.Fx + gen.Fx; \tag{11}$$

$$prs.Fx = \sum_{i=1}^{k} prs.F_i; \tag{12}$$

$$srv.Fx = crr.Fx + prs.Fx. \tag{13}$$

V. A Causal Structure of a Device with a Limited Capacity

A more general and real presentation of a service network portion '*x*' is in Fig. 7. It contains a virtual device *srv.x* with limited capacity. This may cause requests rejection due to lack of service place (call attempts blocking).

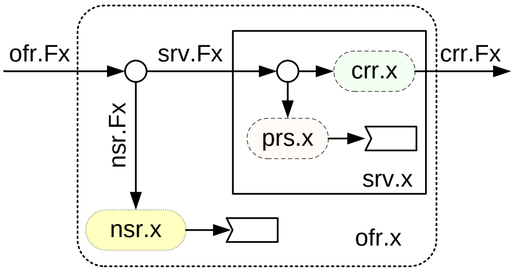

Figure 7. A causal structure of a network portion with device *x* with a limited capacity

The fictive virtual device *nsr.x* corresponds to the not served traffic, due to blocking or other reasons that would be served, with the same time duration so:

$$nsr.Tx = srv.Tx, \tag{14}$$

so the not served traffic intensity, following the theorem of Little, is:

$$nsr.Yx = nsr.Fx\ srv.Tx. \tag{15}$$

Therefore, the equivalent traffic offered [11] to device *x* is:

$$ofr.Yx = nsr.Yx + srv.Yx. \tag{16}$$

The probability of not service (blocking) *nsr.Px* may be predicted using a blocking formula. e.g., the B-formula of Erlang.

The offered traffic concept leads to necessity of definition of a conceptual device called *ofr.x* with its parameters P, F, T, Y, and V (the capacity of the offered traffic device is not considered usually). Hence, in the Fig. 7 network portion *x* is presented by 5 generalized conceptual devices: *prs.x*, *crr.x*, *srv.x*, *nsr.x* and *ofr.x* and 2 terminators, 2 transitions and 8 directors.

Obviously (see Fig. 7, (14)):

$$\begin{aligned}
&srv.Px = \left(1 - nsr.Px\right);\\
&crr.Px = \left(1 - prs.Px\right);\\
&ofr.Fx = srv.Fx + nsr.Fx;\\
&srv.Fx = ofr.Fx\ srv.Px;\\
&srv.Fx = ofr.Fx\ srv.Px = prs.Fx + crr.Fx;\\
&srv.Tx = prs.Px\ prs.Tx + crr.Px\ crr.Tx;\\
&ofr.Tx = srv.Px\ srv.Tx + nsr.Px\ nsr.Tx = srv.Tx.
\end{aligned} \tag{17}$$

The served traffic *srv.Yx* in the network portion *x* is expressed in (10).

VI. Causal Quality Indicators Formulation

The proposed conceptual presentation and naming system enable suitable definitions and simple expressions of traffic, flow and time causal quality indicators. The indicators are ratios between parameters of an inside and a comprising concepttual device, illustrated on the example of the virtual device *x* shown in Fig. 7. The abbreviation of the indicator name is considered as a device *x* parameter.

A. Traffic Quality Indicators

Traffic Quality Indicator 1: Carried / Served Traffic Intensity (Y) Ratio (CSYR) (see (10), (13) and (17)):

$$CSYRx = \frac{crr.Yx}{srv.Yx} = \frac{crr.Px\ crr.Tx}{srv.Tx}. \tag{18}$$

Traffic Quality Indicator 2: Carried / Offered Traffic Intensity (Y) Ratio (COYR) (see (11) and (16)):

$$COYRx = \frac{crr.Yx}{ofr.Yx} = \frac{srv.Px\ crr.Px\ crr.Tx}{srv.Tx}. \tag{19}$$

Traffic Quality Indicator 3: Served / Offered Traffic Intensity (Y) Ratio (SOYR) (see (10) and (17)):

$$SOYRx = \frac{srv.Yx}{ofr.Yx} = \frac{ofr.Fx\ srv.Px\ srv.Tx}{ofr.Fx\ srv.Tx} = srv.Px. \tag{20}$$

B. Flow Quality Indicators

Flow Quality Indicator 1: Carried / Served Flow Intensity (F) Ratio (CSFR) (see Fig. 7 and (16)):

$$CSFTx = \frac{crr.Fx}{srv.Fx} = \frac{ofr.Fx\ srv.Px\ crr.Px}{ofr.Fx\ srv.Px} = crr.Px. \quad (21)$$

Flow Quality Indicator 2: Carried / Offered Flow Intensity (F) Ratio (COFR) (see Fig. 7):

$$COFRx = \frac{crr.Fx}{ofr.Fx} = srv.Px\ crr.Px. \quad (22)$$

Flow Quality Indicator 3: Served / Offered Flow Intensity (F) Ratio (SOFR) (see Fig. 7):

$$SOFRx = \frac{srv.Fx}{ofr.Fx} = \frac{ofr.Fx\ srv.Px}{ofr.Fx} = srv.Px. \quad (23)$$

C. Time Quality Indicators

Time Quality Indicator 1: Carried / Served Partial Times (T) Ratio (CSTR) of device x (see (11) and Fig. 7):

$$CSTRx = \frac{prt.crr.Tx}{prt.srv.Tx} = \frac{crr.Px\ crr.Tx}{srv.Tx}. \quad (24)$$

Time Quality Indicator2: Carried / Offered Partial Times (T) Ratio (COTR) (see Fig. 7, (14), (17)):

$$COTRx = \frac{prt.crr.Tx}{ofr.Tx} = \frac{srv.Px\ crr.Px\ crr.Tx}{srv.Tx}. \quad (25)$$

Traffic Time Quality Indicator 3: Served / Offered Partial Times (T) Ratio (SOTR) of device x (see (13) and (15)):

$$SOTRx = \frac{prt.srv.Tx}{prt.ofr.Tx} = \frac{srv.Px\ srv.Tx}{srv.Tx} = srv.Px. \quad (26)$$

Proposition 2: The values of traffic and time causal quality indicators coincide.

Proof: Compare the equations: (18) and (24); (19) and (25); (20) and (26). In general, if a numerator and denominator of a causal time indicator are multiplied by the flow intensity of the comprising conceptual device we receive the respective causal traffic indicator.

Proposition 3: The values of the Served / Offered causal traffic, time and flow indicators coincide.

Proof: Compare the equations: (20) and (23); and (26). In general, the reason for this is in the definition of equivalent offered traffic [11], leading to the equation (15).

As a result, only 5 causal indicators are different and important: the 3 causal traffic quality and 2 flow quality indicators (CSFR and COFR).

VII. Conceptual Model of a Queueing System

Queuing systems have been mathematically investigated from many years, since the time of Erlang, but their graphical causal conceptual models are not considered in details.

A. Concepts and Assumptions

We propose the following causal conceptual model of a network portion, consisting of queuing system named '*qs*' (Fig. 8).

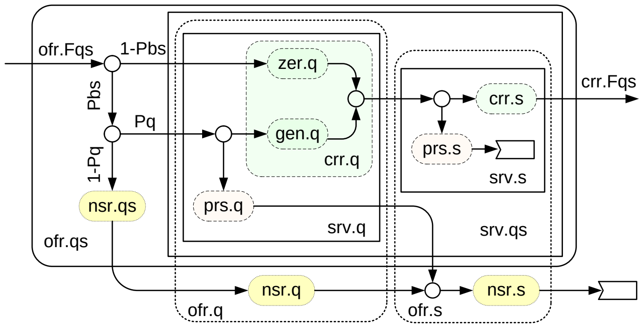

Figure 8. A causal structure of a network portion with queuing system *qs* with a limited capacity of the buffer and the server.

The buffer and the server are with limited capacities. The requests to the queuing system with flow intensity *ofr.Fqs* try to enter the server (*s*). If there is a free place in the server, the requests pass through the buffer without waiting. If in the server there is no free place for service, the requests may wait for service in the buffer device. So, the buffer service may be with or without queuing. In Fig. 8 the buffer is notated as "queuing device" (*q*).

We consider the probability 'blocked server' (*Pbs*) i.e. the server to be full and the requests' service to be blocked. The blocked server probability is different from the 'not served in the server' probability because some of the blocked requests, due to the busy server, may be served after waiting in the buffer (the name '*nsr.Ps'* is not shown in Fig. 8).

If the server is not full (the probability of blocked server is less than one), with a probability of $(1 - Pbs)$ the requests pass through the buffer without queuing, in the zero queuing (*zer.q*) device.

With a probability of *Pbs* (blocked server) the server is busy and the requests try to enter the queue in the buffer, with a probability of *Pq*. If they enter the queue, the waiting may be successful in the genuine queuing (*gen.g*) device, or unsuccessful in the parasitic queuing (*prs.q*) causal device.

Assumption 1: All considered processes are in a stationary state. The values of all parameters are random and we consider their means (or mathematical expectations).

Assumption 2: In case of zero queuing, there is no parasitic service in the buffer, due to the little service duration.

The causal device 'carried queuing' (*crr.q*) comprises devices 'zero queuing' (*zer.q*) and 'genuine queuing' (*gen.g*).

With a probability of (1 – *Pq*) the buffer is busy and the requests are not served in the queuing system – they enter the fictive device 'not served in the queuing system' (*nsr.qs*).

After carried queuing, the requests enter the server device (*srv.s*) with a flow intensity of *srv.Fs* (the name is not shown inFig. 8) and they are served: successfully in the carried service (*crr.s*) device or unsuccessfully in the parasitic service (*prs.s*) device. Obviously:

$$srv.Fs = crr.Fq. \tag{27}$$

The flow intensity of the not served in the queuing system requests (*nsr.Fqs*) is common for the queuing device (*nsr.q*):

$$nsr.Fqs = nsr.Fq. \tag{28}$$

Assumption 3: Following the definition of equivalent traffic offered [11], the fictive service times in the 'not serve' devices are:

$$\begin{aligned} &nsr.Tqs = srv.Tqs;\\ &nsr.Tq = srv.Tq;\\ &nsr.Ts = srv.Ts. \end{aligned} \tag{29}$$

VIII. Parameters of the Queuing Systems' Causal Devices

In Fig. 8 there are three comprising devices: *qs*, *q* and *s*. For each of them we'll find the parameters used in the causal traffic indicators described in Section VI. The stationary state of the queuing system is considered.

A. Parameters of the Buffer (Queuing Device)

We will use the intensity of the offered flow to the queuing system *ofr.Fqs* instead of the intensity of the offered flow to the queuing device (*ofr.Fq*) because they coincide (see Fig. 8):

$$ofr.Fq = ofr.Fqs. \tag{30}$$

The intensity of the carried flow *crr.Fq* is a sum of intensities of the incoming flows to the devices 'zero queuing' and 'genuine queuing'. They depend on the probabilities of the blocking service *Pbs* in server *s* and also on the genuine service (*gen.Pq*) in the queuing device:

$$\begin{aligned} &crr.Fq = zer.Fq + gen.Fq;\\ &zer.Fq = ofr.Fqs(1 - Pbs);\\ &gen.Fq = ofr.Fqs\ Pbs\ Pq\ gen.Pq;\\ &crr.Fq = ofr.Fqs\left[(1 - Pbs) + Pbs\ Pq\ gen.Pq\right]. \end{aligned} \tag{31}$$

The intensity of carried traffic in the queuing device (*crr.Ys*) following from the Theorem of Little and from Fig.8 is:

$$crr.Yq = crr.Fq\ crrTq = zer.Yq + gen.Yq. \tag{32}$$

From (31) and (32) we obtain:

$$\begin{aligned} &crr.Yq = zer.Fq\ zer.Tq + gen.Fq\ gen.Tg =\\ &= ofr.Fgs\left[(1 - Pbs)\ zer.Tq + Pbs\ Pq\ gen.Pq\ gen.Tg\right]. \end{aligned} \tag{33}$$

The mean service time of the carried traffic (*crr.Tq*) in the queuing device, from (32), (31) and (33) is:

$$\begin{aligned} &crr.Tq = \frac{crr.Yq}{crr.Fq} =\\ &= \frac{(1 - Pbs)\ zer.Tq + Pbs\ Pq\ gen.Pq\ gen.Tg}{(1 - Pbs) + Pbs\ Pq\ gen.Pq}. \end{aligned} \tag{34}$$

The intensity of the served flow *srv.Fq* is a sum of the intensities of the incoming flows to all causal sub-devices of the queuing device, including also the 'parasitic queuing' device (*prs.q*) and from Fig. 8 and (31) it is shown in the dependencies with number (35).

$$\begin{aligned} &srv.Fq = prs.Fq + crr.Fq;\\ &prs.Fq = ofr.Fqs\ Pbs\ Pq\ prs.Pq;\\ &crr.Fq = ofr.Fqs\left[(1 - Pbs) + Pbs\ Pq\ gen.Pq\right];\\ &srv.Fq = ofr.Fqs\left[(1 - Pbs) + Pbs\ Pq\right]. \end{aligned} \tag{35}$$

The intensity of the served traffic *srv.Yq* is a sum of the traffic intensities in all causal sub-devices of the queuing device, the 'parasitic queuing' device (*prs.q*) included and from Fig. 8 and (31):

$$\begin{aligned} &srv.Yq = srv.Fq\ srv.Tq =\\ &= zer.Fq\ zer.Tq + gen.Fq\ gen.Tg + prs.Fq\ prs.Tq;\\ &srv.Yq = ofr.Fgs\left[(1 - Pbs)\ zer.Tq +\right.\\ &\left.+ Pbs\ Pq\left(gen.Pq\ gen.Tg + prs.Pq\ prs.Tq\right)\right]. \end{aligned} \tag{36}$$

The mean service time of the served traffic (*srv.Tq*) in the queuing device, from (36) and (35) is:

$$\begin{aligned} &srv.Tq = srv.Yq / srv.Fq =\\ &= \left[(1 - Pbs)\ zer.Tq + Pbs\ Pq\left(gen.Pq\ gen.Tg +\right.\right.\\ &\left.\left.+ prs.Pq\ prs.Tq\right)\right] / \left[(1 - Pbs) + Pbs\ Pq\right]. \end{aligned} \tag{37}$$

The offered to the queuing device traffic (*ofr.Yq*) following the definition of Equivalent Offered Traffic, is a sum of served and not served (*nsr.Yq*) traffic [10]:

$$ofr.Yq = srv.Yq + nsr.Yq. \tag{38}$$

From Fig. 8 and Assumption 2 (28), not served flow of requests (*nsr.Fq*) in the queuing device is:

$$nsr.Fq = nsr.Fqs = ofr.Fqs\ Pbs\ (1 - Pq). \tag{39}$$

From (39), (29) and (37) we may express the *nsr.Yq*:

$$nsr.Yq = nsr.Fq\ srv.Tq = ofr.Fqs\ Pbs\ (1 - Pq)\ srv.Tq. \tag{40}$$

From (38), (39) and (40):

$$\begin{aligned} &ofr.Yq = srv.Yq + nsr.Yq =\\ &= ofr.Fqs\{\left[(1 - Pbs)\ zer.Tq +\right.\\ &\left.+ Pbs\ Pq\left(gen.Pq\ gen.Tg + prs.Pq\ prs.Tq\right)\right] +\\ &+ Pbs\ (1 - Pq) + \left[(1 - Pbs)\ zer.Tq + Pbs\ Pq\ (gen.Pq\ \cdot\right.\\ &\left.\cdot (gen.Pq\ gen.Tg + prs.Pq\ prs.Tq)\right] /\\ &/ \left[(1 - Pbs) + Pbs\ Pq\right]\}. \end{aligned} \tag{41}$$

B. Parameters of the Server

From Fig. 8 follows that in the intensity of the offered flow to the server (*ofr.Fs*) coincides with *crr.Fq* and *srv.Fs*:

$$crr.Fq = ofr.Fs = srv.Fs. \tag{42}$$

From (42) and (31):

$$srv.Fs = ofr.Fqs\left[\left(1-Pbs\right)+Pbs\;Pq\;gen.Pq\right]. \qquad (43)$$

The intensity of the carried flow $crr.Fs$ depends on the probabilities of the parasitic service $prs.Ps$ in the server s (see Fig. 8 and (43)). It is shown in formula (44).

$$\begin{aligned} &crr.Fs = srv.Fs\;crr.Ps;\\ &crr.Fs = ofr.Fqs\left[\left(1-Pbs\right)+Pbs\;Pq\;gen.Pq\right]crr.Ps. \end{aligned} \qquad (44)$$

The mean service time of the carried traffic $crr.Ys$ in the server device, from Fig. 8 simply is the service time of the carried service device $crr.s$ – $crr.Ts$.

The intensity of carried traffic in the server device ($crr.Ys$), following the Theorem of Little, Fig. 8 and (44) is:

$$\begin{aligned} &crr.Ys = crr.Fs\;crr.Ts =\\ &= ofr.Fqs\left[\left(1-Pbs\right)+Pbs\;Pq\;gen.Pq\right]crr.Ps\;crr.Ts. \end{aligned} \qquad (45)$$

The intensity of the served traffic $srv.Ys$ is a sum of the traffic intensities of the 'carried service' and the 'parasitic service' ($prs.q$) devices and from Fig. 8 and (44):

$$\begin{aligned} &srv.Ys = srv.Fs\;srv.Ts;\\ &srv.Ts = crr.Fs\;crr.Ts + prs.Fs\;\;prs.Ts;\\ &srv.Ys = srv.Fs\left(crr.Ps\;crr.Ts + prs.Fs\;\;prs.Ts\right). \end{aligned} \qquad (46)$$

Finally:

$$\begin{aligned} &srv.Ys = ofr.Fqs\left[\left(1-Pbs\right)+Pbs\;Pq\;gen.Pq\right]\cdot\\ &\cdot\left(crr.Ps\;crr.Ts + prs.Ps\;\;prs.Ts\right). \end{aligned} \qquad (47)$$

The mean service time of the served traffic $srv.Ts$ in the service device, from (43) and (47) is:

$$srv.Ts = \frac{srv.Ys}{srv.Fs} = crr.Ps\;crr.Ts + prs.Ps\;\;prs.Ts. \qquad (48)$$

From Fig. 8 and Assumption 2 (28), the not served flow of requests $nsr.Fs$ for the service device is the sum of both quantities $nsr.Fqs$ and $prs.Fq$:

$$\begin{aligned} &nsr.Fs = nsr.Fqs + prs.Fq;\\ &nsr.Fs = ofr.Fqs\left[Pbs\left(1-Pq\right)+Pbs\;Pq\;\;prs.Pq\right];\\ &nsr.Fs = ofr.Fqs\;Pbs\left(1-Pq\;gen.Pq\right). \end{aligned} \qquad (49)$$

From (29), (48) and (49) we may express the $nsr.Ys$:

$$\begin{aligned} &nsr.Ys = nsr.Fs\;srv.Ts =\\ &= ofr.Fqs\;Pbs\left(1-Pq\;gen.Pq\right)\cdot\\ &\cdot\left(crr.Ps\;crr.Ts + prs.Ps\;\;prs.Ts\right). \end{aligned} \qquad (50)$$

The offered to the service device traffic $ofr.Ys$ from (46), (47), (48), (49) and (50) is shown in 47.

The obvious result (51) is a confirmation of the consistency of the conceptual and analytical presentations.

$$\begin{aligned} &ofr.Ys = srv.Ys + nsr.Ys =\\ &= ofr.Fqs\left[\left(1-Pbs\right)+Pbs\;Pq\;gen.Pq\right]\cdot\\ &\cdot\left(crr.Ps\;crr.Ts + prs.Ps\;\;prs.Ts\right)+\\ &+ pfr.Fqs\;Pbs\left(1-Pq\;gen.Pq\right)\cdot\\ &\cdot\left(crr.Ps\;crr.Ts + prs.Ps\;\;prs.Ts\right);\\ &ofr.Ys = ofr.Fqs\;srv.Ts. \end{aligned} \qquad (51)$$

C. Parameters of the Queing system

The carried requests flow $crr.Fqs$ in the queuing system follows from Fig. 8, (27) and (44):

$$\begin{aligned} &crr.Fqs = crr.Fs = crr.Fq\;crr.Ps =\\ &= ofr.Fqs\left[\left(1-Pbs\right)+Pbs\;Pq\;gen.Pq\right]crr.Ps. \end{aligned} \qquad (52)$$

The successful (carried) service time $crr.Tqs$ in the queuing system is a sum (see Fig. 8) of the consecutive successful service times in the buffer $crr.Tq$ (34) and in the server $crr.Ts$:

$$crr.Tqs = crr.Tq + crr.Ts. \qquad (53)$$

The served traffic $srv.Yqs$ in the queuing system follows from Fig. 8, (36), (47) and the Theorem of Little:

$$srv.Yqs = srv.Fqs\;srv.Tqs = srv.Yq + srv.Ys. \qquad (54)$$

Note that $crr.Yqs \neq crr.Yqs + crr.Ys$ because a part of the served requests in the buffer are terminated after parasitic service in $prs.s$ device. Therefore the carried traffic in the queuing system ($crr.Yqs$) is:

$$crr.Yqs = crr.Fqs\;crr.Tqs, \qquad (55)$$

where $crr.Fqs$ is expressed in (52), and $crr.Tqs$ – in (53).

The served traffic in the queuing system $srv.Yqs$ is a sum of the served traffic in the buffer $srv.Yq$ (36) and in the server $srv.Ys$ (47) (cf. Fiq. 8):

$$srv.Yqs = srv.Yq + srv.Ys. \qquad (56)$$

The offered traffic to the queuing system ($ofr.Yqs$) is a sum of the served traffic in the system ($srv.Yqs$) (56) and the not served ($nsr.Yqs$) (cf. Fig. 8):

$$ofr.Yqs = srv.Yqs + nsr.Yqs. \qquad (57)$$

The not served $nsr.Yqs$ traffic intensity, from the equivalent offered traffic definition, Fig. 8 and (54) is:

$$nsr.Yqs = nsr.Fqs\;srv.Tqs = nsv.Fqs\frac{srv.Yqs}{srv.Fqs}. \qquad (58)$$

The not served flow intensity $nsr.Fqs$, from (28) and (39) is:

$$nsr.Fqs = ofr.Fqs\;Pbs\left(1-Pq\right). \qquad (59)$$

The served flow intensity $srv.Fqs$, from Fig. 8 is:

$$srv.Fqs = ofr.Fqs\left[\left(1-Pbs\right)+Pbs\;Pq\right]. \qquad (60)$$

From (58), (59) and (60) we obtain:

$$nsr.Yqs = \frac{Pbs\left(1-Pq\right)}{1-Pbs\left(1-Pq\right)} srv.Yqs. \qquad (61)$$

Finally, the offered traffic in the queuing system $ofr.Yqs$, from (57), (58) and(61) is:

$$ofr.Yqs = \frac{srv.Yqs}{1-Pbs\left(1-Pq\right)} = \frac{srv.Yq+srv.Ys}{1-Pbs\left(1-Pq\right)}. \qquad (62)$$

IX. Some Causal Quality Indicators of a Queing System

The results received in Section X allow a simple calculation of the causal traffic quality indicators, defined in Section VI. In this section, we shall illustrate this with some new results.

A. Traffic Quality Indicators of a Queuing System

The traffic Quality Indicator Served / Offered Traffic Intensity Ratio (SOYR) (20), using (60) is:

$$SOYRqs = \frac{srv.Yqs}{ofr.Yqs} = 1-Pbs\left(1-Pq\right). \qquad (63)$$

B. Flow Quality Indicatorsof a Queuing System

Flow Quality Indicator 1: Carried / Served Flow Ratio (CSFR) (21).

From Fig. 8, *crr.Fs* = *crr.Fqs* and *srv.Fqs* = *srv.Fq* because *srv.Fs* is a part and a continuation of *srv.Fq*:

$$CSFRqs = \frac{crr.Fqs}{srv.Fqs} = \frac{crr.Fs}{srv.Fq}. \qquad (64)$$

From (30), (31), (32) and (33) follows:

$$CSFRqs = \frac{\left[\left(1-Pbs\right)+Pbs\,Pq\left(1-prs.Pq\right)\right]crr.Ps}{\left(1-Pbs\right)+Pbs\,Pq}. \qquad (65)$$

From the other side, from (31) and (33):

$$CSFRq = \frac{crr.Fq}{srv.Fq} = \frac{\left[\left(1-Pbs\right)+Pbs\,Pq\left(1-prs.Pq\right)\right]}{\left(1-Pbs\right)+Pbs\,Pq}. \qquad (66)$$

And from (27) and (32):

$$CSFRs = \frac{crr.Fs}{srv.Fs} = \frac{srv.Fs\;crr.Ps}{srv.Fs} = crr.Ps. \qquad (67)$$

Therefore, from (34), (35) and (36):

$$CSFRqs = CSFRq\;CSFRs. \qquad (68)$$

X. Conclusion

The proposed approach to traffic quality indicators aggregation leads to new conceptual and graphical tools and aggregation functions derived. The QS is considered as a portion of an informational service network. The causal service structure of QS is extended with non-served traffic devices and consists of six causal virtual devices. The virtual devices' notation system is extended. The concept of partial service time is proposed. This allows easier and clearer graphical and analytical model presentations and a definition of QoS indicators as aggregations of QoS of the nested components. The results obtained are useful for QoS measurement, prediction and management in parts of real informational service networks, including telecommunication networks for multimedia traffic [10], [11], [13]. It allows overall quality estimation in every service composition, considered as a workflow.

In our future work, traffic quality indicators' aggregations of other composition patterns will be researched.

References

[1] M. Varela, P. Zwickl, P. Reichl, M. Xie, H. Schulzrinne,"From service level agreements (SLA) to experience level agreements (ELA): The challenges of selling QoE to the user" IEEE ICCW, 2015, pp. 1741-1746, ISBN: 978-1-4673-6305-1, 978-1-4673-6304-4.

[2] ITU-T Recommendation ITU-T P.10/G.100 (11/2017). Vocabulary for performance, quality of service and quality of experience.

[3] H. Zheng, J. Yang, and W. Zhao, "General probability distribution-based QoS analysis for web service composition,"International Symposium on Web Intelligent Systems & Services. D.K.W. Chiu et al, Eds., WISE 2010 Workshops, LNCS 6724, pp. 98–111, 2011, Springer-Verlag.

[4] M. Gatnau, D., Catania, A. F. Cattoni, J. S. Ashta, and P. Mogensen, "A multi-QoS aggregation mechanism for improved fairness in WLAN,"In Vehicular Technology Conference (VTC Fall), 2013 IEEE 78th, 2013, pp. 1-5, https://doi.org/10.1109/VTCFall, 6692041.

[5] S. Poryazov, E. Saranova,"User-oriented, overall traffic and time efficiency indicators in telecommunications," TELFOR 2016 International IEEE Conference #39555, November 22-23, Belgrade, Serbia , IEEE Catalog Number: CFP1698P-CDR, IEEE, ISBN:978-1-5090-4086-5/16, DOI: 10.1109/TELFOR.2016.7818729, 2016.

[6] S. Poryazov, E. Saranova, I.Ganchev,"Scalable traffic quality and system efficiency indicators towards overall telecommunication system's QoE management," In: I. Ganchev, R.van der Mei, H.van den Berg, Eds., Autonomous Control for a Reliable Internet of Services. State-of-the-Art Survey, LN in Computer Science, vol. 10768, Springer, Cham, DOI: https://doi.org/10.1007/978-3-319-90415-3_4, 2018, pp 81-103.

[7] S. Poryazov, E. Saranova, I. Ganchev,"Conceptual and analytical models for predicting the quality of service of overall telecommunication systems," In: Ganchev I., van der Mei R., van den Berg H.,Eds., Autonomous Control for a Reliable Internet of Services. State-of-the-Art Survey, Lecture Notes in Computer Science, vol. 10768, Springer, Cham, DOI https://doi.org/10.1007/978-3-319-90415-3_7, 2018.

[8] Cisco Unified Communications Manager Call Detail Records Administration Guide, Release 8.5(1). Text Part Number: OL-22521-01, pp. 172. Cisco Systems 2010.

[9] J. D. C. Little,"A proof of the queueing formula $L = \lambda W$," Operations Research, 9, 1961, pp.383-387.

[10] ITU E.501. ITU-T Recommendation E.501: Estimation of Traffic Offered in the Network, 26th of May, 1997.

[11] I. Nedyalkov, A. Stefanov and P. Apostolov, "Modeling of the convergence time of an IP-based network with different traffic loads," IEEE EUROCON 2019 – 18th International Conference on Smart Technologies, Novi Sad, Serbia, 2019, pp. 1-6, DOI:10.1109/EUROCON.2019.8861735.

[12] I. Nedyalkov, A. Stefanov, G. Georgiev, "Studying and Characterization of the Data Flows in an IP-Based Network", International Journal on Information Technologies and Security, vol. 1, No. 1, 2019, pp. 3-12, WOS000460045600001, ISSN: 1313-8251.

[13] A. A. Barakabitze, N. Barman, A. Ahmad, S. Zadtootaghaj, L. Sun, M. G. Martini et al., QoE Management of Multimedia Streaming Services in Future Networks: A Tutorial and Survey. DOI: 10.1109/COMST.2019.2958784 IEEE, 2019