INTERNATIONAL TELECOMMUNICATION UNION

CCITT

E.733

THE INTERNATIONAL
TELEGRAPH AND TELEPHONE
CONSULTATIVE COMMITTEE

TELEPHONE NETWORK AND ISDN

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

METHODS FOR DIMENSIONING RESOURCES IN SIGNALLING SYSTEM No. 7 NETWORKS

Recommendation E.733

Geneva, 1992

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.733 was prepared by Study Group II and was approved under the Resolution No. 2 procedure on the 16th of June 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.

METHODS FOR DIMENSIONING RESOURCES IN SIGNALLING SYSTEM No. 7 NETWORKS

1 Introduction

This Recommendation provides a methodology for the planning of Signalling System No. 7 networks, which may be used for circuit-related signalling [e.g. to convey telephone User Part (TUP) and most ISDN User Part (ISUP) messages] and non-circuit-related signalling [e.g. to convey transaction capabilities Application Part (TCAP) messages]. Fundamentally different methods are required from those used for planning circuit switched telephone networks since Signalling System No. 7 is essentially a delay system and the service times are much shorter.

Section 2 describes the reference traffic and reference period to be used to dimension the number of signalling links and to ensure that the capacity of network switching elements is not exceeded. The factors for determining a maximum design link utilization, ρ_{max}, are given, which ensure that the end-to-end delay objectives described in Recommendation E.723 are met for the reference connection described therein. Initial values for ρ_{max} being used are described. Methods are then given for determining the number of signalling links required and the switching capacity required.

It is important to note that the efficiency of signalling links should not be the primary consideration when planning signalling networks. The performance of the network under failure and traffic overload has greater significance than when planning circuit switched telephone networks.

2 Reference traffic

2.1 This section defines the reference traffic needed to give the basis for dimensioning of Signalling System No. 7 networks. The Recommendation is based on Recommendation E.500 which gives the basis for the measurements needed for the design of circuit switched networks.

2.2 The reference period in Recommendation E.500 is one hour. In Signalling System No. 7 networks this is considered too long because the holding times are expected to be far shorter than in the circuit switched case. A reference period of shorter than one hour (e.g. five minutes) is suggested but the actual value is for further study. The question of whether the traffic process can be considered stationary over this period is for further study.

2.3 As an example, the recommended reference traffic is determined in the following way:

For each out of five busiest days the traffic intensity is measured in one minute periods. The intensity values are processed daily to find out the five consecutive one minute periods with the highest intensity value sum. Only this daily peak five minute traffic intensity is registered.

The reference load is the average taken over the five busiest days.

[This is based on the philosophy of the average daily peak hour (ADPH) method recommended in Recommendation E.500.]

The recommended method is for further study.

2.4 To account for the uncertainty concerning the stationarity of the traffic process, it may be useful to introduce a factor K multiplying the load used in § 3 for signalling link dimensioning [depending on whether a signal transfer point (STP) is used or not].

3 Dimensioning objectives

This section describes the objectives that should be used in dimensioning the links and nodes in Signalling System No. 7 networks.

Signalling networks provide the high availability required (see Recommendation Q.709) by providing diverse extra capacity to handle the load of any failed component. The amount of redundant capacity depends on the signalling network architecture. The nodes and links should be dimensioned to meet objectives specified for failure conditions that fully utilize the redundant capacity.

3.1 *Signalling link dimensioning objectives*

Signalling links should be dimensioned so that the link utilization, ρ, does not exceed a maximum utilization, ρ_{max}, when there are no failures in the network. To handle failures, the link should be able to support a utilization of $2\ \rho_{max}$. The signalling load that determines the link utilization, ρ, is determined as described in § 2.

3.1.1 Criteria for determining ρ_{max}

ρ_{max} is determined so that link performance criteria are met under the following network conditions:

– normal error condition;

– extreme error condition;

– transient conditions.

In determining ρ_{max} as described below, it is assumed that the signalling processing capacity in the receiving signalling terminal is not exceeded.

The performance criteria given below ensure the grade of service (GOS) objectives given in Recommendation E.723 are met, and in addition, provide additional protection against poor performance. The performance objectives given below shall apply for both basic and preventive cyclic retransmission (PCR) error correction (see Recommendation Q.703).

The following notation is used:

$\bar{m}$ is the mean message signal unit (MSU) length;

$\bar{s}$ is the mean MSU service time;

T_L is the signalling link loop propagation delay;

P_b is the bit error probability;

P_{SU} is the signal unit error probability;

ρ is the link utilization;

$Q(\rho)$ is the mean queueing delay (not including the emission time) on a signalling link operating at utilization ρ;

$Q^{99}(\rho)$ is the 99 percentile of the queueing delay on a signalling link operating at utilization ρ.

3.1.1.1 *Normal error condition*

The normal error condition on the signalling data link (level 1) is assumed to be random bit errors that occur at the rate of one error in 10^6 bits transmitted. Under this error condition the following should be satisfied:

a) $Q(2\rho_{max}) < D_1$ where $D_1 = \text{Max}(40, 0.4T_L)$ ms (provisional value);

b) $Q^{99}(2\rho_{max}) < D_1^{99}$ where $D_1^{99} = \text{Max}(200, 2T_L)$ ms (provisional value);

c) $\frac{dQ}{d\rho}(2\rho_{max}) < L_1$ where $L_1 = 200$ ms/E (provisional value);

d) $\frac{dQ^{99}}{d\rho}(2\rho_{max}) < L_1^{99}$ where $L_1^{99} = 1000$ ms/E (provisional value),

where $Q(\rho)$ and $Q^{99}(\rho)$ are the mean and 99% queueing delays on a signalling link operating at utilization ρ. These delays are deduced by mixing all the traffic streams offered to the considered link.

It is important to note that the above limits on delay are for severe load conditions. Under normal conditions the load will be at or below ρ_{max} and delays will be much smaller.

3.1.1.2 *Extreme error condition*

The extreme error condition is defined to be when the signalling link is operating at an error rate that puts it at the boundary of changeover, which is at a signal unit error probability $P_{SU} = 0.004$ (see Recommendation Q.706). When a link is operating at utilization $2\rho_{max}$ with PCR, the link will not be sending fill-in signal units (FISUs) when the error rate is high, and therefore all signal units will be new or retransmitted message signal units (MSUs). As a result the MSU error probability, P_m, will equal P_{SU}. For basic error correction, FISUs will be present and the bit error probability, P_b, and the average signal unit error probability, P_{SU}, are related by:

$$P_b = \left(\frac{(1 - \rho_{eff})}{6} + \frac{\rho_{eff}}{\bar{m}} \right) P_{su}$$

where

$$\rho_{eff} = \rho \left(\frac{1 + P_m T_L / \bar{s}}{1 - P_m} \right)$$

Thus, for the extreme error condition:

$$P_m = 0.004 \, [\rho_{eff} + (1 - \rho_{eff}) m/6]$$

Under the above defined extreme error condition, the following should be satisfied:

a) $Q(2\rho_{max}) < D_2$ where $D_2 = \text{Max}(60, 0.6T_L)$ ms (provisional value);

b) $Q^{99}(2\rho_{max}) < D_2^{99}$ where $D_2^{99} = \text{Max}(300, 3T_L)$ ms (provisional value);

c) $\frac{dQ}{d\rho}(2\rho_{max}) < L_2$ where $L_2 = 300$ ms/E (provisional value);

d) $\frac{dQ^{99}}{d\rho}(2\rho_{max}) < L_2^{99}$ where $L_2^{99} = 1500$ ms/E (provisional value), 3.1.1.3 *Transient conditions*

When a high error rate or signalling link changeover occurs, a transient results in signalling link buffers.

ρ_{max} should be chosen so that when these transients occur, the mean transient queueing delay on working links should be less than D_3 = 500 ms (provisional value) when all links are operating at ρ_{max} prior to the high error rate condition or signalling link changeover.

3.1.2 *Models used to determine queueing delays*

To evaluate $Q(\ )$ and $Q^{99}(\ )$ to apply the above criterion, a model or simulation results must be used. $\frac{dQ(\ )}{d\rho}$ and $\frac{dQ^{99}(\ )}{d\rho}$ can be evaluated from $Q(\ )$ and $Q^{99}(\ )$ using graphical methods. Models are available in Recommendation Q.706 for both basic and PCR error correction. These assume Poisson (non-correlated) arrivals for an isolated signalling link. For various network configurations these models may not be sufficient, and network effects such as variation in the inter-arrival time due to upstream transfer times, need to be considered.

Additional models are for further study. When there is uncertainty in the accuracy of the model, the choice of value for ρ_{max} needs to include some safety margin.

3.1.3 *Choosing between basic and PCR error correction*

The choice between basic and PCR should be based on the largest loop propagation delay T_L expected in the network. ρ_{max} is then determined to meet the criteria in §§ 3.1.1.1, 3.1.1.2 and 3.1.1.3. The choice between basic and PCR should be for the error correction method that gives the largest ρ_{max}.

3.2 *Node dimensioning*

Node dimensioning parameters recommended in this section are from the point of view of the network provider rather than the network element manufacturer.

Delay and congestion are the most important criteria for node dimensioning. These criteria should be applied considering evolution of the signalling network from the point of view of handling increasing traffic levels and the accepted traffic characteristics. Traffic characteristics are modelled by the message arrival processes and message length distributions. In other words, it is not just the forecasted amount of traffic which dictates the node dimensioning, but also the type of services deployed are important.

Other factors to be considered in node dimensioning are reliability, security and survivability considerations. For example, given certain amount of forecasted expansion of a signalling network load, there are several approaches to provision nodes for increased load as follows:

– one could increase the capacity of existing nodes;

– one could add additional nodes of larger capacity;

– one could reduce the number of nodes and add even larger capacity at remaining nodes;

– one could plan for a large number of smaller nodes.

The last option, though possibly more expensive, is more secure, reliable and survivable. It is more reliable and secure because of diversification of nodes,. i.e. failure of a single node affects a lesser amount of traffic. It is more survivable because in the event of natural or man made disasters the probability of larger amounts of traffic being affected is less. Obviously, quantification of such factors is not an easy task. These factors should be taken into account in node dimensioning by the network planners depending on specific circumstances and individual network requirements.

Also, network topology considerations have an effect on node dimensioning. For example, duplication of some signalling points, such as data bases, may have an effect.

Another complexity factor in this problem is that in a future intelligent network environment there will be a variety of nodes in the signalling network with varying specialized functions. For example, all or a subset of the following types of nodes could be present in future Signalling System No. 7 networks:

1) simple exchanges;

2) simple signal transfer points (STPs);

3) simple data base nodes;

4) nodes with both exchange and STP functions;

5) nodes with both data base and STP functions (e.g. co-location of global title translation data bases and STPs);

6) special purpose nodes (e.g. announcement nodes);

7) nodes with combinations of the above functions.

Obviously, covering all these combinations with a single set of criteria is not practical. It seems that the most practical method of approaching this problem is to define the common dimensioning criteria when applied to well known types of signalling points. We leave the more specific criteria for further study when better understanding of specific functions are available.

Common dimensioning criteria for signalling points are as follows:

3.2.1 *Signalling point capacity*

For exchanges, the notion of signalling capacity is not easily separable from the capacity of the exchange in terms of the number of circuits. What could be recommended from the signalling point of view is that the exchange must have enough signalling capacity so that when working at maximum call volume, it should be able to support the signalling process and support enough links to support the signalling messages and network architecture.

For STPs, the capacity could be defined as the number of MSUs that can be switched in unit time without causing processor congestion or undue cross-office delays. Also, it should be able to handle enough links to support the network architecture and carry the traffic load.

For Signalling connection Control Part (SCCP) relay points, the capacity could be defined as the number of MSUs that can be relayed in unit time without causing processor congestion or undue cross-office delays. Also, it should be able to handle enough links to carry the traffic.

For data bases, the capacity could be defined as the number of queries that could be processed in unit time without causing processor congestion or undue cross-office delays. Naturally, this capacity is closely related to the type of application. Also, it should be able to handle enough links to support the network architecture and carry the traffic load.

3.2.2 ***cross office signalling delay***

For exchanges, this is the time interval between the time some information is received by the Signalling System No. 7 sub-system from the user application (e.g. call processing) and the time the last bit of a corresponding message is delivered to Message Transfer Part (MTP) level 1 in that office. For POTS calls this delay includes processing times for ISUP (or TUP) and MTP processing delay. For messages bound to or from data bases, this delay includes processing times for TCAP, SCCP and MTP.

For STPs, this is the time interval between the moment the last bit of the incoming MSU is put in the receive buffer of the incoming link and the moment the last bit of the MSU is transmitted on the outgoing link.

For SCCP relay points, this is the time interval between the moment the last bit of the incoming MSU is put in the receive buffer of the incoming link and the moment the last bit of the response MSU is transmitted on the outgoing link.

For data bases, this is the time interval between the moment the last bit of the incoming MSU is put in the receive buffer of the incoming link and the moment the last bit of the response MSU is transmitted on the outgoing link minus the time needed for the application processing. This consists of MTP, SCCP and TCAP processing delays in both directions.

Values for cross-office signalling delays appear in Recommendations Q.706, Q.766 and Q.776.

3.2.3 *Signalling links*

The number of signalling links a signalling point can accommodate, is an important parameter for network planning. This parameter is especially important for STPs.

3.2.4 ***availability***

The availability of a signalling point is defined as the fraction of time that the signalling point is in full working condition.

3.3 Value for ρ_{max}

Presently the values used for ρ_{max} vary from 0.2 to 0.4.

4 Signalling link dimensioning methods

4.1 *Load calculation*

Recommendation E.713 gives the procedure to evaluate the signalling load between two nodes signalling point and/or signal transfer point (SP and/or STP) over a reference period. Dividing these quantities by the reference period length, yields under no failure conditions:

L' is the total load in bit/s in one direction,

L'' is the total load in bit/s in the opposite direction.

For dimensioning purposes, the significant parameter is the largest of the two. This is because a signalling link is actually a pair of unidirectional channels.

$L = \text{Max}\,(L', L'')$

4.2 *Single link capacity*

The capacity, C, of a single link is defined to be the maximum bit rate that a signalling link can carry with no failures in the network. It is calculated as:

$$C = S_L\,\rho_{max}$$

where

S_L is the link speed in bit/s and

ρ_{max} was defined in § 2 of this Recommendation.

4.3 *Link set capacity*

In Signalling System No. 7, load sharing over link sets is done using the 4 bit signalling link selection (SLS) field, and due to modularity effects this procedure does not always allow a completely balanced load distribution over links within a link set. As a result, not all of the signalling link capacity is available for use. Consequently, capacity of a link set is the maximum signalling load that can be shared without exceeding the capacity of any single link.

The number of SLS bits that are available for load sharing over a link set, depends on network architecture.

The Table 1/E.733 provides the link set capacity C_m, as a function of the single link capacity, C, the number of links in a link set m, and the number of SLS bits available for load sharing:

TABLE 1/E.733

Number of links m	Link set capacity (C_m)	
	4 SLS bits used	3 SLS bits used
1	C	C
2	$2C$	$2C$
3	$(8/3)C$	$(8/3)C$
4	$4C$	$4C$
5	$4C$	$4C$
6	$(16/3)C$	$4C$
7	$(16/3)C$	$4C$
8	$8C$	$8C$

Note – The link set capacity in Table 1/E.733 is the maximum allowed load when there are no failures in the network.

4.4 *Dimensioning procedure*

Given the calculated load, L (§ 4.1), and the single link capacity, C (§ 4.2), the number of links, m, required in the link set is obtained from the Table 1/E.733, assuring that $L \leq C_m$.

5 Recommendation history

Recommendation E.733 – First issue 1992.

ANNEX A

(to Recommendation E.733)

Alphabetical list of abbreviations used in this Recommendation

ADPH	Average daily peak hour
FISU	Fill-in signal unit
GOS	Grade-of-service
ISUP	ISDN User Part
MSU	Message signal unit
MTP	Message Transfer Part
PCR	Preventive cyclic retransmission
SCCP	Signalling connection Control Part
SLS	Signalling link selection
SP	Signalling point
STP	Signal transfer point
TCAP	Transaction Capabilities Application Part
TUP	Telephone User Part