INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

Q.543

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

(03/93)

DIGITAL EXCHANGES

DIGITAL EXCHANGE PERFORMANCE DESIGN OBJECTIVES

ITU-T Recommendation Q.543

(Previously "CCITT Recommendation")

FOREWORD

The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of the International Telecommunication Union. 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, established the topics for study by the ITU-T Study Groups which, in their turn, produce Recommendations on these topics.

ITU-T Recommendation Q.543 was revised by the ITU-T Study Group XI (1988-1993) and was approved by the WTSC (Helsinki, March 1-12, 1993).

NOTES

1 As a consequence of a reform process within the International Telecommunication Union (ITU), the CCITT ceased to exist as of 28 February 1993. In its place, the ITU Telecommunication Standardization Sector (ITU-T) was created as of 1 March 1993. Similarly, in this reform process, the CCIR and the IFRB have been replaced by the Radiocommunication Sector.

In order not to delay publication of this Recommendation, no change has been made in the text to references containing the acronyms "CCITT, CCIR or IFRB" or their associated entities such as Plenary Assembly, Secretariat, etc. Future editions of this Recommendation will contain the proper terminology related to the new ITU structure.

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

CONTENTS

DIGITAL EXCHANGE PERFORMANCE DESIGN OBJECTIVES

(Melbourne 1988, modified at Helsinki 1993)

1 General

This Recommendation applies to digital local, combined, transit and international exchanges for telephony in Integrated Digital Networks (IDN) and mixed (analogue/digital) networks, and also to local, combined, transit and international exchanges in an Integrated Services Digital Network (ISDN).

The field of application of this Recommendation is more fully defined in Recommendation Q.500. As to the application in an ISDN, transit connections and exchange connection types I, II, III and IV as defined in Recommendation Q.522 are covered (see Notes 1 and 2 in 2.1). Other types of connection and variants of these connections may be feasible in ISDN and will be the subject of further study.

These performance design objectives are applicable to all exchange implementations at all points in the growth cycle up to the maximum size. These reference loads and performance objectives may be used by manufacturers in designing digital switching systems and by Administrations in evaluating a specific exchange design or for comparing different exchange designs for potential use in the Administration's intended implementation.

These recommended performance design objectives relate to the technical capabilities of exchange design. They are intended to assure that exchanges operating in their intended implementation will be capable of supporting the network grades of service recommended in the E.500-Series Recommendations and will offer a level of performance consistent with the overall network performance objectives given in the I-Series Recommendations. The recommended parameters are design objectives which should not be construed to be grade of service or operating requirements. In actual operation, exchanges will be engineered to provide adequate grades of service as economically as possible and the performance requirements (delays, blocking, etc.) of the exchange in operation will differ from the recommended values for these performance design objectives.

2 Performance design objectives

2.1 Reference loads

The given reference loads are traffic load conditions under which the performance design objectives stated in 2.2 to 2.7 are to be met. In order to have a comprehensive characterization of exchange reference loads, supplementary services and other types of services must be taken into account. Administrations may specify hypothetical models for use in computing exchange loading. These models should characterize the sets of traffic parameters and services that are considered to be typical in the intended application of the exchange, and should include the traffic mix (originating-internal, originating-outgoing, incoming-terminating, transit, abandoned, busy non-answer, etc.), the mix of service classes (residential, business, PABX, coin, etc.), the types and volume of supplementary services (call waiting, call forwarding, etc.) and any other pertinent characteristics. Using the above information, it should be possible to "engineer" the exchange to produce the model. It should also be possible to determine the maximum size of the exchange by the computations discussed in 2.1.4.

Reference load A is intended to represent the normal upper mean level of activity which Administrations would wish to provide for on customer lines and inter-exchange activities. Reference load B is intended to represent an increased level beyond normal planned activity levels.

NOTES

1 For the time being, the following definitions and corresponding values are only applicable to 64 kbit/s circuit switched connections, i.e., including transit connections and connection types I, II and III option a). Other rates and transfer modes require further study.

2 The applicability of this Recommendation to connections originating or terminating on PABXs is for further study.

2.1.1 Reference load on incoming interexchange circuits

a) *Reference load A*

– 0.7 erlangs average occupancy on all incoming circuits

$$\text{Call attempts/h} = \frac{0.7 \times \text{number of incoming circuits}}{\text{Average holding time in hours}}$$

NOTE – Ineffective call attempts must be included in reference call attempts.

b) *Reference load B*

– 0.8 erlangs average occupancy on all incoming circuits

with 1.2 times the call attempts/h for reference load A.

2.1.2 Reference load on subscriber lines (originating traffic)

Characteristics of traffic offered to local exchanges vary widely depending upon factors such as the proportions of residence and business lines that are served. Table 1 provides reference load characteristics for lines typical of four possible local exchange applications. Also provided are representative ISDN cases which are discussed below. Administrations may elect to use other models and/or loads that are more suitable for their intended application.

In the following text, ISDN lines will be referred to as digital lines and non-ISDN lines as analogue lines.

2.1.2.1 Reference load A

TABLE 1a/Q.543

Subscriber line traffic model – Non-ISDN subscriber lines with or without supplementary services

Exchange type	Average traffic intensity	Average BHCA
W	0.03 E	1.2
X	0.06 E	2.4
Y	0.10 E	4
Z	0.17 E	6.8

The following ISDN models and traffic parameters are provisional and may be revised in subsequent study periods.

TABLE 1b/Q.543

Subscriber line traffic model – ISDN digital subscriber access 2B + D

Line type	Average traffic intensity per B channel	Average BHCA per B channel	Average packets per second per D channel
Y′	0.05 E	2	0.05 (signalling) + Data packets[a)]
Y″	0.10 E	4	0.1 (signalling) + Data packets[a)]
Y‴	0.55 E	2	0.05 (signalling) + Data packets[a)]

BHCA Busy hour call attempts.

[a)] Data packet rates are for further study. These include teleaction and packet services data.

Even though only limited ISDN traffic data is available, the specification of the corresponding reference load remains an important factor in exchange evaluation. For the case of digital subscriber lines in Table 1b), access is assumed to utilize the Basic Access with 2B + D channels. The B channels are available for circuit-switched calls, while the D channel is used to carry signalling information or may be used to carry teleaction data and packet switched data. It is assumed that digital lines typically carry traffic comparable with the heavy-traffic analogue lines designated as case Y in Table 1a). Three cases representing likely ISDN applications are included in the table.

Case Y′ traffic per pair of B channels comparable to 1 Case Y line.

Case Y″ traffic per pair of B channels comparable to 2 Case Y lines.

Case Y‴ traffic per pair of B channels comparable 1 Case Y line plus some very high traffic (e.g. circuit switched data traffic at 1 erlang).

Each of these digital lines also carries the associated ISDN signalling and data services on the D channel. For the circuit switched calling rates specified in Table 1b), ISDN signalling is expected to contribute less than 0.05 packet per second per digital subscriber line. The packet rates for D channel ISDN data services can be much larger than this; however, these are left for further study.

2.1.2.2 Reference load B

Reference load B is defined as a traffic increase over reference load A of: +25% in erlangs, with +35% in BHCA.

Reference load B levels for D channel activity are for further study.

2.1.3 Impact of supplementary services

If the reference model exchange assumes that significant use is made of supplementary services, the performance of the exchange can be strongly affected, especially in exchange designs where processor capacity can become a limiting item. The performance delays recommended in 2.3 and 2.4 can be significantly lengthened at a given call load under such circumstances. The Administration or Operating Agency defining the reference model should estimate the fractions of calls which use various supplementary services so that an average processor impact relative to a basic telephone call can be calculated (e.g. possibly by a methodology similar to that of Annex A).

2.1.4 Exchange capacity

In order to evaluate and compare exchange designs, an Administration will usually want to know the maximum possible size of the exchange for the intended implementation. While several factors may limit exchange capacity, processing capacity will frequently be the limiting factor. The maximum possible number of lines and circuits served by an exchange, while meeting performance objectives, will depend on the mix, volumes and types of traffic and the services expected in the particular implementation.

Two methods of determining exchange processing capacity are provided in the annexes to this Recommendation:

– Annex A provides an example of methodology for computing processing capacity of an exchange using information provided by the manufacturer and estimates of traffic mix and load provided by the Administration.

– Annex B provides an example of methodology for estimating the capacity of an exchange by making projections from measurements made on a functioning exchange in the laboratory or in the field. The test exchange must be representative of mix and load of traffic and services expected at maximum size.

2.1.5 Reference loads on other accesses and interfaces

At this time, other applications, such as n × 64 kbit/s on the Primary Rate Interface, are left for further study.

2.2 inadequately handled call attempts

2.2.1 Definition

Inadequately handled call attempts are attempts which are blocked (as defined in the E.600-Series Recommendations) or are excessively delayed within the exchange. "Excessive delays" are those that are greater than three times the "0.95 probability of not exceeding" values recommended in the tables in 2.3 and 2.4. (See Note.)

For originating and transit calls, this inadequately handled call attempt parameter applies only when there is at least one appropriate outlet available.

NOTE – Provisionally, call request delay is not included in this parameter. Further study is required.

2.2.2 Probability of inadequately handled call attempts occurring

The values in Table 2 are recommended.

TABLE 2/Q.543

Type of connection	Reference load A	Reference load B
Internal	10^{-2}	4×10^{-2}
Originating	5×10^{-3}	3×10^{-2}
Terminating	5×10^{-3}	3×10^{-2}
Transit	10^{-3}	10^{-2}

2.3 Delay probability – non-ISDN or mixed (ISDN – non-ISDN) environment

The non-ISDN environment is composed of analogue subscriber lines and/or circuits that use either channel associated or common channel signalling.

The ISDN environment is composed of digital (ISDN) subscriber lines and/or circuits that use common channel signalling.

This subclause defines delay parameters related to non-ISDN environment and mixed (ISDN – non-ISDN) environment.

When a delay parameter in this subclause is also applicable to the pure ISDN environment, a reference to the appropriate part of 2.4 (delay probability – ISDN environment) is provided.

In the following delay parameters, it is understood that delay timing begins when the signal is "recognizable", that is, after the completion of signal verification, where applicable. It does not include line-dependent delays for the recognition of induced voltage conditions or line transients.

The term "mean value" is understood to be the expected value in the probabilistic sense.

Where several messages are received at the exchange from a digital subscriber line signalling system (e.g. several alert messages are received from a multi-user configuration), the message that is accepted for call handling is the one considered in determining the start of a given delay interval.

Where common channel signalling (including inter-exchange and subscriber line signalling) is involved, the terms "received from" and "passed to" the signalling system are used. For CCITT Signalling System No. 7, this is designated as the instant the information is exchanged between the signalling data link (layer 1) and the signalling link functions (layer 2). For digital subscriber line signalling, this is designated as the instant the information is exchanged by means of primitives between the data link layer (layer 2) and the network layer (layer 3). Thus, the time intervals exclude the above layer 1 (CCITT Signalling System No. 7), and layer 2 (D channel) times. They do, however, include queuing delays that occur in the absence of disturbances but not any queuing delays that occur in the absence of disturbances but not any queuing delays caused by re-transmission.

2.3.1 incoming response delay – transit and terminating incoming traffic connections: incoming response delay is a characteristic that is applicable where channel associated signalling is used. It is defined as the interval from the instant an incoming circuit seizure signal is recognizable until a proceed-to-send signal is sent backwards by the exchange.

The values in Table 3 are recommended.

TABLE 3/Q.543

	Reference load A	Reference load B
Mean value	≤ 300 ms	≤ 400 ms
0.95 probability of not exceeding	400 ms	600 ms

2.3.2 local exchange call request delay – originating outgoing and internal traffic connections

2.3.2.1 For ANALOGUE SUBSCRIBER LINES, call request delay is defined as the interval from the instant when the off-hook condition is recognizable at the subscriber line interface of the exchange until the exchange begins to apply dial tone to the line. The call request delay interval is assumed to correspond to the period at the beginning of a call attempt during which the exchange is unable to receive any call address information from the subscriber.

The values in Table 4 are recommended.

TABLE 4/Q.543

	Reference load A	Reference load B
Mean value	≤ 400 ms	≤ 800 ms
0.95 probability of not exceeding	600 ms	1000 ms

NOTES

1 The above values are understood to apply when a continuous tone, i.e., without a cadence, is used and do not include delays caused by functions such as line tests, which may be used in national networks.

2 For Systems with waiting probability less than 0.05, the values for the "0.95 probability of not exceeding" might be meaningless.

2.3.2.2 For DIGITAL SUBSCRIBER LINES using overlap sending, call request delay is defined as the interval from the instant at which the SETUP message has been received from the subscriber signalling system until the SETUP ACKNOWLEDGE message is pased back to the subscriber signalling system.

NOTE – In this case this parameter is equivalent to the user signalling acknowledgement delay (see 2.4.1).

The values in Table 5 are recommended.

TABLE 5/Q.543

	Reference load A	Reference load B
Mean value	≤ 400 ms	≤ 800 ms
0.95 probability of not exceeding	600 ms	1000 ms

2.3.2.3 For DIGITAL SUBSCRIBER LINES using en-bloc sending, call request delay is defined as the interval from the instant at which the SETUP message is received from the subscriber signalling system until the call proceeding message is passed back to the subscriber signalling system.

The values in Table 6 are recommended.

TABLE 6/Q.543

	Reference load A	Reference load B
Mean value	≤ 600 ms	≤ 900 ms
0.95 probability of not exceeding	800 ms	1200 ms

2.3.3 exchange call set-up delay – transit and originating outgoing traffic connections: Exchange call set-up delay is defined as the interval from the instant that the information is required for outgoing circuit selection is available for processing in the exchange, or the signalling information required for call set-up is received from the signalling system, until the instant when the seizing signal has been sent to the subsequent exchange or the corresponding signalling information is passed to the signalling system.

2.3.3.1 Exchange call set-up delay for transit connections

2.3.3.1.1 For transit traffic connections that involve circuits that use channel associated signalling or a mix of channel associated and common channel signalling, the values in Table 7 are recommended.

TABLE 7/Q.543

	Reference load A	Reference load B
Mean value	≤ 250 ms	≤ 400 ms
0.95 probability of not exceeding	300 ms	600 ms

2.3.3.1.2 For transit traffic connections between circuits that use CCITT Signalling System No. 7 signalling exclusively, the requirements of the appropriate signalling system Recommendation should apply, e.g. Recommendations Q.725 and Q.766 for T_{cu} value (case of a processing intensive message).

2.3.3.2 Exchange call set-up delay for originating outgoing traffic connections

2.3.3.2.1 For outgoing traffic connections originating from ANALOGUE SUBSCRIBER LINES, the values in Table 8 are recommended.

TABLE 8/Q.543

	Reference load A	Reference load B
Mean value	≤ 300 ms	≤ 500 ms
0.95 probability of not exceeding	400 ms	800 ms

2.3.3.2.2 For outgoing traffic connections originating from DIGITAL SUBSCRIBER LINES using overlap sending, the time interval starts when the INFORMATION message received contains a "sending complete indication" or when the address information necessary for call set-up is complete.

The values in Table 9 are recommended.

TABLE 9/Q.543

	Reference load A	Reference load B
Mean value	≤ 400 ms	≤ 600 ms
0.95 probability of not exceeding	600 ms	1000 ms

2.3.3.2.3 For outgoing traffic connections originating from DIGITAL SUBSCRIBER LINES using en-bloc sending, the time interval starts when the SETUP message has been received from the digital subscriber signalling system.

The values in Table 10 are recommended.

TABLE 10/Q.543

	Reference load A	Reference load B
Mean value	≤ 600 ms	≤ 800 ms
0.95 probability of not exceeding	800 ms	1200 ms

2.3.4 through-connection delay: Through-connection delay is defined as the interval from the instant at which the information required for setting up a through-connection is available for processing in an exchange, or the signalling information required for setting up a through-connection is received from the signalling system, to the instant at which the appropriate transmission path is available for carrying traffic between the incoming and outgoing exchange terminations.

The exchange through-connection delay does not include an inter-office continuity check, if provided, but does include a cross-office check if one occurs during the defined interval.

When the through-connection is established during call set-up, the recommended values for exchange call set-up delay apply. When the through-connection in an exchange is not established during the exchange call set-up interval, the through-connection delay may then contribute to the network call set-up delay.

2.3.4.1 For transit and originating outgoing traffic connections

The values in Table 11 are recommended.

The requirements for multi-slot connections require further study.

TABLE 11/Q.543

	Reference load A		Reference load B	
	Without ancillary equipment	With ancillary equipment	Without ancillary equipment	With ancillary equipment
Mean value	≤ 250 ms	≤ 350 ms	≤ 400 ms	≤ 500 ms
0.95% probability of not exceeding	300 ms	500 ms	600 ms	800 ms

2.3.4.2 For internal and terminating traffic connections

For connections terminating on ANALOGUE SUBSCRIBER LINES, the through-connection delay is the interval from the instant at which the called subscriber off-hook condition (answer) is recognizable at the subscriber line interface of the exchange until the through-connection is established and available for the carrying traffic or a consequent signal is sent backwards by the exchange.

The maximum values applying to this parameter are included with those for incoming call indication sending delay in 2.3.5.

For connections terminating on DIGITAL SUBSCRIBER LINES, the through-connection delay is the interval from the instant at which the CONNECT message is received from the signalling system until the through-connection is established and available for carrying traffic as those indicated by passing to the respective signalling systems of the ANSWER and CONNECT ACKNOWLEDGE messages.

The values in Table 12 are recommended.

TABLE 12/Q.543

	Reference load A	Reference load B
Mean value	≤ 250 ms	≤ 400 ms
0.95 probability of not exceeding	300 ms	600 ms

2.3.5 incoming call indication sending delay – (for terminating and internal traffic connections)

2.3.5.1 For calls terminating on ANALOGUE SUBSCRIBER LINES, the incoming call indication sending delay is defined as the interval from the instant when the last digit of the called number is available for processing in the exchange until the instant that ringing signal is applied by the exchange to the called subscriber line.

It is recommended that the sum of the values for ringing signal sending delay and through-connection delay for internal and teminating traffic connection should not exceed the values in Table 13. In addition, it is recommended that the value of the incoming call indication sending delay should not exceed 90% of these values nor the through-connection delay exceed 35% of these values.

TABLE 13/Q.543

	Reference load A	Reference load B
Mean value	≤ 650 ms	≤ 1000 ms
0.95 probability of not exceeding	900 ms	1600 ms
NOTE – The above values assume that "immediate" ringing is applied and do not include delays caused by functions such as line teste, which may be used in national networks.		

2.3.5.2 For calls terminating on DIGITAL SUBSCRIBER LINES, the incoming call indication sending delay is defined as the interval from the instant at which the necessary signalling information is received from the signalling system to the instant at which the SETUP message is passed to the signalling system of the called digital subscriber line.

In the case of overlap sending in the incoming signalling system, the values in Table 14 are recommended.

TABLE 14/Q.543

	Reference load A	Reference load B
Mean value	≤ 400 ms	≤ 600 ms
0.95 probability of not exceeding	600 ms	1000 ms

In the case of en-bloc sending in the incoming signalling system, the values in Table 15 are recommended.

TABLE 15/Q.543

	Reference load A	Reference load B
Mean value	≤ 600 ms	≤ 800 ms
0.95 probability of not exceeding	800 ms	1200 ms

2.3.6 Alerting sending delay – terminating and internal traffic connections

2.3.6.1 alerting sending delay for terminating traffic

2.3.6.1.1 For calls terminating on ANALOGUE SUBSCRIBER LINES, alerting sending delay is defined as the interval from the instant when the last digit is available for processing in the exchange until the ringing tone is sent backwards toward the calling user.

The values in Table 13 are recommended.

2.3.6.1.2 For calls termining on DIGITAL SUBSCRIBER LINES, the alerting sending delay is defined as the interval from the instant that an ALERTING message is received from the digital subscriber line signalling system to the instant at which an ADDRESS COMPLETE message is passed to the interexchange signalling system or ringing tone is sent backward toward the calling user.

The values in Table 16 are recommended.

TABLE 16/Q.543

	Reference load A	Reference load B
Mean value	≤ 200 ms	≤ 350 ms
0.95 probability of not exceeding	400 ms	700 ms

2.3.6.2 alerting sending delay for internal traffic

2.3.6.2.1 For calls terminating on ANALOGUE SUBSCRIBER LINES, alerting sending delay is defined as the interval from the instant that the signalling information is available for processing in the exchange until ringing tone is applied to an ANALOGUE calling subscriber line or an ALERTING message is sent to a DIGITAL calling subscriber line signalling system.

For calls from ANALOGUE SUBSCRIBER LINES to ANALOGUE SUBSCRIBER LINES, the values in Table 13 are recommended.

For calls from DIGITAL SUBSCRIBER LINES to ANALOGUE SUBSCRIBER LINES, the values in Table 17 are recommended.

TABLE 17/Q.543

	Reference load A	Reference load B
Mean value	≤ 300 ms	≤ 500 ms
0.95 probability of not exceeding	500 ms	800 ms

2.3.6.2.2 For internal calls terminating on DIGITAL SUBSCRIBER LINES originating from ANALOGUE SUBSCRIBER LINES, alerting sending delay is defined as the interval from the instant that an alerting message is received from the signalling system of the called subscriber's line until ringing tone is applied to the calling subscriber line.

The values in Table 13 are recommended.

Alerting sending delay on internal calls between DIGITAL SUBSCRIBER LINES are covered by Table 28.

2.3.7 **ringing tripping delay – internal and terminating traffic connections**: Ringing tripping delay is a characteristic that is applicable for calls terminating on ANALOGUE SUBSCRIBER LINES only. It is defined as the interval from the instant that the called subscriber off-hook condition is recognizable at the subscriber line interface until the ringing signal at the same interface is suppressed.

The values in Table 18 are recommended.

TABLE 18/Q.543

	Reference load A	Reference load B
Mean value	≤ 100 ms	≤ 150 ms
0.95 probability of not exceeding	150 ms	200 ms

2.3.8 **exchange call release delay**: Exchange call release delay is the interval from the instant at which the last information required for releasing a connection is available for processing in the exchange to the instant that the switching network through-connection in the exchange is no longer available for carrying traffic and the disconnection signal is sent to the subsequent exchange, if applicable. This interval does not include the time taken to detect the release signal, which might become significant during certain failure conditions, e.g., transmission system failures.

2.3.8.1 For transit traffic connections involving circuits using channel associated signalling or a mix of channel associated and common channel signalling, the values in Table 19 are recommended.

For transit traffic connections involving circuits using CCITT Signalling System No. 7 signalling exclusively, the values in Table 35 are recommended.

TABLE 19/Q.543

	Reference load A	Reference load B
Mean value	≤ 250 ms	≤ 400 ms
0.95 probability of not exceeding	300 ms	700 ms

2.3.8.2 For originating, terminating and internal traffic connections, the values in Table 20 are recommended.

TABLE 20/Q.543

	Reference load A	Reference load B
Mean value	≤ 250 ms	≤ 400 ms
0.95 probability of not exceeding	300 ms	700 ms

2.3.9 exchange signalling transfer delay – other than answer signal: Exchange signalling transfer delay is the time taken by the exchange to transfer a signal, no other exchange action being required. It is defined as the interval from the instant that the incoming signal is recognizable, or the signalling information is received from the signalling system, until the instant when the corresponding outgoing signal has been transmitted, or the appropriate signalling information is passed to the signalling system.

2.3.9.1 For transit traffic connections involving circuits using channel associated signalling or a mix of channel associated and common channel signalling, the values in Table 21 are recommended.

For transit traffic connections between circuits that use CCITT Signalling System No. 7 signalling exclusively, the requirements of the appropriate signalling system Recommendations should apply, e.g., Recommendations Q.725/Q.726 for T_{cu} value (case of a simple message).

TABLE 21/Q.543

	Reference load A	Reference load B
Mean value	≤ 100 ms	≤ 150 ms
0.95 probability of not exceeding	150 ms	300 ms

2.3.9.2 Exchange signalling transfer delay for originating, terminating and internal traffic involving a mix of ANALOGUE and DIGITAL SUBSCRIBER LINES is left for further study. Exchange signal transfer delay between DIGITAL SUBSCRIBER signalling systems or between DIGITAL SUBSCRIBER LINE signalling systems and CCITT Signalling System No. 7 is covered in 2.4.2.

2.3.10 answer sending delay: Answer sending delay is defined as the interval from the instant that the answer indication is received at the exchange to the instant that the answer indication is passed on by the exchange toward the calling user. The objective of this parameter is to minimize the possible interruption of the transmission path for any significant interval during the initial response by the called user.

2.3.10.1 For transit traffic involving circuits that use channel associated signalling or a mix of channel associated and common channel signalling, the values in Table 22 are recommended.

TABLE 22/Q.543

	Reference load A	Reference load B
Mean value	≤ 100 ms	≤ 150 ms
0.95 probability of not exceeding	150 ms	300 ms

More stringent values are recommended where in-band line signalling may be encountered in the national part of a built-up connection. The recommended values are given in Table 23.

TABLE 23/Q.543

	Reference load A	Reference load B
Mean value	≤ 50 ms	≤ 90 ms
0.95 probability of not exceeding	100 ms	180 ms

For transit traffic connections involving circuits that use CCITT Signalling System No. 7 exclusively, the requirements of the appropriate signalling system Recommendations should apply, e.g., CCITT Recommendations Q.725 and Q.766 for T_{cu} value (case of a simple message).

2.3.10.2 For connections in a terminating exchange, exchange answer sending delay is defined as the interval from the instant that the off-hook condition is recognizable at the ANALOGUE SUBSCRIBER LINE interface on an incoming call or a CONNECT message is received from a DIGITAL SUBSCRIBER LINE signalling system until the instant that an answer indication is sent back toward the calling user.

The values in Table 24 are recommended.

TABLE 24/Q.543

	Reference load A	Reference load B
Mean value	≤ 250 ms	≤ 350 ms
0.95 probability of not exceeding	300 ms	700 ms

2.3.10.3 For connections in an originating exchange, exchange answer sending delay is defined as the interval from the instant that the answer indication is received from the outgoing circuit signalling system or in the case of an internal call, from the called subscriber's line, until the instant that the answer indication is sent to the calling user. In the case of a call originated from a DIGITAL SUBSCRIBER LINE, the answer indication is a CONNECT message that is sent to the DIGITAL SUBSCRIBER LINE signalling system. If an ANALOGUE SUBSCRIBER LINE originated the call, the answer indication may not be sent.

The values in Table 25 are recommended.

TABLE 25/Q.543

	Reference load A	Reference load B
Mean value	≤ 250 ms	≤ 400 ms
0.95 probability of not exceeding	300 ms	700 ms

For ISDN operation involving DIGITAL SUBSCRIBER LINES and CCITT Signalling System No. 7 exclusively, the values in Table 28 are recommended.

2.3.11 timing for start of charging (circuit switched calls): When required, timing for charging at the exchange where this function is performed, shall begin after receipt of an ANSWER indication from a connecting exchange or the called user. The start of timing for charging should occur within the intervals recommended in Table 26.

TABLE 26/Q.543

	Reference load A	Reference load B
Mean value	≤ 100 ms	≤ 175 ms
0.95 probability of not exceeding	200 ms	350 ms

2.4 Delay probability – ISDN environment

The following notes apply to the delay parameters included in this subclause:

1) The term "mean value" is understood as the expected value in the probabilistic sense.

2) Where several messages are received at the exchange from a digital subscriber line signalling system (e.g. several alert messages are received from a multi-user configuration), the message that is accepted for call handling is the one considered in determining the start of a given delay interval.

3) The terms "received from" and "passed to" the signalling system are used. For CCITT Signalling System No. 7 this is designated as the instant the information is exchanged between the signalling data link (layer 1) and the signalling link functions (layer 2). For digital subscriber line signalling, this is designated as the instant the information is exchanged by means of primitives between the data link layer (layer 2) and the network layer (layer 3). Thus, the time intervals exclude the above layer 1 (CCITT Signalling System No. 7) and layer 2 (D channel) times. They do, however, include queuing delays that occur in the absence of disturbances but not any queuing delays caused by re-transmission.

2.4.1 user signalling acknowledgement delay: User signalling acknowledgement delay is the interval from the instant a user signalling message has been received from the subscriber line signalling system until a message acknowledging the receipt of that message is passed back from the exchange to the user line signalling system. Examples of such messages are SETUP ACKNOWLEDGEMENT to SETUP, CONNECT ACKNOWLEDGEMENT to CONNECT and RELEASE ACKNOWLEDGEMENT to RELEASE.

The values in Table 27 are recommended.

TABLE 27/Q.543

	Reference load A	Reference load B
Mean value	≤ 400 ms	≤ 800 ms
0.95 probability of not exceeding	600 ms	1000 ms

2.4.2 signalling transfer delay: The exchange signalling transfer delay is the time taken for the exchange to transfer a message from one signalling system to another with minimal or no other exchange actions required. The interval is measured from the instant that a message is received from a signalling system until the moment the corresponding message is passed to another signalling system. Examples of messages are ALERT to ADDRESS COMPLETE, ADDRESS COMPLETE to ADDRESS COMPLETE, CONNECT to ANSWER, RELEASE to DISCONNECT, etc.

The values in Table 28 are recommended for originating and terminating connections.

TABLE 28/Q.543

	Reference load A	Reference load B
Mean value	≤ 200 ms	≤ 350 ms
0.95 probability of not exceeding	400 ms	700 ms

For transit connections, the requirements of the appropriate signalling system Recommendation should apply, e.g. Recommendations Q.725 and Q.766 for T_{cu} value (case of a simple message).

NOTE – User-to-user signalling may imply additional functions in the exchanges, e.g. charging, flow control, etc. The requirements for user-to-user signalling transfer delay and the impact of user-to-user signalling on exchange performance is for further study.

2.4.3 call set up delay: Call set-up delay is defined as the interval from the instant when the signalling information required for outgoing circuit selection is received from the incoming signalling system until the instant when the corresponding signalling information is passed to the outgoing signalling system.

2.4.3.1 For originating 64 kbit/s circuit switched connections [(types I, II and III option a)].

i) If overlap sending is used, the interval starts when the information message received contains a "sending complete" indication or the address information for call set up is complete.

ii) If en-bloc sending is used, the time interval starts when the SETUP message has been received from the user signalling system.

For call attempts using overlap sending, the values in Table 29 are recommended.

TABLE 29/Q.543

	Reference load A	Reference load B
Mean value	≤ 400 ms	≤ 600 ms
0.95 probability of not exceeding	600 ms	1000 ms

For call attempts using en-bloc sending, the values in Table 30 are recommended.

TABLE 30/Q.543

	Reference load A	Reference load B
Mean value	≤ 600 ms	≤ 800 ms
0.95 probability of not exceeding	800 ms	1200 ms

2.4.3.2 For originating supplementary service call attempts:

For further study.

2.4.3.3 For transit 64 kbit/s circuit switched connections between circuits that use CCITT Signalling System No. 7, the requirements of Recommendations Q.725 and Q.766 should apply for T_{cu} value (case of a processing intensive message).

2.4.4 through connection delay

2.4.4.1 For originating outgoing and transit traffic 64 kbit/s switched circuit connections, through connection delay is defined as the interval from the instant that the signalling information required for setting up a connection through the exchange is received from the incoming signalling system to the instant that the transmission path is available for carrying traffic between the incoming and outgoing terminations on the exchange.

Usually, both directions of transmission will be switched through at the same time. However, at an originating exchange, on certain calls, there may be a requirement to effect switch through in two stages, one direction at a time. In this case, different signalling messages will initiate the two stages of switch through and the recommended delay applies to each stage of switch through.

The values in Table 31 are recommended.

TABLE 31/Q.543

	Reference load A		Reference load B	
	Without ancillary function	With ancillary function	Without ancillary function	With ancillary function
Mean value	≤ 250 ms	≤ 350 ms	≤ 400 ms	≤ 500 ms
0.95% probability of not exceeding	300 ms	500 ms	600 ms	800 ms

2.4.4.2 For internal and terminating traffic 64 kbit/s switched circuit connections the through connection delay is defined as the interval from the instant that the CONNECT message is received from the called line signalling system until the through connection is established and available for carrying traffic and the ANSWER and CONNECT ACKNOWLEDGEMENT messages have been passed to the appropriate signalling systems.

The values in Table 32 are recommended.

TABLE 32/Q.543

	Reference load A	Reference load B
Mean value	≤ 250 ms	≤ 400 ms
0.95 probability of not exceeding	300 ms	600 ms

2.4.5 incoming call indication sending delay – (for terminating and internal traffic connections): The incoming call indication sending delay is defined as the interval from the instant at which the necessary signalling information is received from the signalling system to the instant at which the SETUP message is passed to the signalling system of the called subscriber line.

In the case of overlap sending in the incoming signalling system, the values in Table 33 are recommended.

TABLE 33/Q.543

	Reference load A	Reference load B
Mean value	≤ 400 ms	≤ 600 ms
0.95 probability of not exceeding	600 ms	1000 ms

In the case of en-bloc sending in the incoming signalling system, the values in Table 34 are recommended.

TABLE 34/Q.543

	Reference load A	Reference load B
Mean value	≤ 600 ms	≤ 800 ms
0.95 probability of not exceeding	800 ms	1200 ms

2.4.6 connection release delay: Connection release delay is defined as the interval from the instant when DISCONNECT or RELEASE message is received from a signalling system until the instant when the connection is no longer available for use on the call (and is available for use on another call) and a corresponding RELEASE or DISCONNECT message is passed to the other signalling system involved in the connection.

The values in Table 35 are recommended.

TABLE 35/Q.543

	Reference load A	Reference load B
Mean value	≤ 250 ms	≤ 400 ms
0.95 probability of not exceeding	300 ms	700 ms

2.4.7 Call clearing delay

Disconnect and call clearing will usually be performed at the same time. However, on certain calls it may be necessary for an exchange to retain call references after disconnect has occurred, until a clearing message is received. The exchange may then discard the call reference information. The corresponding RELEASE message must be passed on to other involved signalling systems in the interval allowed for signalling transfer delay (see 2.4.2).

2.4.8 Timing for start of charging (circuit switched calls)

When required, timing for charging at the exchange where this function is performed, shall begin after receipt of an ANSWER indication from a connecting exchange or the called user. The start of timing for charging should occur within the intervals recommended in Table 36.

TABLE 36/Q.543

	Reference load A	Reference load B
Mean value	≤ 100 ms	≤ 175 ms
0.95 probability of not exceeding	200 ms	350 ms

2.5 Call processing performance objectives

2.5.1 64 kbit/s switched connections

2.5.1.1 Premature release

The probability that an exchange malfunction will result in the premature release of an established connection in any one minute interval should be:

$$P \leq 2 \times 10^{-5}$$

2.5.1.2 Release failure

The probability that an exchange malfunction will prevent the required release of a connection should be:

$$P \leq 2 \times 10^{-5}$$

2.5.1.3 Incorrect charging or accounting

The probability of a call attempt receiving incorrect charging or accounting treatment due to an exchange malfunction should be:

$$P \leq 10^{-4}$$

2.5.1.4 Misrouting

The probability of a call attempt misrouted following receipt by the exchange of a valid address should be:

$$P \leq 10^{-4}$$

2.5.1.5 No tone

The probability of a call attempt encountering no tone following receipt of a valid address by the exchange should be:

$$P \leq 10^{-4}$$

2.5.1.6 Other failures

The probability of the exchange causing a call failure for any other reason not identified specifically above should be:

$$P \leq 10^{-4}$$

2.5.2 64 kbit/s semi-permanent connections

This requires further study taking into consideration:

- need to recognize an interruption;
- probability of an interruption;
- requirements for re-establishment of interrupted connection;
- any other unique requirements.

2.5.3 n × 64 kbit/s switched connections

To be recommended if/when specific services are defined.

2.5.4 n × 64 kbit/s semi-permanent connections

To be recommended if/when specific services are defined.

2.6 Transmission performance

2.6.1 64 kbit/s switched connections

The probability of a connection being established with an unacceptable transmission quality across the exchange should be:

$$P \text{ (Unacceptable transmission)} \leq 10^{-5}$$

The transmission quality across the exchange is said to be unacceptable when the bit error ratio is above the alarm condition.

NOTE – The alarm condition has yet to be defined.

2.6.2 64 kbit/s semi-permanent connections

To be recommended.

2.6.3 n × 64 kbit/s switched connections

To be recommended, if/when specific services are defined.

2.6.4 n × 64 kbit/s semi-permanent connections

To be recommended if/when specific services are defined.

2.7 Slip rate

2.7.1 Normal conditions

The slip rate under normal conditions is covered in Recommendation Q.541.

2.7.2 Temporary loss of timing control

The case of temporary loss of timing control corresponds to the "holdover operation" defined and recommended in Recommendation G.812. The allowable slip rate will correspond to the maximum relative TIE also recommended therein.

2.7.3 Abnormal conditions at the exchange input

The slip rate in case of abnormal conditions (wide phase deviations, etc.) at the exchange input is the subject of further study taking into account the requirements of Recommendation G.823.

3 Exchange performance during overload conditions

This clause applies to digital exchanges operating during periods when the number of call attempts presented to the exchange exceeds its call processing capacity for a significant period of time, excluding momentary peaks. Under these conditions the exchange is said to be operating in an overload condition.

This Recommendation identifies requirements for exchange performance during overload and for overload mechanisms in the exchange. Network management functions to be supported by an exchange are defined in 5/Q.542.

3.1 Explanation of terms used in definition of overload parameters

- **load**: The total number of call attempts presented to an exchange during a given interval of time (i.e. offered load).
- **overload**: That part of the total load offered to an exchange, in excess of the engineered traffic processing capacity of the exchange. Overload is usually expressed as a percentage of engineered capacity.
- **throughput**: The number of call attempts processed successfully by an exchange per unit time.
- **engineered capacity**: The mean offered load at which the exchange just meets all grade of service requirements used by the Administration to engineer the exchange.

3.2 Call processing performance during overload

An exchange must continue to process a specified load even when the offered call attempts exceed its available call processing capacity. The number of call attempts handled during an overload condition should not be significantly lower than the engineered capacity of the exchange for a specified Grade Of Service (GOS), as noted in 3.7.

Two basic requirements for exchange performance during overload are:

- to maintain adequate exchange throughput in sustained overload;
- to react sufficiently quickly to load peaks and the sudden onset of overload.

As the offered load increases beyond the engineered attempt capacity of the exchange, the throughput or the carried attempt load may exhibit a behaviour shown by curve A in Figure 1, i.e. processor throughput may be reduced drastically if the offered load increases well beyond the engineered load. Curve B in Figure 1 represents the maximum throughput, where the throughput remains at the nominal design level under overload. Appropriate overload protection mechanisms should be included in the overall exchange design so that the throughput performance of the processor under overload resembles the curve C in Figure 1.

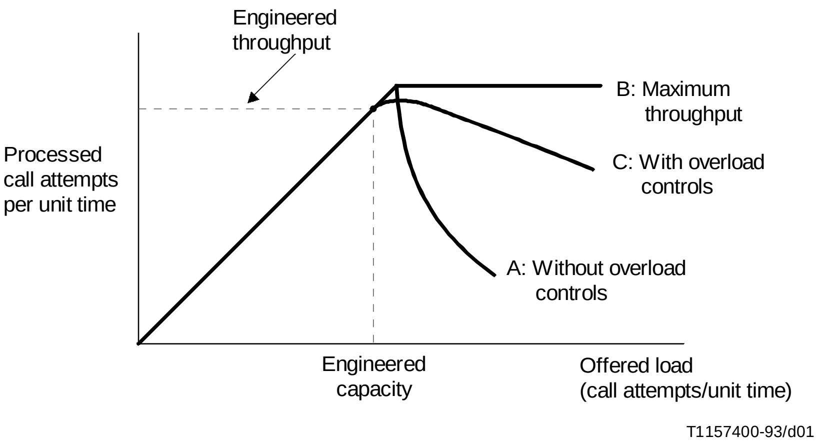


FIGURE 1/Q.543

Throughput characteristics

3.3 Engineered exchange capacity

Exchange engineered capacity is the maximum load that the exchange can handle while operating in a "normal" mode (i.e. performing all required operating and administrative functions) while meeting performance requirements specified in clause 2 or those specified by the Administration. It is not necessarily the point of maximum throughput (see Figure 1).

Overload controls, when applied, may have a significant effect on exchange capacity. Overload throughput performance should be specified relative to the engineered capacity of the exchange when overload controls are operating.

3.4 Overload control strategy

An effective overload control strategy will prevent the rapid decrease in processed call attempts with increasing overload (see Curve A in Figure 1); the relatively gradual decrease with overload controls enabled (Curve C in Figure 1) is due to the increasing processing overhead in exercising the overload controls.

Overload is defined as the level of call attempts offered to the exchange in excess of the exchange engineered capacity. For example, when the exchange is offered call attempts at a rate of 10% greater than the engineered capacity, the exchange is said to have 10% overload.

The exchange throughput at an overload of Y% above the engineered capacity load should be at least X% of the throughput at engineered capacity. This concept is shown in Figure 2 which shows the region of unacceptable throughput performance. Any throughput curve which remains above the X% level until reaching the point of Y% overload is acceptable. The recommended values are Y = 50% and X = 90%. Beyond Y% overload the exchange should continue to process calls in an acceptable manner.

As long as the level of overload does not exceed Y% above the exchange engineered capacity, then the exchange throughput should be no less than X% of engineered capacity, as depicted in Figure 2.

Measurements that can provide data as the basis for calculation of X and Y, are identified in 3.8.

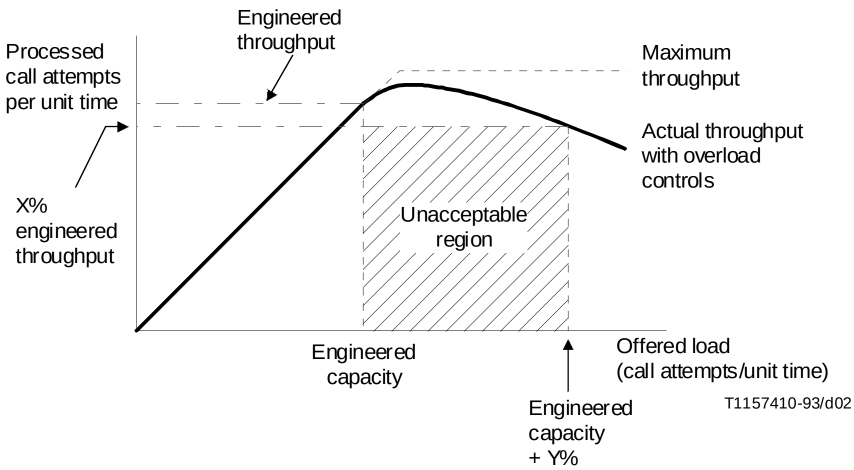


FIGURE 2/Q.543

Throughput performance with overload control activated

3.5 Detection of overload

The exchange should incorporate suitable means for detecting overload conditions.

The onset of an overload state should be recognized by the exchange processing logic which in turn will invoke strategies to avoid a severe degradation in throughput load. During overload, both severe delays and processing delays will increase and will normally exceed the performance objectives given for Reference load B.

Overload indications may, for example, be provided by: a continuous measurement of the occupancy of the resources used for call handling over short periods (e.g. a few seconds); monitoring the queue lengths for the various call handling processes, etc. Overload control activation indications should be given to the administration staff.

3.6 Overload protection

The internal overload control methods used in an exchange are dependant on the particular technical arrangement of the switching system, and are not subject to CCITT Recommendations. Overload controls used in conjunction with adjacent exchanges are discussed under “Network management design objectives” in clause 5/Q.542.

In order to reduce the load on the exchange caused by calls that cannot be processed during overload, it may be necessary to discourage further attempts by customers during this situation. Methods used to achieve this reduction should not significantly increase the load on exchange processors, as for example, routing calls to recorded announcements.

Overload controls, once applied, should be removed as quickly as possible when the degree of overload reduces, consistent with the need to avoid oscillatory behaviour which might prolong the period of degraded service.

As a guideline to providing service during overload conditions, the following general principles are applicable:

- give preference to the processing of terminating calls;

- give preference to priority class lines, calls to priority destinations based on digit analysis and incoming calls with priority indications in, for example, the Initial Address Message of a call using CCITT Signalling System No. 7, if an essential service protection capability has been invoked;

- defer some or all activities non-essential to handling offered traffic; examples are some administration and maintenance processes in the exchange. (Nevertheless the man-machine communications essential for priority operational tasks should always be preserved. In particular, network management terminals and functions associated with interfaces to network management support systems should be afforded high priority, since network management actions can play an important role in reducing exchange overloads);

- maintain normal charging and supervisory functions, and established connections until the receipt of the appropriate release signal;

- assign priorities to specific exchange measurements, such that low priority measurements cease at a predetermined level of congestion. Higher priority measurements may be ceased at a higher level of congestion, or may be run continuously, depending on their importance to the call handling functions;

- give preference to calls already being processed, before accepting new calls.

3.7 Grade of service during overload

In general the overall grade of service seen by the subscribers will deteriorate when the exchange experiences severe overload conditions and the overload protection mechanisms have been invoked. This may be due to the fact that the overload protection procedures may require that the exchange not accept all the call attempts offered.

Accepted calls may or may not receive a grade of service equal to that received by calls at Reference load B of clause 2. In terms of the exchange overload performance, it is sufficient that calls be accepted in such a way that throughput is maximized.

3.8 Performance monitoring during overload control activation

The operational measurements in the exchange should be sufficient to determine the number of call attempts accepted by the exchange, and the number that are successfully being completed, from the exchange point-of-view. Separate measurements should be available to count the number of attempts rejected by the exchange during overload, so that the total load can be estimated.

An accepted call attempt is defined to be a call attempt which is accepted for processing by the exchange. This does not necessarily mean that an accepted call attempt will complete or receive an acceptable grade of service.

The call completion rate can vary statistically with time, according to the specific call attempt acceptance process invoked by the overload controls. Therefore the call completion rate estimated from the operational measurements needs to be taken over a sufficiently long period of time to verify conformance to the X% throughput requirement.

Annex A

An example of methodology for computing the call processing capacity of a Digital Exchange, taking into account ISDN services, including packet data handling

(This annex forms an integral part of this Recommendation)

A.1 General

Exchanges will generally be required to handle many types of calls as they provide basic telephony service, supplementary telephony service, ISDN bearer service and ISDN supplementary services. A variety of signalling types will be used on subscriber lines and for handling calls over interexchange circuits. Performance objectives have been recommended and are applicable over the full range of exchange sizes and loads up to the limit of exchange "engineered" capabity at its maximum size for the mix of call types handled and signalling types used in the exchange. Different mixes of call types and signalling types require different amounts of processing capacity. Thus the maximum number of subscriber lines that can be served and the number of calls that can be handled will be different for each mix on the same switching system. This annex serves as an example of a methodology that makes it possible to compute the processing capacity of an exchange for any particular mix of call types and signalling expected to be encountered in its implementation. Of course, other possible limiting factors such as allowable hardware configuration, memory capacity, etc., must also be taken into account when determining the capacity of the exchange.

The method of calculating call processing capacity illustrated herein is for a particular multi-processor exchange design shown in Figure A.1. However, the principles used can be applied to any processor controlled exchange design for any mix of services, traffic and signalling handled by the exchange. This method requires that manufacturers provide information and data about their exchange designs in terms that Administrations can use in the formulae derived below and that Administrations make measurements and/or estimates to forecast the expected traffic volumes and mix of services, call types and signalling.

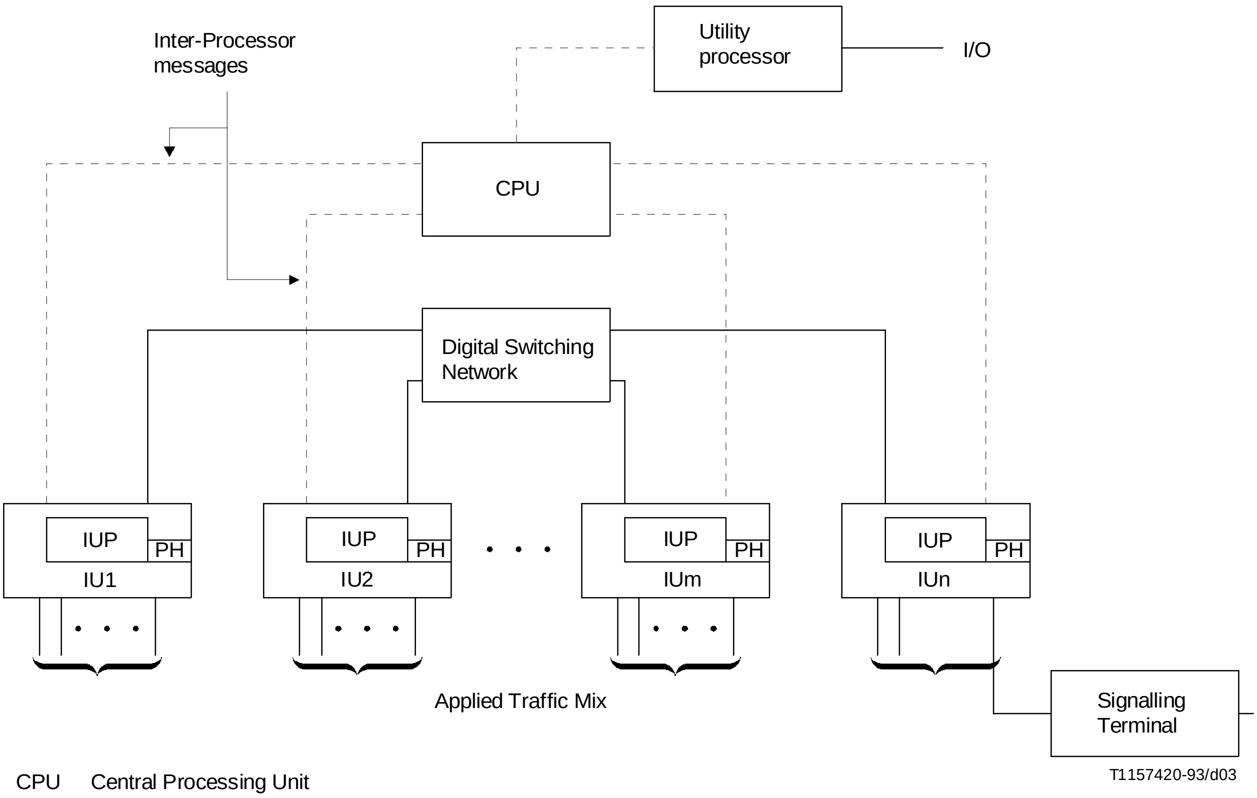
It is important to examine the exchange architecture and to understand how calls are processed in order to recognize potential limiting elements. For example, ISDN calls involving packet switching will have two separate elements to be considered, call set-up and packet handling. Packet call set up can be dealt with in the same manner as circuit switched call set-up by considering these types of call attempts in and with the circuit switched call attempt originations and dispositions. However, subsequent packet handling requires continuing processing capacity, occasionally for long periods of time, may be handled by processors other than those involved in call set-up and thus, must be dealt with separately.

Figure A.1 shows a block diagram of an exchange design with several processors, which is used as an example in this annex.

a) The Interface Unit 1 through n provide interfaces to user lines, interexchange circuits, signalling terminals and any other interfaces to entities outside the exchange. A certain amount of call processing (e.g. handling signalling to or from lines or interexchange circuits, digit analysis, etc.) can be performed by processors in these interface units. In this example, each Interface Unit also contains its own packet handler (shown as PH). The Interface Units communicate with a Central Processing Unit over high capacity inter-processor lines.

b) The Central Processing Unit directs call processing by the exchange. It receives information about call attempts from the Interface Units, determines how they should be handled and routed and directs their disposition by the appropriate Interface Units. In connection with packet switching calls, it is assumed that the Central Processing Unit is involved only in call set-up and call release and that ongoing packet handling requires no significant amount of CPU processing capacity. The CPU also performs other call related and administrative tasks, such as maintaining charging information, and performs other administrative and operation functions for the exchange.

To determine the capacity of this design it is necessary to know how many Interface Units can be connected to an exchange. Then it is necessary to compute the call processing capacity of the Central Processing Unit and the capacity of the Interface Units to determine which is the limiting factor. In some designs, other elements, such as a utility processor or the switching network, can limit the size of the exchange. Thus, it is necessary to understand the exchange design and then to make appropriate computations involving the limiting elements to determine the processing capacity of the exchange for the traffic mix envisioned.



CPU Central Processing Unit
IU Interface Unit
IUP Interface Unit Processor
PH Packet Handler

FIGURE A.1/Q.543

Example of an exchange design with several processors

A.2 Definitions

A.2.1 capacity unit: The processing capacity required in an exchange (or processing unit) to process a call attempt consisting of the originating portion plus the terminating (or disposition) portion.

A.2.2 half unit: The processing capacity required to process either the originating or terminating (disposition) portion of a call attempt handled by an exchange or a processing unit, e.g. an Interface Unit in the exchange design shown.

A.2.3 originating type: A type of call attempt entering the exchange (e.g. a telephone call from a line class-marked for basic telephone service, or one from a line marked for supplementary services, or basic ISDN services, or ISDN supplementary services, or a call entering the exchange on an incoming interexchange circuit, etc.).

A.2.4 terminating (disposition) type: A type of call attempt leaving or disposed of by the exchange (e.g. a call attempt terminating to a line class marked for basic telephone service, or one to a line with supplementary or ISDN services assigned, or to an outgoing interexchange circuit, etc.).

A.2.5 reference capacity unit: The processing capacity required for processing an arbitrarily selected pair of half units, one an originating type attempt and one a terminating (disposition) type attempt, usually a pair that is expected to be involved in a significant portion of the traffic load in the exchange. The reference capacity unit uses a standard against which capacity units for other types of attempts are compared. (It is suggested that an originating outgoing "local" telephone call attempt from a basic telephone line and disposed of by routing it to an interexchange circuit using CCITT Signalling System No. 7 as the reference capacity unit.)

A.2.6 reference capacity half-unit: The processing capacity required in an interface unit to process an arbitrarily selected half-unit, either an originating or a terminating (disposition) type (usually one that is involved in a significant portion of traffic that interface units handle, e.g. an originating telephone call attempt from a basic telephone line). The reference capacity half-unit is used as the standard against which half-units of other types of attempts are compared. When separate calculations for different interface units are necessary, which occurs when different mixes of line classes and traffic are served by the different interface units, the same reference capacity half-unit should be used for all calculations.

A.2.7 central processor unit (CPU) reference capacity unit: The processing capacity required in the CPU to process the portions of attempts associated with one reference capacity unit. The reference capacity unit is assigned unit value. Thus, if F is the fraction of one reference capacity unit for processing the originating portion and F' is the fraction of one reference capacity unit required for processing the terminating (disposition) portion, the sum is unity ($F + F' = 1$).

A.2.8 interface unit (IU) reference capacity unit: The amount of processing capacity required in the IU in the exchange design shown, to properly handle one reference capacity half-unit.

A.2.9 weighting factor: The ratio of the relative amount of processing capacity required to handle either portion, originating or terminating (disposition), of any attempt type, to the capacity required in that processor to perform the same functions for reference capacity unit, [originating and terminating (disposition) portions]. For example, if a complete reference capacity unit requires 1000 processor cycles in the CPU and the originating portion of a call attempt entering the exchange requires 430 cycles in the CPU, the weighting factor (CPU) for that originating attempt type would be 0.43.

Similarly, in the interface unit, a weighting factor is the ratio of the amount of IU processing capacity required to handle a particular half-unit to the amount of IU processing capacity required to handle a reference capacity half-unit. Thus if an IU requires 600 cycles to handle a reference capacity half-unit and another type of call entering the exchange via the IU requires 725 IU processor cycles, the weighting factor (IU) for that half-unit attempt type would be 1.21.

Weighting factors for all originating and terminating (disposition) types of capacity units and half-units, are required for each processing unit in the exchange in order to make capacity computations. These weighting factors must be furnished by the manufacturer.

A.2.10 reference unit (and half-unit) processing capacity (RUPC): Is capacity information that should be furnished by the manufacturer. RUPC is the total number of reference capacity units (and half-units) that can be performed by a processor (or processing unit) in one hour in an exchange while meeting performance criteria specified by the Administration and at the same time performing all the operations and administrative tasks required for normal operation of the exchange. Thus, RUPC is the processing capacity available for call handling. It is the total installed capacity diminished by an amount required for overhead, administrative tasks, etc. In addition to accounting for the overhead of administrative tasks, it may also be desirable to "reserve" a certain percentage of capacity for program growth additions that would be needed in a maximum size exchange for adding new features in the future. To be able to make a realistic comparison of different systems, it is necessary that the Administration learn from the manufacturers, the non-call handling functions that are accounted for and the percent of capacity that is being reserved for growth.

A.3 Processing capacity computation (for a central processing unit)

Capacity information and weighting factors are furnished by the manufacturer.

Let F_i = weighting factor for originating type i

F'_j = weighting factor for terminating (disposition) type j.

Traffic mix on the CPU is specified by the Administration.

Let P_i = fraction of call attempts expected to be originating type i

P'_j = fraction of call attempts expected to be terminating (disposition) type j.

where

$$\sum_{i=1}^{n} P_i = 1.0$$

and

$$\sum_{j=1}^{m} P'_j = 1.0$$

If, R = the call attempt rate expressed in terms of busy hour call attempts, then the amount of processing capacity required for originating type work units associated with the i-th call attempt type traffic is:

$$P_i F_i R$$

Similarly, the processing capacity required for disposition work associated with the j-th call type traffic is:

$$P'_j F'_j R$$

In order to satisfy the performance design objectives in this Recommendation, the reference unit processing capacity (RUPC) must be equal to or greater than the total originating type work plus the total terminating (disposition) type work:

$$RUPC\,(CPU) \geq \left[\sum_{i=1}^{n} P_i\ F_i + \sum_{j=1}^{m} P'_j\ F'_j \right] R$$

From which:

$$R\ (\text{maximum}) = \frac{RUPC\,(CPU)}{\sum_{i=1}^{n} P_i\ F_i + \sum_{j=1}^{m} P'_j\ F'_j}$$

A.4 Processing capacity computation (for an interface unit)

Capacity information and weighting factors are furnished by the manufacturer.

Let H_i = weighting factor for half-unit type i.

Traffic mix on the interface unit is specified by the Administration.

Let P_i = fraction of attempts to be half-unit type i.

where

$$\sum_{i=1}^{n} P_i = 1.0$$

If, R = the attempt rate in terms of busy hour half-units, the processing capacity required for i-th type half-units is:

$$P_i H_i R$$

In order to satisfy performance criteria, the reference unit call processing capacity (RUPC) must be equal to or greater than the total processing load:

$$RUPC\,(IU) \geq \left[\sum_{i=1}^{n} P_i \; H_i \right] R$$

From which:

$$R \text{ (maximum)} = \frac{RUPC\,(IU)}{\sum_{i=1}^{n} P_i \; H_i}$$

A.5 Examples of processing capacity computations

A.5.1 For a central processing unit

Inputs

Information furnished by manufacturer:

- RUPC = 100 000 central processor reference capacity units per hour;
- Weighting factors (see Table A.1).

TABLE A.1/Q.543

Termination type	Originating portion (F)	Termination (disposition) portion (F′)
Basic analogue access line	0.60	0.40
Analogue access line with supplementary services	0.72	0.48
ISDN access line	0.72	0.56
Interexchange circuit (IXC)	0.50	0.40

Information furnished by the Administration.

Expected traffic mix (see Table A.2).

TABLE A.2/Q.543

Originating call type	From – termination type	Traffic mix (fraction of total)
Telephone	Basic analogue access line	0.28
Telephone	Analogue access line with supplementary services	0.32
64 kbit/s switched	ISDN access line	0.05
Packet switched (set-up)	ISDN access line	0.02
Incoming-circuit switched	Interexchange circuit (IXC)	0.33
	Total	1.00
Terminating call type	**To – termination type**	**Traffic mix (fraction of total)**
Telephone	Basic analogue access line	0.26
Telephone	Analogue access line with supplementary services	0.30
64 kbit/s switched	ISDN access line	0.05
Packet switched (set-up)	ISDN access line	0.02
Outgoing-circuit switched	Interexchange circuit (IXC)	0.37
	Total	1.00

Computation (see Table A.3).

TABLE A.3/Q.543

Termination type	Originating portion	Terminating portion
Basic analogue access line	$0.28 \times 0.60 = 0.168$	$0.26 \times 0.40 = 0.104$
Analogue access line with supplementary services	$0.32 \times 0.72 = 0.230$	$0.30 \times 0.48 = 0.144$
ISDN access line – circuit switched	$0.05 \times 0.72 = 0.036$	$0.05 \times 0.56 = 0.028$
ISDN access line – packet switched	$0.02 \times 0.72 = 0.014$	$0.02 \times 0.56 = 0.011$
Interexchange circuit (IXC)	$0.33 \times 0.50 = 0.165$	$0.37 \times 0.40 = 0.148$
Total	0.613	0.435

Maximum call attempt rate for the central processor for the specified mix of traffic:

$$R \text{ maximum } = \frac{100\,000}{0.613 + 0.435} = 95\,420 \text{ call attempts per hour.}$$

At this point in the computation, it would be wise to examine the exchange design to verify that hardware configuration, memory capacity, or any other possible limitations do not prevent reaching this computed capacity.

A.5.2 Example of a processing capacity computation for an interface unit (see Table A.4)

Weighting factors are furnished by the manufacturer.

Traffic mix is estimated by the Administration.

TABLE A.4/Q.543

	Call type	Weighting factor		Traffic mix (fraction of total)	
From:					
Basic analogue access line	Telephone (reference call)	1.00	×	0.14	= 0.140
	False start/abandon	1.16	×	0.005	= 0.006
Analogue access line	Telephone	1.15	×	0.10	= 0.115
	False start/abandon	1.20	×	0.005	= 0.006
	Supplementary service No. 1	1.52	×	0.05	= 0.076
	Supplementary service No. 2	1.31	×	0.01	= 0.013
	Supplementary service No. n	1.++	×		
ISDN access line	64 kbit/s switched	1.20	×	0.025	= 0.030
	Packet call set-up	1.15	×	0.01	= 0.012
	Supplementary service No. 1	1.44	×	0	
	Supplementary service No. 2	1.20	×	0.01	= 0.012
	Supplementary service No. n	1.++	×		
IXC – CCITT No. 5	Incoming	1.30	×	0.07	= 0.091
IXC – CCITT No. 7	Incoming	0.90	×	0.08	= 0.072
To:					
Basic analogue line	Telephone	0.65	×	0.13	= 0.085
Analogue line	Telephone	0.75	×	0.12	= 0.090
	Supplementary service No. 4	0.80	×	0.035	= 0.028
ISDN	64 kbit/s switched	0.75	×	0.02	= 0.015
	Packet call set-up	0.75	×	0.01	= 0.008
	Supplementary service No. 5	0.80	×	0.01	= 0.008
IXC – CCITT No. 5	Outgoing	1.62	×	0.08	= 0.130
IXC – CCITT No. 7	Outgoing	0.83	×	0.10	= 0.083
				Total	1.020

Information from the manufacturer.

Reference capacity for an interface unit = 15 000 reference capacity half-units per hour.

Computation:

$$R \text{ maximum } = \frac{15\,000}{1.020} = 14\,705 \text{ half-units per hour or 7352 call attempts per hour.}$$

If the traffic load is distributed in the above proportions across all interface unit the number of interface units required to fully load the central processing unit would be 13 (95 420 divided by 7352). In this case it would probably be wise to plan on a maximum of 14 interface units in order to reserve some processing capacity for future program growth. At this point in the computation, it would be wise to examine the exchange design to verify that hardware configuration, memory or any other possible limitations do not prevent reaching this computed capacity.

The above capacity computation methodology can also be used to study the effects of different traffic mixes on interface units.

A.6 Packet handling

A.6.1 Definitions

A.6.1.1 packet: the unit of information exchanged between processors at layer 3.

A.6.1.2 user packet: A packet of information exchanged between the originating and terminating users in a packet switched connection. The length of packets may vary, depending on the protocol used. The number of user packets transferred between the originating and terminating users measures the amount of information transferred. The fundamental measure of packet switching capacity is expressed as the number of some agreed standard length user packets per second.

A.6.1.3 acknowledgement packet: Packet switching protocols have various strategies to ensure the reliable transmission of packets between users. These strategies involve sending packets not containing user data to verify the successful transmission of user packets. Such packets are called acknowledgement packets. The acknowledgement strategy depends on the packet switching protocol being used.

A.6.1.4 reference packet type: An arbitrarily selected user packet type, usually one of a protocol that is expected to be involved in a significant portion of the packet traffic an exchange might handle.

A.6.1.5 reference packet work unit: The amount of processor capacity required to handle one packet of the reference packet type together with its "share" of capacity required to handle associated acknowledgement packets. The reference packet work unit is assigned unit value.

A.6.1.6 weighting factor: The ratio of the amount of processing capacity required to handle any type of packet (including its "share" of associated acknowledgement packets) to the amount of processing required to handle one reference packet (including its "share" of associated acknowledgement packets). For example, if a complete reference packet requires 1000 processor cycles and a complete X.25 message packet requires 1200 cycles, the weighting factor for that packet type would be 1.2. The weighting factors must be furnished by the manufacturer for each packet type handled by the exchange.

A.6.1.7 reference packet processing capacity (RPPC): The total number of reference type user packets that can be handled by the processor in one second while meeting the specified performance criteria. This number should be furnished by the manufacturer. It is important to note that RPPC derives from that processing capacity reserved for packet handling and generally is the installed capacity diminished by an amount required for overhead, administrative tasks, etc.

A.6.2 Packet calls

Packet calls consist of two parts: packet call set-up (and disconnect) and ongoing packet exchanging (packet handling stage).

A.6.2.1 Packet call set-up can be dealt with in the same manner as that described previously for circuit switched call set-up. Appropriate weighting factors for the various types of packet call set-up and estimates of packet type calls in the traffic mix are used for computing the capacity of the processor involved. (See A.5. Packet call set-up was included in the example of call attempt processing capacity computations). Just as with circuit switched services, there may be packet calls with different processing requirements and therefore it will be necessary to treat the different type packet calls individually in the computation.

A.6.2.2 After packet call set-up, each packet exchanged between users during the call requires processing at the originating and terminating exchanges. The total amount of processing work required during a packet switched call is a function of the number of packets exchanged throughout the call. If a processor is dedicated to handling packets, the processing capacity is usually expressed in terms of number of user packets of a standard length handled per second. To account for the packet processing capacity that will be needed in an exchange during a busy hour, data on the average number (and type) of packets per call must be forecast. Note that for very long duration calls, e.g. permanent virtual circuits, only packets offered during the busy hour need to be considered. Also, packets from long duration calls originated prior to but extending into the busy hour, must be included.

In the exchange architecture shown in Figure A.1, it is assumed that each interface unit has a separate packet handling processor (shown as PH) within the unit. This processor interacts with digital line or digital circuit units to handle the protocols involved in packet switching. Once a packet call has been set-up, there is no further demand for processing work on the interface unit processor nor the central processing unit processor until call disconnect. Thus, the only potential capacity limitation due to packet handling in the exchange will be that imposed by the processing capacity of the packet handling processor in the interface unit. (For systems that use the same processor for call set-up and packet handling, see A.7.)

A.6.2.3 Processing capacity computation for a packet handling processor

Weighting factors are furnished by the manufacturer. Let G_k be the weighting factor for handling a user packet of type k (including the handling of an appropriate "share" of associated acknowledgement packets).

The data traffic mix (fractions of total) and volumes is forecast by the Administration.

Let Q_k be the fraction of user packets of type k. Note that:

$$\sum_{k=1}^{n} Q_k = 1$$

If R_p = user packet arrival rate, then the amount of processing capacity required for work associated with user packet traffic of the k-th type is:

$$Qk \ Gk \ Rp$$

In order to satisfy performance criteria the reference packet processing capacity (RPPC) must be equal to or greater than the total packet handling work. Thus:

$$RPPC \geq R_p \left[\sum_{k=1}^{n} Q_k \ G_k \right]$$

From which the maximum packet processing capacity R_p max is:

$$R_p \ \max = \frac{RPPC}{\sum_{k=1}^{n} Q_k \ G_k} \text{ packets per second.}$$

A.6.2.4 Example of a packet processing computation for an interface unit packet processor

Information furnished by the manufacturer:

a) RPPC = 1000 reference packet work units per second

b) Weighting factors (G):

- X.25 type data = 1.00 (reference type)
- X.75 type data = 0.70

Estimated data traffic mix (furnished by the Administration):

Type	Traffic portion (Q)
X.25	0.52
X.75	0.48

Computation:

Packet type	Processing factor
X.25 data	$1.00 \times 0.52 = 0.520$
X.75 data	$0.70 \times 0.48 = \underline{0.336}$
	Total 0.856

Maximum processing capacity for the above data traffic mix:

$$R_p \text{ max } = \frac{1000}{0.856} = 1168 \text{ packets per second.}$$

If the estimated data packet arrival rate (R_p) does not exceed the above number, then packet handling capacity in the interface unit will not limit the number of digital lines or circuits that generate data packets terminated on the unit. If it does exceed the above number, the digital lines and circuits generating the packet traffic will have to be spread over more interface units.

A.7 Capacity computation for exchange architectures other than that assumed in Figure A.1

If the same processor is used for both call set-up (circuit switched calls and packet calls) and for handling data packet traffic, the capacity of the processor must be allocated between the two functions. This can be done by computing the capacity of the processor for each function separately [with zero capacity used for the other function] and then allotting capacity between the two functions as required. Thus, if a processor has a maximum call processing capacity of 100 000 calls per hour or 1000 packets per second, for every 100 packets per second of packet handling capacity required, the call processing capacity will be reduced by 10 000 calls.

A.8 Conclusion

The methodology shown here illustrates a possible approach for determining the limiting factors in an exchange design and for computing its processing capacity. It is most important that the exchange architecture be understood, that capacity limiting elements be identified and that the proper computations be made to determine the true capacity of the exchange. These procedures can be used in engineering and loading the exchange most effectively. Trade-offs can be made between the use of capacity for various purposes. For example, in Figure A.1, a signalling terminal is shown connected to an interface unit. In that IU, the available processing capacity will be reduced by the amount of work required by the interface unit to support that terminal. The remainder of the processing capacity can be allocated effectively by using information generated in the call processing computation methodology.

It is also very important that the capacity of an exchange should not be calculated using the entire capacity for call processing. It should be made using the processing capacity available under “normal” operating conditions with the exchange performing all the operations and administrative functions expected of it during the busy hour.

Annex B

An example of a methodology for measuring exchange capacity

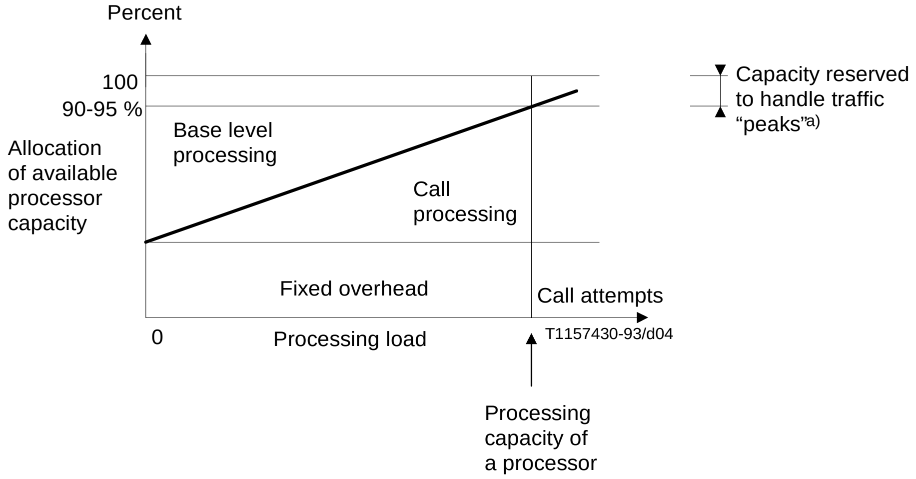
(This annex forms an integral part of this Recommendation)

B.1 General

The capacity of an exchange used for call processing can be measured in a laboratory or in the field and projections can be made to predict the maximum processing capacity of the exchange design for the configuration and load characteristics involved in the measurements. This annex serves as an example of a methodology that makes it possible to measure the processing capacity of an exchange for the configuration and load characteristics involved in the measurement.

B.2 Theory behind the measurement method

The call handling capacity of a processor can be expressed in terms of the maximum number of calls (or call attempts) which can be processed in a fixed interval of time while meeting all service criteria. In normal conditions, the work functions performed by a switching system processor can be divided into three categories (one fixed level and two variable) as shown in Figure B.1.



a) The amount of reserved capacity depends on system architecture and hierarchical position of the processor.

FIGURE B.1/Q.543

Allocation of processing capacity

At normal loads, a linear relationship is usually observed between offered load and processor utilization. However, at heavy loads, some system components may become overloaded and this can be reflected in non-linearity in the processor utilization versus load characteristic.

In the case of a single processor controlled system, Figure B.1 represents the processing capacity of the exchange. In a multi-processor system, the capacity is distributed among processors and the exchange capacity is related to the system configuration and the exchange processing capacity is a function of the processors involved in call handling functions.

As shown in Figure B.1, the processing capacity of a processor is divided between three elements:

1) fixed overhead related to mandatory tasks (e.g. task scheduling and scanning);

2) call processing work (including traffic-related overhead tasks);

3) deferrable (base-level) tasks (e.g. routine maintenance).

The tasks which a processor executes are assigned to three levels of priorities, base, medium and high-level tasks [see diagrams a) and b) of Figure B.2].

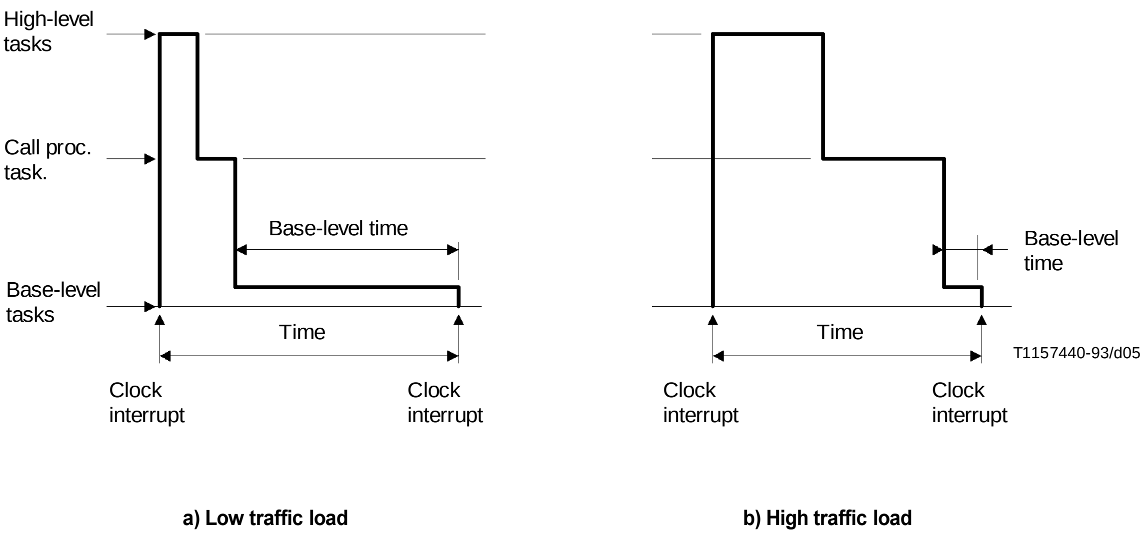


FIGURE B.2/Q.543

Allocation of processor time to tasks

As the traffic load (call attempts) increases call processing work expands and the processing of deferrable tasks decreases.

Measurement of the percentage of time spent by the processor performing base-level tasks gives an indication of the percent or processing capacity required for a particular load on the processor.

As shown in diagram a) of Figure B.2, at low traffic load, the percentage of time used to perform base-level tasks is relatively high. In diagram b) of Figure B.2, at high traffic load, the percentage of time at base-level is relatively low. Thus the measurement of percentage of time used to perform base-level tasks can be used to determine call processing capacity.

B.3 Capacity measurement methodology for exchanges

Measurements can be performed on exchanges in laboratories or in the field to measure capacity usage for various load levels and then to project the data to estimate the call processing capacity of a processor.

The collection of data will depend on facilities available to perform the required measurements. The exchange may be designed to provide indications of time spent performing base-level tasks or it may be necessary to access the bus system of a processor in order to measure this time. Equipment will be needed to create loads, or loads in a working exchange must be measured in order to establish load points. Various level loads for the various types of calls (or services) should be observed in order to establish a basis for projecting the load line to determine the maximum processing capacity for the mix of traffic services assumed or measured. In projecting call capacity care must be taken not to extrapolate beyond the linear region of the processor utilization versus offered call attempts relationship (see Figure B.3).

Where multi-processors are involved, the exchange configuration, the distribution of traffic types and processing capacity of each processor must be examined to determine the limiting factors that control the exchange capacity (as discussed in Annex A, "An example of methodology for computing the call processing capacity of a digital exchange, taking into account ISDN services, including packet data handling").

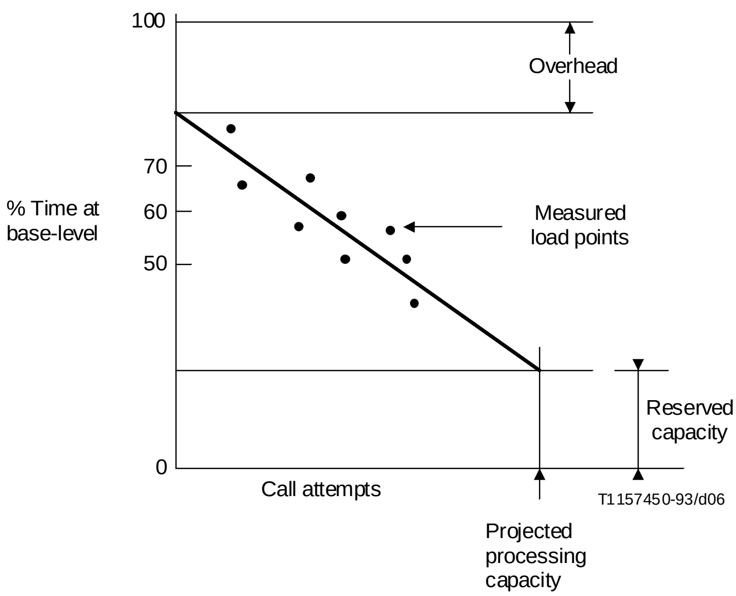


FIGURE B.3/Q.543

Measurement of processing capacity