INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

E.733

(11/98)

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

SERIES E: OVERALL NETWORK OPERATION, TELEPHONE SERVICE, SERVICE OPERATION AND HUMAN FACTORS

Quality of service, network management and traffic engineering – Traffic engineering – ISDN traffic engineering

Methods for dimensioning resources in Signalling System No. 7 networks

ITU-T Recommendation E.733

(Previously CCITT Recommendation)

ITU-T E-SERIES RECOMMENDATIONS

OVERALL NETWORK OPERATION, TELEPHONE SERVICE, SERVICE OPERATION AND HUMAN FACTORS

OPERATION, NUMBERING, ROUTING AND MOBILE SERVICES	
INTERNATIONAL OPERATION	
Definitions	E.100–E.103
General provisions concerning Administrations	E.104–E.119
General provisions concerning users	E.120–E.139
Operation of international telephone services	E.140–E.159
Numbering plan of the international telephone service	E.160–E.169
International routing plan	E.170–E.179
Tones in national signalling systems	E.180–E.199
Maritime mobile service and public land mobile service	E.200–E.229
OPERATIONAL PROVISIONS RELATING TO CHARGING AND ACCOUNTING IN THE INTERNATIONAL TELEPHONE SERVICE	
Charging in the international telephone service	E.230–E.249
Measuring and recording call durations for accounting purposes	E.260–E.269
UTILIZATION OF THE INTERNATIONAL TELEPHONE NETWORK FOR NON-TELEPHONY APPLICATIONS	
General	E.300–E.319
Phototelegraphy	E.320–E.329
ISDN PROVISIONS CONCERNING USERS	E.330–E.399
QUALITY OF SERVICE, NETWORK MANAGEMENT AND TRAFFIC ENGINEERING	
NETWORK MANAGEMENT	
International service statistics	E.400–E.409
International network management	E.410–E.419
Checking the quality of the international telephone service	E.420–E.489
TRAFFIC ENGINEERING	
Measurement and recording of traffic	E.490–E.505
Forecasting of traffic	E.506–E.509
Determination of the number of circuits in manual operation	E.510–E.519
Determination of the number of circuits in automatic and semi-automatic operation	E.520–E.539
Grade of service	E.540–E.599
Definitions	E.600–E.699
ISDN traffic engineering	**E.700–E.749**
Mobile network traffic engineering	E.750–E.799
QUALITY OF TELECOMMUNICATION SERVICES: CONCEPTS, MODELS, OBJECTIVES AND DEPENDABILITY PLANNING	
Terms and definitions related to the quality of telecommunication services	E.800–E.809
Models for telecommunication services	E.810–E.844
Objectives for quality of service and related concepts of telecommunication services	E.845–E.859
Use of quality of service objectives for planning of telecommunication networks	E.860–E.879
Field data collection and evaluation on the performance of equipment, networks and services	E.880–E.899

For further details, please refer to ITU-T List of Recommendations.

ITU-T RECOMMENDATION E.733

METHODS FOR DIMENSIONING RESOURCES IN SIGNALLING SYSTEM No. 7 NETWORKS

Summary

This Recommendation describes the methods for dimensioning resources in Signalling System No. 7 networks. It defines the reference traffic to be used for dimensioning and then describes the dimensioning objectives and dimensioning methods for links and nodes in signalling networks. It considers methods for dimensioning signalling links and nodes in Signalling System No. 7 networks.

Source

ITU-T Recommendation E.733 was revised by ITU-T Study Group 2 (1997-2000) and was approved under the WTSC Resolution No. 1 procedure on the 9th of November 1998.

FOREWORD

ITU (International Telecommunication Union) is the United Nations Specialized Agency in the field of telecommunications. The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of the ITU. The ITU-T 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 World Telecommunication Standardization Conference (WTSC), which meets every four years, establishes the topics for study by the ITU-T Study Groups which, in their turn, produce Recommendations on these topics.

The approval of Recommendations by the Members of the ITU-T is covered by the procedure laid down in WTSC Resolution No. 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation the term *recognized operating agency (ROA)* includes any individual, company, corporation or governmental organization that operates a public correspondence service. The terms *Administration, ROA* and *public correspondence* are defined in the *Constitution of the ITU (Geneva, 1992)*.

INTELLECTUAL PROPERTY RIGHTS

The ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. The ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, the ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementors are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database.

CONTENTS

Recommendation E.733

METHODS FOR DIMENSIONING RESOURCES IN SIGNALLING SYSTEM No. 7 NETWORKS

(revised in 1998)

1 Scope

This Recommendation considers methods for dimensioning signalling links and nodes in Signalling System No. 7 networks. Certain cases (e.g. due to new services) involving segmentation of long messages or mixtures of very long and short messages require further study to give good approximations for signalling link dimensioning purposes.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; all users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published.

– ITU-T Recommendation E.492 (1996), *Traffic reference period.*

– ITU-T Recommendation E.500 (1998), *Traffic intensity measurement principles.*

– CCITT Recommendation E.713 (1992), *Control plane traffic modelling.*

– CCITT Recommendation E.721 (1991), *Network grade of service parameters and target values for circuit-switched services in the evolving ISDN.*

– CCITT Recommendation E.723 (1992), *Grade-of-service parameters for Signalling System No. 7 networks.*

– ITU-T Recommendation E.734 (1996), *Methods for allocating and dimensioning Intelligent Network (IN) resources.*

– ITU-T Recommendation Q.703 (1996), *Signalling link.*

– ITU-T Recommendation Q.706 (1993), *Message transfer part signalling performance.*

– ITU-T Recommendation Q.709 (1993), *Hypothetical signalling reference connection.*

– ITU-T Recommendation Q.716 (1993), *Signalling Connection Control Part (SCCP) performance.*

– ITU-T Recommendation Q.766 (1993), *Performance objectives in the integrated services digital network application.*

3 Definitions

This Recommendation defines the following terms:

3.1 queueing delay: The queueing delay of a signalling message at the transmit queue of a signalling link is defined as the time from when the last bit of the message is placed in the transmit buffer until the first bit of the message is transmitted and the message is not subsequently retransmitted for level 2 error correction. Note that with this definition, in the case that a message is retransmitted one or more times for level 2 error correction, the queueing delay includes the time it takes to wait for a successful transmission.

3.2 emission time: The emission time of a signalling message on a signalling link is the time it takes to place all the bits of the message onto the transmission media. The emission time is equal to the length of the message (octets) divided by the transmission speed (octets/sec).

3.3 sojourn time: The sojourn time of a signalling message being sent over a signalling link is defined as the time from when the last bit of the message is placed into the transmit buffer until the last bit of the message is transmitted and the message is not subsequently retransmitted for level 2 error correction (i.e. the sum of queueing delay and emission time). The end-to-end sojourn time of a signalling message going over a path of signalling links and intermediate nodes is the time from when the last bit of the message is placed in the transmit queue of the first signalling link in the path until the last bit of the message is received at the far end of the path and the message is correctly received (i.e. it does not fail the level 2 CRC error check).

3.4 propagation time: The propagation time of a correctly received signalling message over a signalling link is the time from when the first bit of the message is placed on the transmission media until the first bit of the message is placed in the receive buffer at the receiving end of the link.

4 Abbreviations

This Recommendation uses the following abbreviations:

CRC	Cyclic Redundancy Check
DPP	Daily Peak Period
FISU	Fill-In Signal Unit
GOS	Grade of Service
IN	Intelligent Network
ISDN	Integrated Services Digital Network
ISUP	ISDN User Part
MSU	Message Signalling Unit
MTP	Message Transfer Part
PCR	Preventive Cyclic Retransmission
POTS	Plain Old Telephone Service
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

5 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.

Clause 6 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.

6 Reference traffic

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

6.2 The average traffic intensity is measured over time intervals called *read-out periods* (see Recommendation E.492). The length of the read-out period must be chosen so that acceptable estimates of traffic intensity are obtained. Specific considerations that must be made in choosing the read-out period are given in Recommendation E.500 and in 6.4 below.

6.3 The recommended reference traffic is determined in the following way:

The traffic intensity is measured continuously over a day for consecutive read-out periods. The highest intensity value for each day is retained.

The recommended daily traffic intensity measurement method is the one developed in Recommendation E.492, and it is called the Daily Peak Period (DPP) method. In this method the traffic intensity is measured for consecutive read-out periods of each day, and the peak traffic intensity for the day is recorded. Possible variations of this traffic measurement method are discussed in Recommendation E.500.

The reference traffic load for dimensioning is defined over a monthly time interval. A set of days is chosen (e.g. working days) that are approximately statistically homogeneous. The *normal* and *high* reference loads are determined by the following steps:

1) Order the chosen days from lowest to highest daily peak traffic intensity measurement.
2) The normal reference load is the *fourth* highest daily peak traffic intensity measurement, and the high reference load is the *second* highest peak traffic intensity measurement.

The signalling network dimensioning can be done using either normal or high loads. The dimensioning objectives developed in 7.1.1.1, 7.1.1.2, and 7.1.1.3 were developed assuming normal loads would be used. If high load is used, different (less stringent) performance objectives would need to be chosen.

6.4 To account for the variability of traffic intensity within the read-out period, it is useful to introduce a factor K which multiplies the reference load to determine the load user for signalling link dimensioning. The factor K represents the degree to which the actual traffic intensity is greater than for stationary Poisson traffic.

The factor K must be determined by special studies that take traffic measurements over shorter time intervals to determine the traffic variability within a read-out period for the network being considered. Studies have shown that this factor will vary considerably between networks (e.g. between 1.08 and 1.23 for two reported studies), so a single value cannot be recommended for all networks.

One method that has been used to estimate the factor K is as follows:

A sliding window of width w is used with window sizes, w, ranging from 1 second to 100 seconds. For each read-out period considered, the average load is computed over the sliding window when it is placed at intervals (0, w) (w, 2w), etc. For each window size, w, the maximum sliding window average load is determined, and the ratio of this maximum sliding window average to the average load over the read-out period is determined. This is called the measured peak-to-mean ratio for window size w.

The factor K is determined by comparing the measured peak-to-mean ratios for different window sizes to the corresponding expected peak-to-mean ratios assuming the message arrival process is a stationary Poisson process and a traffic intensity equal to the average load measured over the read-out period. The factor K for window size w, K_w, is the measured peak-to-mean ratio divided by the expected Poisson peak-to-mean ratio. The factor K is then the maximum of K_w over the considered window sizes, w. A method for determining the expected Poisson peak-to-mean ratios is now described.

For a window size w, the peak interval is the interval with the largest average load out of t_{ro}/w intervals, where t_{ro} is the length of the read-out period. Therefore, the peak interval is estimated to be the $(1 - w/t_{ro}) \times 100$ percentile of the distribution for the average load over an interval of length w. The average load over an interval of width w is a random variable that is most conveniently described in terms of the random variable N_w, which is the number of message arrivals in a period of length w. If the arrival process is Poisson, N_w has a Poisson distribution with mean and variance wL_{ro}/M, where L_{ro} is the measured average traffic load (octets/sec) over the read-out period and M is the average message length. The average load over an interval of length w can be expressed as the random variable $(M/w)N_w$, which has mean L_{ro} and variance $(M/w)L_{ro}$. When the Poisson random variable N_w is approximated by a normal distribution, the Poisson peak-to-mean ratio is estimated by the $(1 - w/t_{ro}) \times 100$ percentile of the normal distribution with unity mean and variance $M/(w\ L_{ro})$. If the Poisson distribution for N_w is used, the expected Poisson peak-to-mean ratio is $M/(w\ L_{ro})$ times the $(1 - w/t_{ro}) \times 100$ percentile of value for N_w.

If the length of read-out period, t_{ro}, is such that the K value is large (e.g. greater than 1.3), it is recommended that t_{ro} be reduced so the measured loads become more accurate and there is less dependence on the estimation of the K value.

7 Dimensioning objectives

This clause 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.

7.1 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 clause 6.

7.1.1 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 ensures 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 notations are used:

$\overline{m}$ is the mean Message Signal Unit (MSU) length (in octets);

$\overline{S}$ 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 transmit buffer queueing delay (not including the emission time) on a signalling link operating at utilization ρ;

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

7.1.1.1 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 = Max (40, $0.4T_L$) ms (provisional value);

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

c) $\dfrac{dQ}{d\rho}(2\rho_{max}) < L_1$ where L_1 = 200 ms/unit of utilization (provisional value);

d) $\dfrac{dQ^{(99\%)}}{d\rho}(2\rho_{max}) < L_1^{(99\%)}$ where $L_1^{(99\%)}$ = 1000 ms/unit of utilization (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.

7.1.1.2 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:

$$P_b = \frac{1}{8}\left(\frac{1-\rho_{eff}}{6} + \frac{\rho_{eff}}{m}\right)P_{su}$$

where

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

Thus, for the extreme error condition:

$$P_m = 0.004\left(\rho_{eff} + \left(1-\rho_{eff}\right)\frac{\overline{m}}{6}\right)$$

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

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

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

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

d) $\frac{dQ^{(99\%)}}{d\rho}(2\rho_{max}) < L_2^{(99\%)}$ where $L_2^{(99\%)}$ = 1500 ms/unit of utilization (provisional value).

7.1.1.3 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.

7.1.2 Models used to determine queueing delays

7.1.2.1 Conditions for using M/G/1 models

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.

M/G/1 models are available in Recommendation Q.706 for both basic and PCR error correction. These models assume that the message arrival process is Poisson. In actual networks this assumption is not precisely met, but the model is still an acceptable approximation when the following conditions are met:

1) A Poisson call arrival process is a good approximation to the actual original call arrival process.
2) The time separation between messages in the same direction associated with the same call are greater than 1 second for most calls. This is needed in order for the known signalling message correlation to not significantly alter the queueing behaviour from the M/G/1 model.
3) The signalling point processing does not significantly distort message interarrival times (i.e. does not cause significant batching or smoothing of messages).

When signalling messages traverse multiple links prior to entering a link being considered, the M/G/1 model may not be accurate at the considered link depending on the network architecture and message length distribution. Simulation results have been developed to examine these effects, and in particular the effects of long messages have been studied. The following points summarize the findings from those simulation studies:

- When there is a reasonable amount of random mixing of traffic onto a signalling link from other signalling links, the M/G/1 model gives a good approximation for computing queueing delay statistics.
  - A "reasonable amount of mixing" means that no more than 10 to 20% of any one link's load is sent to the considered link, and no one link contributes more than 10 to 20% of the total load on the considered link.
  - The M/G/1 model "gives a good approximation" means that the mean and 95% queueing delays predicted by the M/G/1 model are within 10 to 20% of the actual values.
- When there are tandem paths of signalling links (i.e. most of the load from one link goes to the next link in the path), the M/G/1 model will not be accurate.
  - There will be a pronounced effect of long messages in that batches of short messages will form behind long messages.
- When there is a noticeable impact of long messages, the effect is, in general, to:
  - make the queueing delays for long messages less than the M/G/1 models would predict;
  - make the queueing delays for short messages higher than the M/G/1 models would predict, except for rare exceptions.
- The percentage changes from M/G/1 behaviour decrease as the considered link's utilization increases.

7.1.2.2 Segmentation of long messages

When segmentation of long messages takes place to transport such messages over a Signalling System No. 7 network, the segments of a long message will have correlated interarrival times as they flow through the signalling network. Therefore, the message segments cannot be assumed to be independent arrivals, and thus the Q.706 M/G/1 models may not be accurate. As the segments of a long message flow through a signalling network, the interarrival times between segments become larger due to other messages arriving and coming in between segments. If the separation between segments is large enough, an M/G/1 model in which all Message Signal Units (MSUs) are assumed to have Poisson arrivals will be a good approximation. This approximation is identified as Model 1. On the other hand, if the MSUs from a segmented message stay very close together, an M/G/1 model in which the segmented message is considered as a batch of MSUs (one long message) would be a good approximation. This approximation is identified as Model 2.

Simulation studies have shown that Models 1 and 2 can be used to determine lower and upper bounds for queueing delays and end-to-end sojourn times when Conditions 1) to 3) above are met.

Consider queueing delays for MSUs of non-segmented messages and the first MSU of segmented messages. For signalling links for which all the segmented traffic is being carried on a signalling link for the first time (so the time separation between segments from a long message are small enough that there is negligible probability that other MSUs will arrive in between segments), the assumptions of Model 2 are met and so it is a good approximation. For signalling links for which all the segmented traffic has gone through at least one previous signalling link, simulation studies have shown that Model 1 is a good approximation when link utilizations are somewhat less than 0.6. For larger utilizations the queueing delays lie between Model 1 and Model 2, with neither being a close approximation.

Simulation studies have also shown Models 1 and 2 can be used to provide lower and upper bounds on segmented message end-to-end sojourn times. The Model 1 lower bound is determined by computing the sojourn time of the first segment of the long message using queueing delays from Model 1 (Model 2 can be used for the first link if it carries only first offered segmented messages) and then adding to that sojourn time the emission time of each subsequent segment of the long message. The Model 2 upper bound is determined by treating the long message as a single message and computing its end-to-end sojourn time using queueing delays from Model 2. Neither of these bounds gives a good approximation, except at low link utilizations (less than 0.2 for all links) when the Model 1 bound gives a close approximation.

7.1.3 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. If ρ_{max} is determined to meet the criteria in 7.1.1.1, 7.1.1.2 and 7.1.1.3, then the choice between basic and PCR should be for the error correction method that gives the largest ρ_{max}.

Analytic and measurement studies have shown that for links carrying typical ISUP traffic (average message length of 23 octets), and for any other signalling traffic having higher average message lengths, basic error correction will have a higher ρ_{max} (if ρ_{max} is determined to meet the criteria in 7.1.1.1, 7.1.1.2 and 7.1.1.3) for loop propagation delay, T_L, up to 250 ms. For signalling traffic with larger average message lengths, the limit on T_L for using basic error correction will be higher than 250 ms. For these cases an analysis to determine ρ_{max} for basic and PCR should be done for the specific traffic characteristics and maximum expected T_L to determine whether basic or PCR should be used. Annex A provides some more detail on the analytic studies comparing basic and PCR error correction.

Considering the provisional nature of the limiting values in the conditions given in 7.1.1.1, 7.1.1.2 and 7.1.1.3, it should be noted that the cross-over point of 250 ms significantly depends on those limiting values. Reducing those values will also reduce the cross-over point.

If ρ_{max} is determined by means other than described above, the choice between basic and PCR method should be for the method which minimizes the mean queueing delay for a given link utilization. This may lead to results different from those above. Since the queueing delay depends heavily on the traffic parameters (bit error rate and mean message length for example), an analysis considering the specific traffic characteristics should be done. If the bit error rate is very small, the difference between the BEC and the PCR method becomes negligible. For extreme error conditions, Annex A provides a comparison of the basic and the PCR methods.

7.2 Node dimensioning

Node dimensioning parameters recommended in this subclause are from the point of view of the network provider rather than of 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. Anyway, in dimensioning signalling points and their components, failure situations in which links of the node may be at load $2\rho_{max}$ must be taken into account.

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

Another complexity factor in this problem is that in an 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 Signalling System No. 7 networks:

1) simple exchanges;

2) simple Signal Transfer Points (STPs);

3) simple database nodes;

4) nodes with both exchange and STP functions;

5) nodes with both database and STP functions (e.g. co-location of global title translation databases 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. Similar considerations on dimensioning IN resources are presented in Recommendation E.734.

Common dimensioning criteria for signalling points are as follows:

7.2.1 Capacity

The signalling capacity of a switch depends on the number of signalling relations and the traffic volume of call-related and non-call-related signalling.

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 databases, 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.

7.2.2 Cross-office signalling delay

For a transit exchange, 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 POTS calls this delay includes processing times for ISUP (or TUP) and MTP processing delay. For messages bound to or from databases, 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 databases, 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.716.

7.2.3 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.

7.2.4 Availability

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

7.3 Value for ρ_{max}

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

8 Signalling link dimensioning methods

8.1 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.

– $L = \text{Max}\,(L', L'')$.

8.2 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} is defined in clause 7.

8.3 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.

Table 1 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 is the maximum allowed load when there are no failures in the network.

In determining the link set capacity given in Table 1, it has been assumed that signalling traffic load between every signalling point pair is uniformly distributed over the SLS codes (in terms of both traffic intensity and message length distribution). If this is not the case, a more detailed analysis is required that takes into account the different traffic characteristics of different SLS codes.

8.4 Dimensioning procedure

Given the calculated load, L (see 8.1), and the single link capacity, C (see 8.2), the number of links, m, required in the link set is obtained from Table 1, assuring that $L \le C_m$.

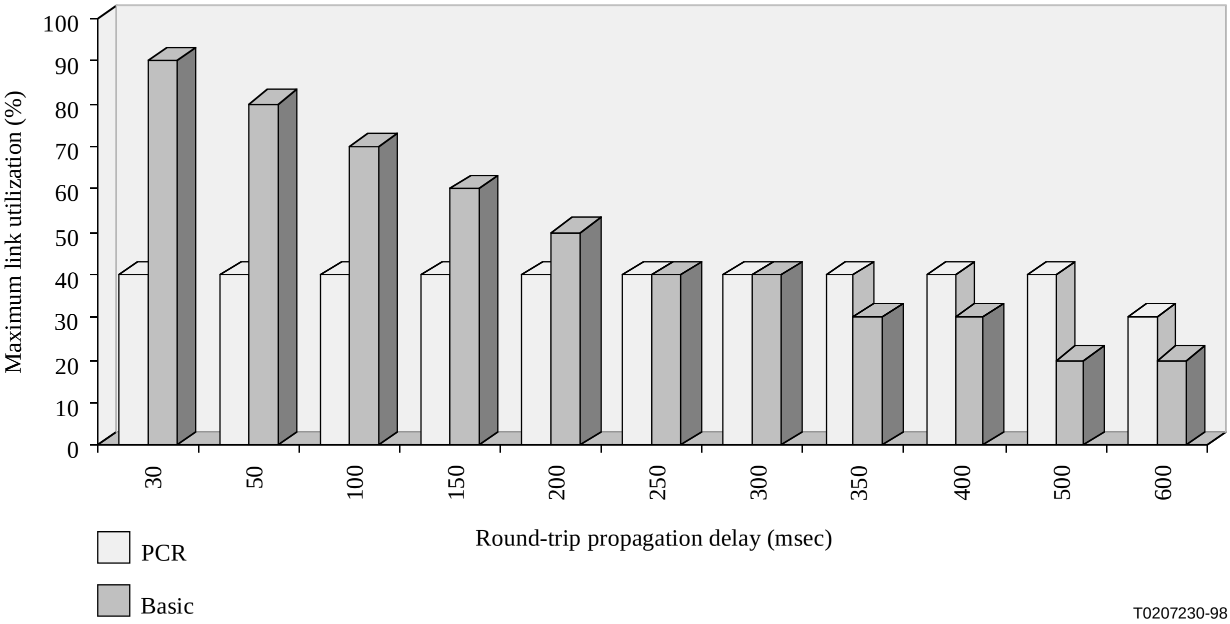
9 Recommendation history

Recommendation E.733 – First issued in 1992; revised in 1996, revised in 1998.

ANNEX A

Analytic results comparing basic and PCR error correction

This Annex provides some analytical results that compare the maximum link utilization, ρ_{max}, for basic and PCR error correction. The criterion considered is Criterion a) in 7.1.1.2. That is, the extreme error condition is considered, so the signal unit error probability is set at 0.004; and ρ_{max} is determined by requiring the queueing delay at link utilization $2\rho_{max}$ to be less than Max(60, $0.6T_L$), where T_L is the loop propagation delay. The queueing delays are determined by using the queueing delay formulas for basic and PCR given in Table 2/Q.706. Figure A.1 shows the comparison of $2\rho_{max}$ for basic and PCR when the signalling traffic has a constant message length of 23 octets. The use of a typical message length distribution with the mean message length of 23 octets results in the same choice between the basic and PCR methods.



NOTE – The calculated maximum link utilization values are rounded down to nearest multiples of 10 in the graph.

Figure A.1/E.733 – Calculated maximum link utilization with varying loop propagation delays

Figure A.2 compares the queueing delays for the basic and PCR methods under the extreme error condition. Using the same assumptions as above, the considered traffic is a Poisson traffic with constant message length of 23 octets. A link utilization of 30% is assumed. The queueing delay is plotted versus the loop propagation delay T_L. The PCR method results in a smaller queueing delay than the basic method for loop propagation delays greater than 30 msec.

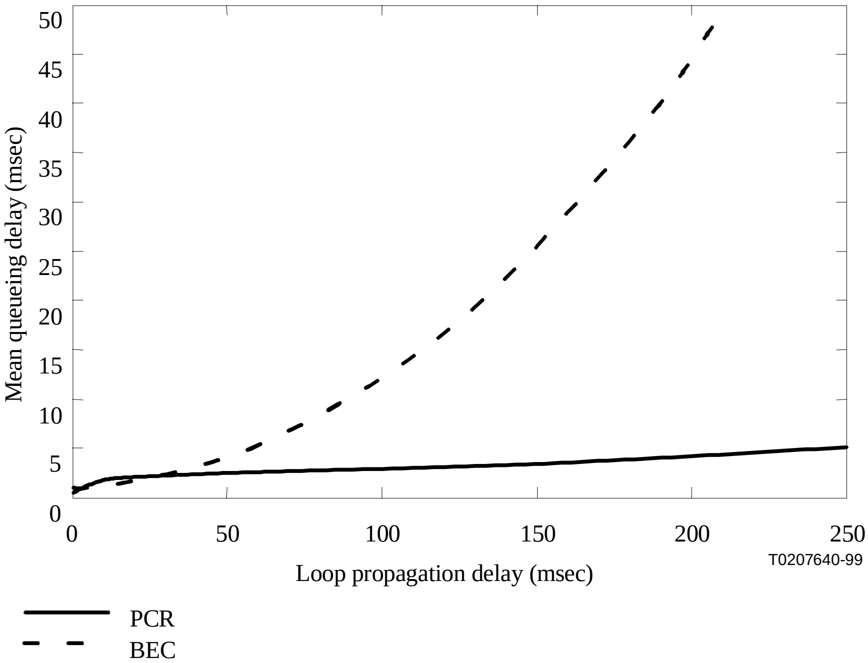


Figure A.2/E.733 – Calculated mean queueing delays with varying loop propagation delays

ITU-T RECOMMENDATIONS SERIES

Series A	Organization of the work of the ITU-T
Series B	Means of expression: definitions, symbols, classification
Series C	General telecommunication statistics
Series D	General tariff principles
Series E	**Overall network operation, telephone service, service operation and human factors**
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	TMN and network maintenance: international transmission systems, telephone circuits, telegraphy, facsimile and leased circuits
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks and open system communications
Series Y	Global information infrastructure
Series Z	Programming languages