INTERNATIONAL TELECOMMUNICATION UNION

CCITT

E.731

THE INTERNATIONAL
TELEGRAPH AND TELEPHONE
CONSULTATIVE COMMITTEE

(10/92)

TELEPHONE NETWORK AND ISDN

**QUALITY OF SERVICE,
NETWORK MANAGEMENT AND TRAFFIC
ENGINEERING**

METHODS FOR DIMENSIONING RESOURCES OPERATING IN CIRCUIT SWITCHED MODE

Recommendation E.731

FOREWORD

The CCITT (the International Telegraph and Telephone Consultative Committee) is a permanent organ of the International Telecommunication Union (ITU). CCITT is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The Plenary Assembly of CCITT which meets every four years, establishes the topics for study and approves Recommendations prepared by its Study Groups. The approval of Recommendations by the members of CCITT between Plenary Assemblies is covered by the procedure laid down in CCITT Resolution No. 2 (Melbourne, 1988).

Recommendation E.731 was prepared by Study Group II and was approved under the Resolution No. 2 procedure on the 30th of October 1992.

____

CCITT NOTES

1) In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized private operating agency.

2) A list of abbreviations used in this Recommendation can be found in Annex A.

Recommendation E.731

METHODS FOR DIMENSIONING RESOURCES OPERATING IN CIRCUIT SWITCHED MODE

(1992)

1 Introduction

This Recommendation provides methods for dimensioning resources operating in circuit switched mode in the ISDN.

It proposes the use of traditional telephone dimensioning methods complemented with specific techniques that should be used for those ISDN issues that are widely at variance with the traditional telephone traffic models. Techniques for the following ISDN issues are provided:

- Multi-slot connection;
- Service protection methods;
- Attribute negotiation;
- Service reservation;
- Multipoint connections.

2 Evolution of network dimensioning methods

2.1 *Telephone traffic modelling*

All telephone network dimensioning procedures are based on mathematical models that approximate the statistical behaviour of telephone traffic in large populations. These models allow straightforward characterization of the traffic demand and network dimensioning by adopting simplifying assumptions concerning:

- stationarity of the traffic process during the reference period;
- call attempt inter-arrival times;
- call holding times;
- disposition of blocked or delayed call attempts;
- dependency of call attempts and holding times from network state and other call attempts.

For example, in the most commonly used Erlang loss formula, the following assumptions are used:

- stationary traffic process during the reference period;
- poissonian arrival of call attempts;
- general distribution of call holding times;
- blocked calls cleared;
- independence of call attempts and holding times on network state and other call attempts.

Assumptions are not actual descriptions of individual caller behaviour, but give rise to statistical patterns that have been experimentally observed to closely approximate the aggregate traffic flows in real telephone networks.

The history of telephone network dimensioning can be viewed as primarily the extension of a few basic traffic models to a widening range of applications, including the development of techniques to replace situations not entirely consistent with the basic models by “equivalent” situations that permit continued application of the same techniques.

The E.520-Series Recommendations provide telephone network dimensioning methods based on traditional traffic modelling.

2.2 *Impact of ISDN circuit switched mode*

Narrow-band ISDNs will evolve from the PSTN and the basic network structure for the circuit switched mode will be identical to that of the PSTN.

However, many ISDNs will offer circuit switched services with modes of operation that are at variance with underlying assumptions of traditional telephone traffic models. For example multi-slot services require different amounts of network resources for different call attempts, an aspect that is not considered in traditional telephone traffic models.

The technologies being introduced in conjunction with ISDN can result in new traffic patterns as well. For example, automatic repetition terminals coupled with fast network signalling may result in a sequence of closely spaced, correlated call attempts that cannot be represented by the independence assumptions in traditional telephone traffic models.

2.3 *Evolution of ISDN dimensioning*

ISDN is an open, permissive concept with a very large range of development paths. Many of the services and features presently identified might never materialize in quantity, and new services and features not yet described might dominate future ISDN evolution. In this situation new traffic models should not be introduced until actual experience indicates system behaviour to be sufficiently approximated.

This Recommendation provides specific dimensioning techniques that should be used for services and features that are widely at variance with the traditional telephone traffic models and are anticipated to affect resource requirements in at least part of ISDN networks. The decisions of when and where to apply these techniques in evolving ISDN structures will be a major challenge for traffic engineering specialists in individual Administrations.

3 **Multi-slot connections**

3.1 *Completely partitioned circuit groups*

Consider a circuit group serving M independent traffic components, with different arrival rates, holding times and bandwidth (equivalently circuit or bit rate) requirement. A call of the component m requires d_m circuits to be carried. The straightforward approach to achieve specified grades of service for different traffic components is to split total circuit group into M parts. The capacity of each part is dedicated to its corresponding traffic component.

Required circuit group capacity for each traffic component should be found to achieve the GOS target. Standard dimensioning methods described in the E.520-Series Recommendations can be used for calculations of required circuit group capacity. Those calculations are based on the traffic defined as $A'_m = \lambda_m \cdot h_m$, λ_m being the call arrival rate and h_m the holding time (see Recommendation E.712, § 3.5), thus without accounting for the bandwidth requirement d_m. Consequently, the required circuit capacity C_m for part m of the traffic is expressed as follows:

$$C_m = N_m \cdot d_m$$

where N_m denotes the value calculated by using methods described in the E.520-Series Recommendations.

Total circuit capacity C is given by the sum of the capacities dedicated to each traffic component.

$$C = \sum_{m=1}^{M} C_m$$

3.2 *Shared circuit groups*

Consider a circuit group that can accommodate *M* independent traffic components. Complete sharing of capacity in a total circuit group is allowed. In case of no overflow traffic, state probabilities have the product form [1]. In order to find minimal total circuit group capacity for achieving specified GOS objectives for different traffic components, blocking probability of each component can be calculated from state probabilities. Capacity of total circuit group should be dimensioned using methods described in Recommendation E.526 and in [1], [2], [3], [4], [5] and [6].

Note that savings in required capacity due to resource sharing depend on the number of different traffic components, their bit rate, traffic intensity and GOS requirements. Hence, completely partitioned circuit group, described in § 3.1, may require fewer circuits than shared group, if traffic components with different bit rate have quite different grades of service requirements.

3.3 *Path choice restrictions*

In the case in which it is required that all slots of a multi-slot call should be in a same circuit sub-group, e.g. in a same primary rate multiplex, methods for calculating blocking values for no overflow traffic are given in [2], [4], [7] and [8].

In some systems multi-slot calls must occupy fixed adjacent slots. Under these paths choice restrictions, precise blocking calculations and dimensioning depend on the individual situation and need further study.

4 Service protection methods

4.1 *Available methods*

It is clear that in the full availability circuit group, lower bit-rate traffic tend to have better GOS than higher bit-rate traffic [9]. To improve the blocking performance of higher bit rate traffic, and more generally, to control blocking performances of all traffic components, it is possible to apply service protection methods. The following service protection methods are suitable for ISDN and are based on the corresponding parts of Recommendation E.525:

– circuit reservation (also called sum limitation) where access of each traffic component is not allowed when the overall number of busy circuits exceed its assigned threshold;

– virtual circuits (also called class limitation) where the number of busy circuits occupied by each traffic component can not exceed its assigned limit.

4.2 *Evaluation and dimensioning*

4.2.1 *Circuit reservation*

For the circuit reservation type of service protection, total capacity of circuit group and parameters for circuit reservation should also be determined so as to achieve the GOS target. For the no overflow case, efficient blocking calculation algorithms and method for optimizing values of trunk reservation parameters are available, [3], [4], [10], [11], [12].

For circuit groups carrying overflow traffic and taking into account some reservation strategies, algorithms for calculating blocking values for each component are given in [6]. Methods for the calculations of optimal circuit reservation parameters are for further study.

4.2.2 *Virtual circuits*

For the virtual circuits type of service protection, state probabilities still have the product form, if the arrival processes are independent Poisson processes. Methods for the calculations of minimal total circuit group capacity to achieve the GOS criteria and optimal virtual circuit parameters are available for the no overflow case [13]. The case of circuit groups carrying overflow traffic is for further study.

4.3 *Performance characteristics*

Further study is required before detailed performance characteristics of service protection methods for ISDN traffic are recommended.

5 **Attribute negotiation**

ISDN offers the possibility of negotiating the service attributes between the user and the service or the network provider. Some Administrations may wish to introduce this facility; this section provides the means of evaluating the impact of the negotiation on the GOS and, in the cases that this impact is significant, the means of revising the dimensioning. Only the negotiation of the connection characteristics of a call demand, i.e. of those service attributes which are significant from a traffic point of view, as defined and listed in Recommendation E.711, § 2.2, will be considered.

If the network cannot accept the service as requested by the user, a negotiation may occur and, as a consequence, a call with some connection characteristics different from the ones initially requested may be served. Thus, negotiation allows a call, which without negotiation would be blocked, to be served. Consequently, the negotiation increases the traffic carried by the network and, therefore, increases the blocking probabilities of all the calls.

The impact of negotiation for services on demand is considered in this section, and for reserved services in § 6.

The values of some connection characteristics as, e.g. the information transfer capability, may affect the choice of the circuit group to be used by the call, as indicated in § 3.2 of Recommendation E.712. On the other hand, the value of other connection characteristics, as e.g. the information transfer rate, may only affect the required number of circuits of a given group.

5.1 *Connection characteristics affecting the choice of circuit group*

If a call cannot be carried by the circuit group required according to the connection characteristics requested by the user, the network may give the chance of carrying the call by another circuit group if the user accepts the implied change of connection characteristics. Thus, the network may work according to an overflow scheme, with a first choice circuit group and its subsequent choice circuit groups for each set of requested connection characteristics. The probabilities that the user accepts the changes of connection characteristics implied by overflow to other circuit groups (or, by the contrary, give up the call if it cannot be carried by the previous choice circuit groups) has to be considered in the evaluation. A safe-side approximation when these probabilities are unknown is to consider them equal to one. Evaluation of the impact on the blocking probabilities and, if necessary, revision of the dimensioning should be made using methods described in the E.520-Series Recommendations.

5.2 *Connection characteristics affecting the required number of circuits*

According to Recommendation E.712, § 3.3, the number of circuits required by a call depends on the information transfer rate and on the communication configuration (point-to-point or multipoint). The negotiation of the information transfer rate is considered here, while the negotiation of the communication configuration is considered in § 7.

Consider a circuit group of C circuits serving M independent traffic components, each component characterized by an arrival rate, λ_m, a holding time, h_m, and a number of circuits required per call, d_m. Complete sharing of capacity of the total circuit group is allowed. (The complete partitioning of the circuit group into sub-groups, each sub-group serving a traffic component, is covered by § 5.1.)

In case of no overflow traffic or negotiation, the state probabilities have the product form [1] and may be evaluated by the recurrent formula:

$$Q(n) = \sum_{m=1}^{M} \frac{\lambda_m \cdot h_m \cdot d_m}{n} \cdot Q(n - d_m)$$

where $Q(n)$ is the probability of having n circuits busy.

The blocking probability for the components m, B_m, is given by:

$$B_m = \sum_{n=C-d_m+1}^{C} Q(n)$$

Consider now that there is negotiation. If a call of the component m is offered and there is only k free circuits, with $k < d_m$, the network gives the chance of carrying the call with only k circuits. Define $\pi_{m,k}$ as the probability that the user accepts this chance. (If $\pi_{m,k}$ is unknown, it could be considered equal to 1 as a safe-side approximation). The state probabilities, $Q'(n)$, with this type of negotiation can be estimated from those without negotiation, $Q(n)$, by the following approximate formula:

$$Q'(n) = \frac{Q(n)}{1 + EB} \qquad \text{for } n < C$$

$$Q'(C) = \frac{Q(C) + EB}{1 + EB}$$

being:

$$EB = \sum_{m=1}^{M} \sum_{k=1}^{d_{m-1}} \frac{\lambda_m \cdot h_m \cdot k \cdot \pi_{m;k}}{n} \cdot Q(n - k)$$

If the change of the requested d_m circuits by only k circuits affects the holding time of the call which becomes $h'_{m,k}$ instead of h_m, h_m must be substituted by $h'_{m,k}$ in the previous formula.

The probability, B'_m, that a call of the component m cannot obtain the initially requested d_m circuits is:

$$B'_m = \sum_{n=C-d_m+1}^{C} Q'(n)$$

Note that B'_m is greater than B_m; thus, the negotiation increases the probability of carrying a call (with a bandwidth equal or smaller than the requested one) but decreases the probability of carrying it with the initially requested bandwidth. From the point of view of the users or services which cannot take advantage of the negotiation (because they cannot accept a bandwidth reduction), the blocking probability has increased without any compensation. In principle, this increase should be avoided, but depending on the increase produced in each case, the Administration should decide if it is worthwhile to revise the dimensioning.

6 Service reservation

Some services (videoconference, data communications, etc.) can be provided, exclusively or not, on an advanced reservation basis: users pre-book the necessary network resources for a planned call in a preliminary dialogue with a reservation centre. Both the user and the network provider may have reasons for using telecommunications services with advanced reservation: the former achieves in this way guaranteed service for scheduled calls, the latter can more easily offer high bit rate services (6, 24, 30×64 kbit/s), that, in many cases, could not be offered on a demand basis.

The user request may be characterized by the following relevant parameters [3]:

- the starting instant (T) and duration (D) of the reserved time;
- the required bandwidth (d).

If the system cannot accept the reservation as requested, a negotiation could be made about the values of these parameters.

The required resources may be either only dedicated to reservation services or also shared by on demand services.

6.1 *Dedicated resources and no negotiation*

The notice interval, defined as the time interval between the reservation request instant and the desired starting instant, is an essential characteristic of the reservation. There are not yet specifications on GOS requirements for reservation services in the E.720-Series Recommendations. It can be assumed that the specifications will be related to the parameters B and either $R(\tau)$ or $R'(\tau)$ defined as follows:

B: probability of blocking of a call whose reservation had been accepted;

$R(\tau)$: probability of rejection of a reservation request made with a notice interval equal to τ;

$R'(\tau)$: Probability of rejection of a reservation request made with a notice interval greater than or equal to τ.

In the case of resources only dedicated to reservation services, B is zero.

The investigation on the effect of the notice interval has led to the conclusion [14] that the system performance is independent of the mean of the notice interval (of the whole population of requests). It means that, e.g. two notice interval distributions $G_1(x)$ and $G_2(x)$, such that $G_1(x) = G_2\ (x - c)$, c being a constant lead to reservation rejection probabilities $R_1(\tau)$ and $R_2(\tau)$ respectively such the $R_1(\tau) = \mathrm{R}_2\ (\tau - c)$.

System performance however depends on the variance of the notice interval. If the notice interval is constant, the system behaves as with on demand services. Dimensioning methods for on demand services should be used. The blocking probabilities obtained using the on demand service models should be interpreted as reservation rejection probabilities, $R(c)$, c being the constant value of the notice interval.

For non-zero variance of notice, a call can be blocked by (previously reserved) calls commencing after its own requested start time. This phenomenon of "retro-blocking" [14] is more significant the greater the notice interval dispersion. There are no available models which consider the "retro-blocking" effect. As a first approximation, the "retro-blocking" effect can be neglected if only $R'(0)$, the global reservation rejection probability, independent of the notice interval, has to be evaluated. With this approximation, dimensioning methods for on demand services may be used. The blocking probabilities obtained from these methods should be interpreted as the global reservation rejection probabilities.

This approximation can be extended to evaluate $R'(\tau)$, the reservation rejection probability, of services requested with a notice interval greater than or equal to τ. This evaluation can be made by the same methods, but only considering the traffic corresponding to reservation requests made before the instant $T - \tau$.

Since these approximations neglect the "retro-blocking", they provide a rejection probability value independent of the duration of the reserved time, D. This value will be an underestimation of the rejection probability of calls with large D, and an overestimation for calls with small D. Further study is required to quantify these effects.

6.2 *Resources shared with on demand services (in the no negotiation case)*

The reservation rejection probabilities, $R(\tau)$ or $R'(\tau)$, defined in § 6.1 for the reservation services, are not affected by a sharing of the resources with on demand services. However, this sharing makes that, B the probability of blocking of a previously reserved call is not zero. It is due to the fact that the duration of the on demand calls is not previously known by the network and an on demand call accepted a certain time before the requested starting instant, T, of a reserved call may have not finished when the instant T arrives. Further studies are required to evaluate the time before T in which on demand calls must not be accepted to assume a low value of the blocking probability B of reserved calls.

A model to evaluate the GOS provided to the on demand calls by shared resources is given in [15]. The model assumes that the reservation traffic is for communications with d slots reserved for whole busy hour. The on demand calls require only one slot. The analysis shows that the grade of service for on demand calls is not the same qualitatively as with resources only dedicated to on demand services since the expected blocking probabilities strongly varies from busy hour to busy hour (depending on the amount of resources which have been reserved for each busy hour). Another result is that the capacity of the system, i.e. the dimensioning load which assumes a certain GOS for the on demand calls, decreases when the reservation percentage increases, mainly if d has a high value. Both results lead to the conclusions that an unrestricted mixing of reservation and demand traffic is not appropriate. A service protection method which guarantees a minimum number of resources for the demand services, thus assuring a certain grade of service for the demand calls in the worst busy hour, is recommended.

6.3 *Parameter negotiation*

The possibility of negotiation of the relevant parameters characterizing a user request is usually inherent to the reservation process. Those parameters are the starting instant, T, the duration, D, and the required bandwidth, d. A general model which considers simultaneous negotiation of the three parameters is for further study.

The impact of bandwidth negotiation can be evaluated by combining the methods described in §§ 5.2 and 6.1.

The impact of the negotiation of the starting instant can be modelled if the negotiation only includes the possibility of postponing the starting instant of the reserved time. If there are not available resources in the reserved time initially requested (T, $T + D$), the starting instant is postponed to the closest instant T' for which there are available resources in the interval (T', $T' + D$). The flexibility of the user to accept such delay may be modelled by a flexibility interval (T, $T + f$) in which the user accepts to start his service.

The system performance may be measured by the time $T' - T$ which the request is delayed and by the reservation rejection probability, i.e. the probability that a call cannot start inside its flexibility interval. If only global performance parameters, independent from the notice interval, have to be evaluated, and the "retro-blocking" effect is neglected, the system may be modelled as a queueing system for on demand services with impatient users: the time $T' - T$ that the starting instant is delayed can be interpreted as a queueing time of a demand service and the situation of a reservation rejected because his flexibility interval is not long enough to go out of the congestioned states can be interpreted as due to the impatience of the user facing a too long queueing time.

This modelling of the reservation services with negotiation of the starting instants by means of a queueing system with impatient users allows to apply Queueing Theory results to solve the problem for a variety of assumptions on arrival laws and distributions of reserved time durations and of lengths of user flexibility intervals. In particular, it has been solved in [16] for the following assumptions:

– continuous time, i.e. time is not slotted;

– Poisson arrivals of requested starting instants;

– exponential distribution of reserved time duration;

– exponential distribution of the length of user flexibility intervals.

7 Multipoint connections

As described in Recommendation E.712, § 3.3, several resources of the same group can be required by a call in multipoint connections, depending on the place where the information is split; Figure 1/E.731 gives two examples of a three points communication; if each single connection requires n slots, n resources are required in each trunk group in example a), while in example b), $2 \times n$ resources are required in the trunk group between the local exchange A and the transit exchange, and n resources in each of the other two trunk groups.

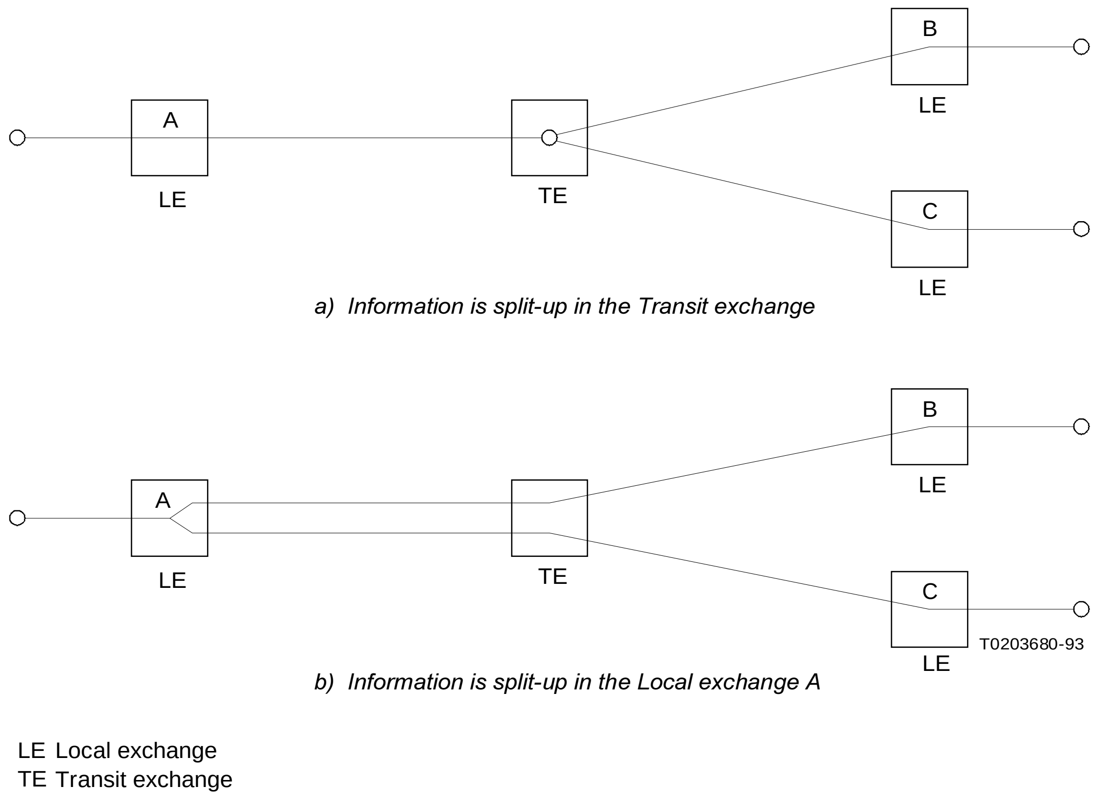

FIGURE 1/E.731

Two examples of implementation of a three-points communication

In general, three cases can be distinguished:

1) The single connections of a multipoint connection have to be established through different and independent trunk groups; e.g. connections of example a) or connections TE to LE B and TE to LE C of example b).

2) Several single connections of a multipoint connection have to be established through the same trunk group; e.g. connection LE A to TE of example b).

3) Several single connections of a multipoint connection have to be established through different trunk groups of the same network cluster.

In case 1), the same methods as for point-to-point connections should be used for dimensioning.

In case 2), the trunk group should be dimensioned as whether the multipoint connection would be a point-to-point multi-slot connection, thus applying the methods described in § 3.2. In the considered example b), this multipoint connection in the LE A to TE trunk group should be considered as a $2 \times n$ slot connection to take into account both the multipoint and the multi-slot natures of the considered connection.

Case 3) requires further study.

Negotiation of the number of points of a multipoint connection may be usual; this negotiation could be implicit: if a user requests connection to, e.g. three points and one of them fails the network could establish the connection to the other two points without an explicit negotiation; the reaction of the user could be to accept the point-to-two-points connection or to cut the communication. From a traffic point of view, this process should be considered as a real negotiation. If a trunk group has to carry several connections of a multipoint connection, this possibility of negotiation should be considered. Its impact can be evaluated using the methods described in § 5.2.

References

[1] KAUFMAN, (J.S.): Blocking in a shared resource environment, *IEEE Trans. Comm.* 29 1981.

[2] CONRADT, (J.) and BUCHHEISTER, (A.): Considerations on loss probability of multi-slot connections, *Proc. 11th ITC, paper 4.4B-2*, Kyoto, Japan (1985).

[3] ROBERTS, (J.W.): Teletraffic models for the Telecom 1 integrated service network, *Proc. 10th ITC, paper 1.1.2*, Montreal 1983.

[4] FICHE, (G.), LE GALL, (P.) and RICUPERO, (S.): Study of blocking for multi-slot connections in digital link systems, *Proc. 11th ITC, paper 5.4B-2*, Kyoto, Japan 1985.

[5] LE GALL, (P.): Overflow traffic combination and cluster engineering, *Proc. 11th ITC, paper 2.2B-1*, Kyoto, Japan 1985.

[6] CHANDRAMOHAN. (J.): An analytic multiservice performance model for a digital link with a wide class of bandwidth reservation strategies, *IEEE-JSAC*, Vol. 9, No. 2, 1991.

[7] MIYAKE, (K.): Traffic study on different bit-rate calls carried by a trunk group with trunk modularity, J. of the *IEICE, J71-B*, 1988.

[8] LUTTEN, (J.L.) and ROBERTS, (J.W.): Traffic performance of multi-slot call routing strategies in an integrated services digital network, *Proc. ISS 1984*, Florence, Italy 1884.

[9] GIMPELSON, (L.): Analysis of mixture of wideband and narrow-band traffic, *IEEE Trans. on Comm.*, Vol. 13, No. 3, 1965.

[10] LINDBERGER, (K.): Blocking for multi-slot heterogeneous traffic streams offered to a trunk group with reservation, *Proc. 5th ITC Seminar, Session on services*, Part II, Lake Como, Italy 1987.

[11] LINDBERG, (P.), NIVERT, (K.) and SAGERHOLM, (B.): Trunk reservation and grade of service issues in circuit switched integrated networks, *Proc. 12th ITC, paper 5.4A.3.1*, Turin, Italy 1988.

[12] TAKAGI, (K.) and SAKITA, (Y.): Analysis of loss probability equalized by trunk reservation for mixtures of several bandwidth traffic, *Proc. 12th ITC, paper 5.1A.5*, Turin, Italy 1988.

[13] KASHPER, (A.N.): Bandwidth allocation and network dimensioning for international multiservice networks, *Proc. 5th ITC Seminar, Session on networks*, Part II, Lake Como, Italy 1987.

[14] LIANG, (Y.), LIAO, (K.), ROBERTS, (J.W.) and SIMONIAN, (A.): Queueing models for reserved set up telecommunications service, *Proc. 12th ITC, paper 4.4B1*, Turin, Italy 1988.

[15] LIAO, (K.) and ROBERTS, (J.W.): Traffic models for telecommunication services with advance capacity reservation, *Proc. 11th ITC, paper 1.3.2*, Kyoto, Japan 1985.

[16] GIACOBBO SCAVO, (G.) and TONIETTI, (A.): Design of videoconference networks considering user flexibility, *Proc. 12th ITC, paper 4.4.B.3*, Turin, Italy 1988.

ANNEX A

(to Recommendation E.731)

Alphabetical list of abbreviations used in this Recommendation

GOS	Grade of service

ISDN	Integrated services digital network

LE	Local exchange

PSTN	Public switched telephone network

TE	Transit exchange