MLADEN BEGOVIĆ, D. Sc.
INA - Naftaplin
Zagreb, Šubićeva
SLAVKO ŠARIĆ, D. Sc.
MIROSLAV MIKULA, D. Sc.
University of Zagreb
Faculty of Transport and Traffic Engineering
Zagreb, Vukelićeva 4

Znanost u prometu - Science in Traffic
Prethodno priopćenje - Preliminary communication
U.D.C. 656:681.3
Primljeno - Accepted: 27.02.1996.
Prihvaćeno - Approved: 28.05.1996.

MODELLING OF TRAFFIC PROCESSES IN TELECOMMUNICATION SYSTEMS

MODELIRANJE PROMETNIH PROCESA U TELEKOMUNIKACIJKIM SUSTAVIMA

Modeliranje prometnih procesa u telekomunikacijskim sustavima u fazi tehnoloških promjena je neophodno u cilju planiranja proširenja kapaciteta sustava te nadgledavanja postojećih prometnih tijekova. U ovom radu obrađen je prometni problem u telekomunikacijskim sustavima u periodu zamjene tehnologije, pri čemu se promet analizira kao slučajni proces, a prometni procesi simuliraju pomoću podvorbenih modela.

1. INTRODUCTION

The requirements for modelling traffic processes originate from the following two needs:

(a) establishing the laws of traffic being accommodated by the existing communication systems, and

(b) designing of communication systems.

In both cases, care should be taken that the properties of traffic processes can be assigned into two groups:

1. time characteristic, and

2. quantity characteristic

In the period of introduction of new technologies in the existing communication system, the old and new technology are in simultaneous coexistence. This means we have to count upon both types of traffic: analog (Erl) and digital (bit). This viewpoint is necessary if we regard traffic as input data for the process of designing new communication systems. Viewed from the economic aspect, more useful is the value presenting traffic cumulatively, namely charged holding time usually expressed in minutes and called the charged minute.

What is to be measured?

The answer to this question requires having in mind that electromechanical commutation disappears rapidly, being increasingly substituted for processor-operated ones. Hence, we are confronted with the processor-operated commutation. The process of recording and processing of traffic information in a processor-operated commutation has the potentials to meet all traffic requirements.

Economists are primarily interested in getting the indices of economic efficiency, hence focusing upon the total time (minutes), because the time of engagement is charged, or more accurately: duration of connection. Engineers, in turn, are primarily interested in quantity of information (bit) or the percentage of total time in which the transmission channels are busy (Erl). Economists find the objective of traffic data processing in monitoring cost effectiveness, while engineers have as their goal the timely expansion of capacities for purposes of keeping given service quality (grade of service).

The potential of modern processor-operated commutations in meeting these requirements can be observed on the example of access to the method of recording and processing of traffic and tariff functions of AXE commutation. OMS (Operation and Maintenance Subsystem is designed for traffic functions, and CHS (Charging Subsystem) for tariff functions. Both subsystems comprise functional blocks that may be used for getting traffic data. In OMS this refers to CHAS, TRAR and TROB, while in CHS this refers to MP and TT. Standard software does not offer all of the potentials being in terms of traffic requirements offered by hardware, so that mostly only charging requirements are met, i.e. charged minutes are recorded. Other potentials are of the nature to meet all engineers' requirements. The subject requirements will be highlighted in the process of analysis of the traffic problem in the communication system.

2. TRAFFIC PROBLEM IN THE COMMUNICATION SYSTEM

Traffic problem in the telecommunication system refers to the establishing of traffic at each and every point of the system, at any one moment (present and past), in the process of designing and operation. In the process of designing, traffic is dealt with solely by means of models and modelling, while the analytical and simulation methods are applied as solution finding ones. In the process of operation traffic is defined by measurement. This still involves models for purposes of timely projections of needs for the expansion of individual segments of the entire communication system. Namely, the failure to complete this, can result in cumulative growth of losses possibly affecting larger segments of the communication system.

In the process of resolving a traffic problem it is necessary to take account of the fact that the character of traffic accommodated by analog and digital communication systems is essentially different, affecting in turn the selection of measuring methods. Traffic in the analog communication system refers to the length of total holding time measured by the length of measurement period. Call is the usual term for the length of holding time. Between calls we have pauses. In analog communication systems the process of circuit switching is applied, in which one communication channel is assigned to a sole information source at the time of call.

The intensity of traffic in a digital environment refers to the number of users served during busy period. The process of commutation refers to the packet (message) switching. In the very definition of traffic in the digital network we have to apply the terms from the queuing theory, best illustrating the applicable models. queuing systems involve a queue, service facility and users or jobs, and are described by means of input process, service process and rules of joining and leaving the queue discipline. Users is the term defining the job to be performed by the service facility and has the physical features of realistic conditions to be modelled by way of the queuing system. Service facility is part of the queuing system in which the provision is effectuated. Traffic refers to the movement of users in the system, so that all features referring to the number of users are at the same time traffic features. In the digital network users are represented by data volumes in bits. In the case of adequate description of a service facility, we will be able to relate the theory of queuing systems to the theory of information in Shannon's terms. The relating term is the capacity of server. In the application of the theory of queuing systems to the communication system, namely, the volumetric aspect of service provision required by individual users is usually called request or job. The units for job measuring are various and depend upon the character of the server and users. If the server is a communication channel and users are messages of packets, the unit for job done is bit or byte. In the majority of cases we may suppose that the users population is homogeneous, meaning that the jobs of individual users are selected by means of a single common distribution. This distribution is called service time distribution. Obviously there are cases in which this is not met and then the users are grouped per types of jobs and the same distribution. Definition: The capacity or the rate of operation of a service facility, C, is the service rate of service facility. The unit for the capacity of service facility depends upon the aspect of the service provided, determined by the physical features of the users and the service facility. We are interested in communication systems so that the units in question are (request/second) and (bit/second). In the first case, we deal with processor-operated segments of the queuing system, while in the second we deal with the communication channel. In the latter, the term of the capacity of the service facility can be matched with the term of the capacity channel as the highest possible transmission rate with respect to the distribution of input messages or packets.

$$C = \max_{p(x_i)} I(X;Y)$$

If the user's request comprises the value S/service units/user/ and the server has the capacity C/service units/second/ then the ratio of S/C /second/user/ is called the service time provision. Its average value $\overline{S}/C$ /second/user/ is referred to as the average service time, while the reciprocal value

$$\mu = C/\overline{S}/\text{users/second}/$$

is the service rate. If C is constant, there is no need to discriminate the term of request for service and the term of service time, and it can be put that its value is equal to C=1. Now, the request for service is measured in seconds, i.e. by time. However, there are numerous cases when C is not constant, but changes with the number of users in the system. The simplest example refers to the multi-server queuing system with a common queue. The capacity of the service facility here depends upon the state of the system, i.e. upon the number of users in the system:

$$C(q) = \min(q,c) \cdot C$$

where q is the number of users in the system and c is the number of servers.

3. TRAFFIC AS RANDOM PROCESS

We have seen that in the existing communication system there are three traffic values (units) for describing the traffic aspected behaviour of the system: total length of holding time measured by total measured time in (Erl), total length of call in the process of measurement expressed in charged minutes and quantities of information in (bit). All these traffic values change in time following the rhythm of human activities in a day, week, month and year. Traffic values are random variables in the spirit of the theory of probability, for being affected by a variety of factors, while only their cumulative result is observed. The behaviour of the random variable in time can be modelled by means of a random process. When we model the behaviour of traffic values in time, we speak of the modelling of time characteristics of traffic processes.

The process of modelling traffic by way of random processes in based upon the results of measurement of traffic flow in time. Measurement is namely necessary for purposes of maintenance of the communication process up to the in advance specified level of quality. This quality must be ensured at any one moment and at any one point of the communication system. The results of measurement of traffic within the time periods specified in advance, must thus be regarded as the realization of corresponding random processes. (See, for instance: Karlin, Taylor: "First Course in Stochastic Processes", Academic Press, 2nd ed., 1975).

Random or stochastic processes are the most suitable models for description of the behaviour of traffic processes in time, whether we deal with utilization or designing. In the process of operation of the existing communication system it shall be necessary to structure the model of traffic process for purposes of planning the extension of capacities and recording the experience acquired in traffic monitoring, being unavoidable in the process of designing of new systems. The model is worth to the degree of its reflecting the realities. From this standpoint, the practice of modelling traffic processes by way of random processes emerges as the most accurate and advantageous approach to designing of the communication systems.

A stochastic process is completely determined by means of the set of all realizations. In the process of traffic modelling by means of stochastic processes we must therefore first determine all realizations. The process of measurement is hence defined as comprising an as high as possible number of basic measurement periods in which traffic values are measured. The set of measurement results of one period refers to one realization of traffic process. The length of the basic measurement period is determined by the physical character of traffic. When dealing with the communication system it appears obvious that the day (24 hours) should be taken as the basic measurement period, being the basic unit of human activity in time. In the light of the fact the measurement is affected by the unavoidable factor of economy, technical requirements should be weighed against economic standards. In the subject case this means that the daily measurement is accurate enough if limited to the busy period of the day, usually between 7.00 a.m. and 10.00 p.m. Obviously there are such communication systems for which it shall be necessary to extend the measurement period to 24 hours, or such for which this period can be reduced to only a few hours or less (for instance: LAN).

In all these considerations we must keep in mind the public communications system using as its transmission background the public PT network. Specific users, having available their own networks are in the position to determine measurement periods according to the physical character of the traffic source. Typical examples refer to the electric-power industry (remote surveillance and control), railway, transport companies, or gas companies, water resources management, etc.

Treatment of traffic within the scope of issues of communication in the communication system is stipulated by CCITT regulation.

In these regulations traffic problem or issue is closely related to the issue of communication system management. The term relating management of traffic is the grade of service, being quantified by among other things boundary congestion values. These parameters are the basic criteria for planning the extension of the capacities of the existing system, dictated by the needs for maintenance of the required quality of service. The former practice of stipulating the boundary values of congestion was not based upon traffic modelling by way of random processes and was therefore at discrepancy with the practice. These regulations were therefore a recurrent theme of argument until the introduction in broad use of computers, which if not available would make the approach to traffic as a random process hardly feasible in practice.

The proces of traffic modelling by way of the theory of random processes proceeds in several steps:

(a) Identifying all completed realizations

(b) Identifying the kind of random process: It is obvious that for reasons of economy as simple as possible models are to be used. Therefore, in the first place we must resolve the issue of its ergodicity and/or stationary. The practice shows that for engineering purposes traffic can be described as an ergodic and stationary random process, at least within the major part of the basic measurement period. This issue is rather complicated and will for some time to come stay not-covered by standards.

(c) Structuring the distribution for the basic measurement period.

(d) Pursuit of the analytical model of actual distribution: adequacy of the analytical model is checked against the standard procedures of statistical analysis for purposes of matching the theoretical with the actual distribution.

(e) Comparison of several measurement periods by way of their distribution: by means of the tests of statistical analysis for checking the sameness of several samples we arrive at the conclusion on the results of traffic measurement in different basic periods (not)belonging to the same probability population. These results have multi-purpose application, for enabling traffic surveying in different periods of the year and in different conditions (seasons, urban or rural environment, comparison of transit areas, traffic in the period of working hours and outside this period, etc.).

(f) Monitoring and forecasting of the parameters to serve for the description of the quality of service within plan periods: this comes as a direct result of traffic modelling by way of random process models.

Modern processor-operated commutations enable detailed measurement of the overall communication process. This means that in conjunction with traffic we can observe and examine the structure of transactions (calls). This refers us to the conclusion that traffic and call structure measurement ensures the approach to the issue of charging of a more realistic quality than the approach considering the total number of minutes charged for a specified period.

All these considerations understand a realistic approach to traffic handling. This in the first place refers to the regular monitoring or follow-up of traffic in the existing communication systems at all hierarchical levels. The former practice in the country failed to even approximately meet this. The PT Company of Zagreb has succeeded to reach farthest in the development of this approach in international traffic, where this aspect of approach was for the first time developed and in part applied in everyday practice. We are positive that the introduction of processor-operated commutations at all levels of the communication system shall considerably improve the approach to traffic and confirm the justifiability of modelling of traffic by way of the models from the theory of random processes.

4. TRAFFIC AS QUEUING PROCESS

In the first section of this paper we have dealt with the kinds of traffic in the analog and digital network. We have discussed the method of characterizing traffic in the analog network by means of a relative or absolute time period, and by means of the quantity of information in the digital one. The actual physical character of traffic in both cases refers to the quantity of information in Shannon's terms. If we examine traffic at this level, the modelling of traffic processes can make use of the models from the theory of random processes in conjunction with essential features of the communication system, i.e. its queuing aspect. Traffic as a quantity of information, namely, is accommodated by the communication system only in case the system is able to accommodate and distribute it by means of its transmission support. All parts of the communication system receive and process traffic and provide its transmission between the sources (originators) and destinations. In all points of the communication system with traffic the queuing process takes place. The only models applicable are the models from the queuing theory. Queuing models are therefore macromodels of excellent use for the illustration of traffic proceeding in individual parts of the communication system. The time characteristic component of these processes is successfully modelled by means of the models from the theory of random processes. Queuing models at each point of the communication system relate the time variability of traffic to the information transmission capacity. We have already given a summarized description of the queuing systems, so that we shall now consider only those features related to traffic:

(a) Offered traffic:

Offered traffic refers to the average number of users reaching the queuing system with their requests, as handled during the average value of service time.

$$a = \frac{T_s}{T_a} = \frac{\text{average value of service time}}{\text{average value of the time between arrivals}}$$

$$a = \lambda T_s = \text{rate of arrivals} \cdot \text{average value of service time}$$

$$a = \frac{\lambda}{\mu} = \frac{\text{rate of arrivals}}{\text{rate of service}}$$

where the following relations exist

$$\lambda = \frac{1}{T_a} \qquad \mu = \frac{1}{T_s}$$

The unit of measure of the offered traffic is Erl.

This exactly matches the definition of the traffic of one communication channel.

(b) Traffic intensity:

Traffic intensity refers to traffic per server.

If we have c parallel identical servers each operating at μ rate, then the intensity of traffic is

$$\rho = \frac{a}{c} = \frac{\lambda}{c\mu}$$

This value sometimes hardly lends to expression by way of the time between arrivals and service.

(c) Carried traffic:

Carried traffic refers to that part or portion of offered traffic that has not been lost upon arrival at the queuing system. Hence, it refers to the measure of the number of those users that have actually been served. If the system enables forming of an indefinite queue, while the users are patient, i.e. do not leave the system prior to being served, the carried traffic is equal to the offered one. If the system has an indefinite queue, the carried traffic shall differ from the offered one for those users reaching the queuing system at the moment where the queue is full. If B(a) represents the probability that the user is in the system of offered traffic yet prevented from entering and being served, the state of the system is described as the blockage, while the carried traffic is

$$a' = a\,(1 - B\,(a))$$

In a common multi-server system under steady-state conditions, the carried traffic is numerically identical with the average number of busy servers.

(d) Utilization factor:

The factor of utilization refers to the carried traffic per one server under equilibrium conditions of the system. For c parallel identical servers it is:

$$\rho' = \frac{a'}{c}$$

If the system has adequate capacity so that no one of the incoming users shall experience blockage and all enjoy the same rate of service m, the utilization factor and the intensity of traffic are numerically mutually identical values:

$$\rho' = \rho = \frac{\lambda}{c\mu}$$

(e) Load level:

Load level refers to the portion of time in which the number of users in the queuing system is higher than the specified number N:

$$P_N = P\{q > N\} = \sum_{j=N+1}^{\infty} p_j = 1 - \sum_{j=0}^{N} p_i$$

where p_j is steady-state probability of the system state.

(f) α - load limit

α - load limit, U_α, is the load the system carries with in time (%) of the total observation time:

$$P\{q \le U_d\} \ge \frac{\alpha}{100} > p\{q \le U_\alpha - 1\}$$

or:

$$\sum_{j=0}^{U_d - 1} p_i < \frac{\alpha}{100} \le \sum_{j=0}^{U_\alpha} p_i .$$

This review of traffic features of the queuing system speaks in favour of the suitability of these models for the study of traffic and other features of all those parts of the communication system in which the phenomenon of servicing the users can be established, in which process the two terms should be taken without their physical content. The theory of queuing systems can thus be applied to the sections of whole commutations, channel bundles, processor-aided operation, etc. These models possess a universal character and are being applied in all traffic systems: from maritime, road, rail, air up to the communications system dealt with in this paper. The queuing systems are indispensable for the estimate of the processor- monitored and controlled systems.

5. CONCLUSIONS

The present treatment of traffic in our communication system can be described as inadequately comprehensive and insufficiently accurate. PT companies deal with traffic mostly from the economic point of view, hence accurately recording only the overall charged minutes for all types of traffic generation sources. The intensity of traffic (Erl) is being measured only at some levels, mostly in spurts. The main cause is seen in the usual scarcity of measuring equipment. The economic indices refer to the essential element for the provision of engineering propositions and these dictate the pace and scope. However, it is our impression that the conscience is missing of how much the technical solutions and economic base stand in the cause-and-effect relationship. Every technical solution must result in the increase of cost effectiveness and profit. In other words: investment in technical equipment must always result in increased profit. We implicitly understand that here we deal with right and not wrong technical propositions. They will be increasingly better if the growth of traffic is forecast as precisely as possible, being feasible only through the improved methods of modelling, able to assume/depict in a more realistic manner the actual circumstances. The models best illustrating the actual conditions in communication systems are the models from the theory of random processes and theory of queuing systems.

A primarily economic approach to traffic has caused the inception of traffic values adequate for individual aspects of traffic sources, so that we deal with charged minutes in telephone traffic and cables and charged minutes for telegraph traffic. From the aspect of charging of traffic sources this is sufficient while from the aspect of expansion of the capacities of the communication system this is obviously insufficient. The data we observe in telegraph traffic can well illustrate the rate of change of the number of originators and scope of economic effect. For the needed timely intervention in the area of expansion of capacities of the segments of communication system, it would be necessary to at least have some knowledge of traffic in Erls in individual groups or lines. The measurement of the behaviour of sources should namely be separated from the measurement of conditions of the connection groups and communications channel. It is our opinion that all these requirements shall be met only after the introduction of processor-operated commutations at all levels of the communication systems.

SUMMARY

Modelling of traffic processes in telecommunication systems in the process of the shift of technologies is indispensable for purposes of planning and designing the extension of capacities and monitoring of the existing traffic flows. The paper deals with the traffic problem in the telecommunication system in the period of the exchange of technologies, discussing traffic as a random process and expounding the need for modelling traffic processes by way of servicing models.

BIBLIOGRAPHY

Traffic related data for communication systems can be found in a variety of publications. Those available to us mostly come from the annual publications of the United Nations. These publications have been compiled on the basis of data sent by PT organizations from all over the world. In addition to the OUN publications there exist those published by some major world companies. There is a common consensus based upon practical experience among the experts interested in these issues that the best publications dealing with traffic, including the OUN ones, are:

[1] Yearbook of Statistics year ITU, Generava
[2] The world's telephones: A statistical compilation as of year, AT & T, Indianapolis, IN,
[3] International Fernsprechstatistik, Siemens, Munich.