TECHNICAL REPORT

ISO/TR 13283

First edition
1998-12-15

Industrial automation — Time-critical communications architectures — User requirements and network management for time-critical communications systems

Automatisation industrielle — Architectures de communications en temps réel — Prescriptions des utilisateurs et gestion de réseau pour les systèmes de communications en temps réel

Reference number
ISO/TR 13283:1998(E)

Contents

Page

International Organization for Standardization
Case postale 56 • CH-1211 Genève 20 • Switzerland
Internet iso@iso.ch

Printed in Switzerland

Foreword

ISO (the International organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The main task of ISO technical committees is to prepare International Standards. In exceptional circumstances a technical committee may propose the publication of a Technical Report of one of the following types:

— type 1, when the necessary support cannot be obtained for the publication of an International Standard, despite repeated efforts;

— type 2, when the subject is still under technical development or where for any other reason there is the future but not immediate possibility of an agreement on an International Standard;

— type 3, when a technical committee has collected data of a different kind from that which is normally published as an International Standard ("state of the art", for example).

Technical Reports of types 1 and 2 are subject to review within 3 years of publication, to decide whether they can be transformed into International Standards. Technical Reports of type 3 do not necessarily have to be reviewed until the data they provide are considered to be no longer valid or useful.

ISO/TR 13283, which is a Technical Report of type 3, was prepared by Technical Committee ISO/TC 184, *Industrial automation systems and integration*, Subcommittee SC 5, *Architecture and communications*.

It is becoming apparent that in their present form the OSI Standards cannot readily accommodate time-critical communications. The network architectures which have been standardized so far were primarily intended for general traffic and are not always capable of providing adequate performance and resilience for time-critical communications, especially where time-critical and non-time-critical traffic coexists. In particular in many CIM and control installations there appears to be a requirement for an intermediate network, between general enterprise-wide networks (e.g. MAP) and the fieldbus-type networks. This intermediate network should carry both bulk data transfer ad time-critical messages, and be able to operate over considerable distance and in hostile environments.

A task force was therefore setup under ISO/TC 184/SC 5/WG 2 to look at requirements for a time-critical communications architecture, concentrating particularly on the requirements in intermediate network and thus complementing work done on requirements for time-critical communications in the fieldbus and to prepare a Technical Report on the topic. Among those requirements some are identified to be considered when managing time-critical communications network where the specified time window must be respected by the communication systems. The management and the support for these requirements are the second part of this Technical Report. The relationship with other standards is depicted in the Figure 1.

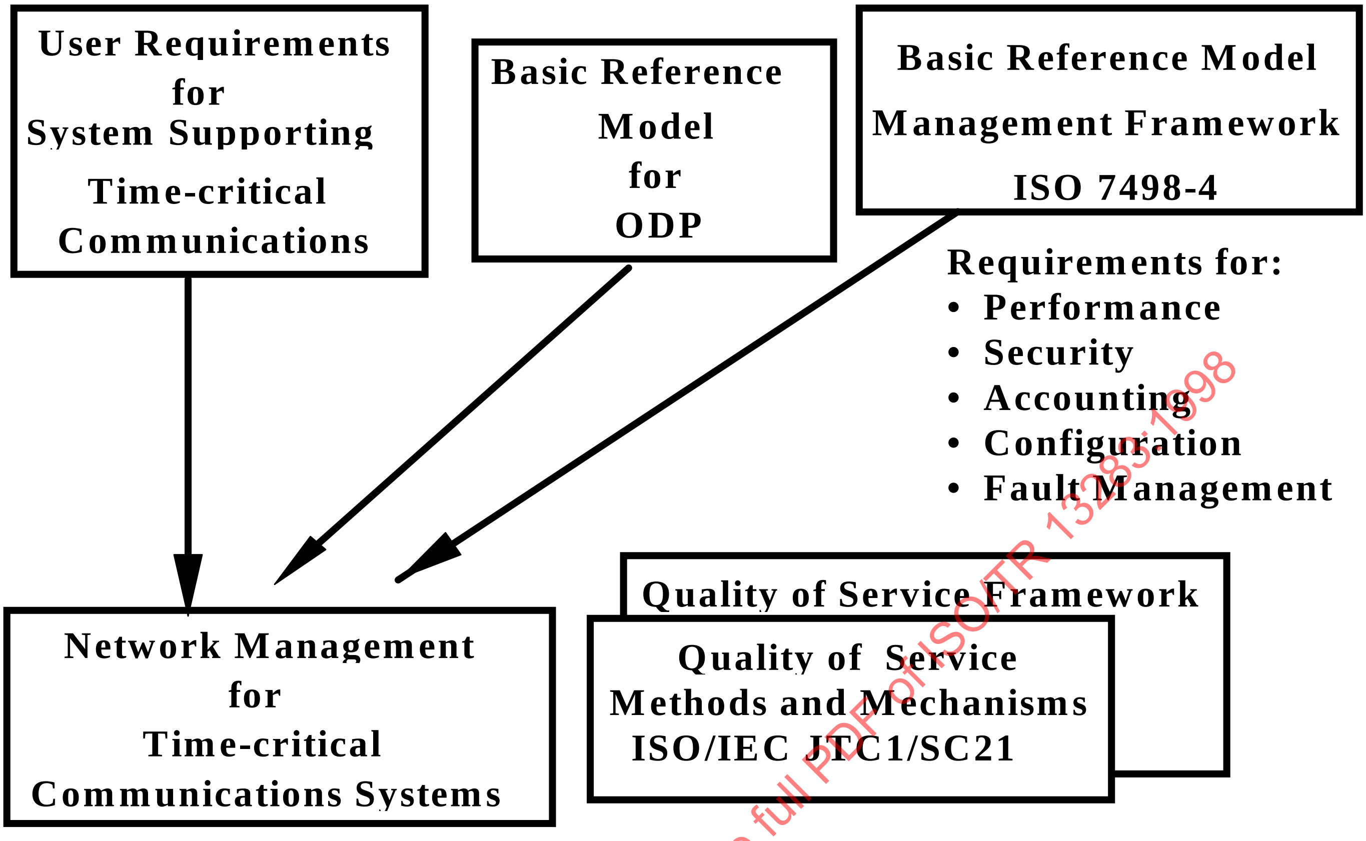

Figure 1 — Relationship with other standards

Time Critical Network Management (TCNM) defines only those additional requirements that are necessary to manage a TCCS network. TCNM does not describe the system management consideration regarding configuration management, accounting, security, and fault management except where these might be required to support the time critical aspects; i.e. support for TCNM directly impacts performance management to a great extent, but also impacts other areas which use time critical services to accomplish their function.

The focus of this Technical Report is to identify the communication needs of tightly coupled control systems, and the management and support of these needs. This effort is intended to complement efforts which have focused on factory information networks (e.g. MAP) which are characterized by high throughput, large packet service between information processors (e.g. mini-computers) and control processors (e.g. special purpose controllers). The effort is also intended to complement efforts to identify sensor networks (e.g. fieldbus) which concentrate on providing a low complexity interface to simple factory floor sensors, actuators and other devices.

There is no clear consensus on whether the time-critical communication architecture should be connection-oriented or connectionless. Either solution is acceptable providing that the requirement are met; however, some requirements may be easier to meet one of these solutions that with the other. Since MMS uses a connection-oriented session layer then the overall service may appear connection-oriented even if the network and transport layers are connectionless.

This Technical Report summarizes the user requirements for time-critical communications systems that have been identified by the task force and provides a short explanation of each of the requirements. Consideration is also given to the management and support of these requirements.

Introduction

General requirements for TCCS

A time-critical communications system (TCCS) is a type of communications network system which is optimized for communications where sequences and traffic patterns vary dynamically over a whole network system. However, such a communications network system also has to support pre-scheduled static sequences of communications.

Time-critical communications are communications in which there is a specific bounded time window within which one or more specified actions must be completed with some defined level of certainly. A TCCS provides the capability for both time-critical messages and non-time-critical messages to coexist. In a distributed time-critical application, algorithms, data and control structures are around networks. Such a distribution is only possible if the communication systems provide a certain quality of service, particularly regarding time constraints, coherence, etc.

It is recognized that all TCC systems differ from other networks in their requirement for resilience. The tighter the time constraints are, the more important strategies for resilience become, since networks capable of meeting tight time constraints need to exhibit great resilience if the data flow is not to be perturbed by minor faults.

A TCC architecture complies with the OSI Basic Reference Model whenever appropriate, meeting the user requirements identified in clause 4. Current OSI standards cannot readily accommodate time-critical communications because their primary objectives are to ensure the transfer of general data traffic or to satisfy intercommunication between delivery of messages. Therefore, the current OSI model and architecture needs to be enhanced to support TCC.

User requirements for TCCS

Automated factories use networks for distributed control applications. These applications require communications systems which are designed to accommodate worst-case performance as opposed to average performance; in such systems timeliness is of paramount importance. The network should be able to cope with both bulk data transfers and time-critical messages and should support methods for achieving resilience.

If necessary, interfacing the TCCS with the main factory data communication network should be a simple procedure, using a linking device. Because of the requirement to operate in very harsh environments, highly reliable media and signaling methods are necessary so that a very low bit error rate results and a minimum number of retransmissions is necessary. The user needs to be able to define communications priorities and control error recovery mechanisms. Network management functions are necessary to allocate resources, control access to TCC groups, detect latent failures and actual failures, etc.

The user requirements that are identified allow the concept of time windows to be realized in OSI. The inter-relationship between TCC architecture, characteristics or attributes, time windows and applications requiring time-critical communications are set out in Figure 2.

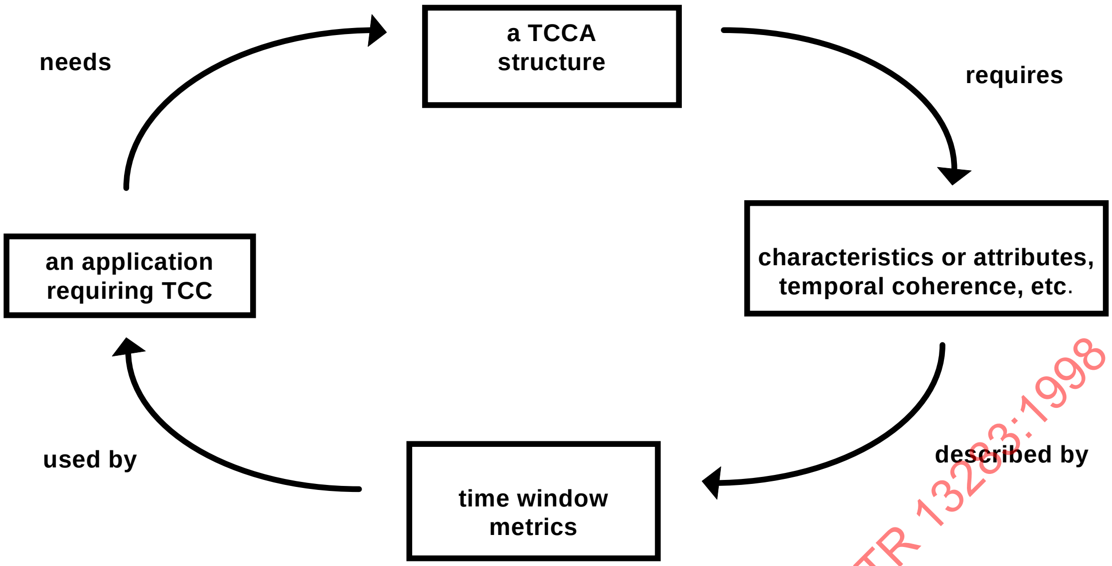

Figure 2 — Inter-relationship between TCC architecture, characteristics, time windows and applications requiring time-critical communications

Network management in TCCA

The requirements on network management are extracted and translated from the user requirements in TCCS in terms of the provision of services which is essential in order to meet the time constraints specified by users under the circumstances in which the behaviour of the time-critical communications network system varies dynamically.

In the sense of the QoS Basic Framework, the QoS requirements are information to express part or all of the network management requirements, and are expressed in terms of QoS parameters as part of a mechanism when a QoS requirement is conveyed between entities. This information closely concerns a policy that is to apply to the operation of a time-critical communications network system.

In some time-critical applications all user requirements need not always be satisfied. It is therefore not essential for a TCCA or the network management in a TCCA to meet all the network management requirements identified herein. This means that there will be allowable classes which meet specific subsets of these requirements. These classes are not addressed in this Technical Report.

Industrial automation — Time-critical communications architectures — User requirements and network management for time-critical communications systems

1 Scope

This Technical Report identifies user requirements for systems supporting time-critical communications systems and the network management aspect which are specific to time-critical communications architecture in peer-to-peer and multipeer communications in or between application processes.

This Technical Report translates user requirements for systems supporting time-critical communications into QoS requirements for the management and support of a time-critical communications network, using the concepts and terminology of the QoS Basic Framework developed by ISO/IEC JTC1/SC21/WG7.

This Technical Report also describes a model, characteristics and functions for network management in time-critical communications systems.

This Technical Report focuses on time-critical communications systems used in discrete parts manufacturing applications; however, these time-critical communications systems may also be used in other applications, including process control.

This Technical Report concentrates primarily on time-critical communications systems for event-driven applications in which traffic flow and network configuration change dynamically; however, this Technical Report may also be applicable to state-driven applications in which traffic patterns and configuration are static.

2 References

The following standards contain provisions which, through reference in this text, constitute provisions of this Technical Report. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this Technical Report are encouraged to investigate the possibility of applying the most recent editions of the standards indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

2.1 International Standards

ISO/IEC 7498-1:1994, *Information technology — Open Systems Interconnection — Basic Reference Model: The Basic Model.*

ISO/IEC 7498-4:1989, *Information processing systems — Open Systems Interconnection — Basic Reference Model — Part 4: Management framework.*

ISO/IEC 9506-1:1990, *Industrial automation systems — Manufacturing Message Specification — Part 1: Service definition.*

ISO/IEC 9506-2:1990, *Industrial automation systems — Manufacturing Message Specification — Part 2: Protocol specification.*

ISO/IEC 9545:1994, *Information technology — Open Systems Interconnection — Application Layer structure.*

ISO/IEC 10040:1992, *Information technology — Open Systems Interconnection — Systems management overview.*

ISO/IEC 10164-1:1993, *Information technology — Open Systems Interconnection — Systems Management: Object management function.*

ISO/IEC 10164-11:1994, *Information technology — Open Systems Interconnection — Systems Management: Metric objects and attributes.*

ISO/IEC 10164-13:1995, *Information technology — Open Systems Interconnection — Systems Management: Summarization function.*

ISO/IEC 10164-20, *Information technology — Open Systems Interconnection — Systems Management: Time management function.*

ISO/IEC 10165-1:1993, *Information technology — Open Systems Interconnection — Management Information Services — Structure of management information: Management Information Model.*

ISO/IEC 10165-2:1992, *Information technology — Open Systems Interconnection — Structure of management information: Definition of management information.*

ISO/IEC 10165-4:1992, *Information technology — Open Systems Interconnection — Structure of management information: Guidelines for the definition of managed objects.*

ISO/IEC 10731:1994, *Information technology — Open Systems Interconnection — Basic Reference Model — Conventions for the definition of OSI services.*

ISO/IEC 10746-2:1996, *Information technology — Open Distributed Processing — Reference Model: Foundations.*

2.2 Documents under development

ISO/IEC JTC 1/SC 21 N 9309, *Quality of Service — Basic Framework.*

ISO/IEC JTC 1/SC 21 N 9310, *Quality of Service — Methods and Mechanisms.*

3 Terms and definitions

For the purposes of this Technical Report, the following terms and definitions apply.

3.1 TCCA terms

3.1.1
application dependent requirement
a requirement related to the usage of the time-critical communications architecture and should be independent of the implementation

3.1.2
application entity
AE
part of the application process that deals with the communications system

3.1.3
implementation-dependent requirement
a requirement which depends on the specific aspects of the implementation of the time-critical communication architecture being used

3.1.4
interaction
any causal relation induced by message transfer

3.1.5
media access control
MAC
the data link layer responsible for scheduling and routing data transmission on a shared medium

3.1.6
multipeer data transmission
the transmission of a PDU to one or more destinations

3.1.7
multipeer group
a group of peer entities which are mutually willing and able to be senders or receivers of multi-peer data transmissions with other members of the group

3.1.8
performance
the behaviour of the system with respect to time

3.1.9
spatial coherence
a property of duplicated lists of variables indicating whether or not all the copies are identical at a given time or within a given time window

3.1.10
temporal coherence
a property of a list of variables indicating whether or not the value of each variables in the list has been produced and transmitted and/or received within a given time window

3.1.11
time-critical communications
TCC
when one or more application process which send a message require it to be received (or received and acted upon, or received and acted upon and confirmed) within a certain bounded time period window after its send request to the system

3.1.12
time-critical communications architecture
TCCA
an architecture, in the sense of the OSI reference model, which supports the identified requirements for time-critical communications

NOTE TCCA is a structure which has layers, entities, service access points, protocols, connections, etc. in accordance with the concepts introduced in ISO 7498.

3.1.13
time-critical communications entity
TCCE
an entity participating in time-critical communications (a time-critical communications entity may or may not map directly to an application entity since a single application entity could have time-critical and non-time-critical components)

3.1.14
time-critical communications group
TCCG
a group of time-critical communications entities in a time-critical communications system

3.1.15
time-critical communications system
TCCS
a system in which there are time-critical communications

NOTE A TCCS is an implementation of a time-critical communications architecture. The system consists both of a user's application processes and a time-critical communications network.

3.1.16
time-critical communications transaction
TCCT
an ordered set of message transfers (or exchanges) coming from different application entities which all have to be completed within a given time window

NOTE This may be a time-critical multipeer communications transaction or a time-critical peer to peer communications transaction. Such transactions are complex information processing operations involving multiple data transmissions for an application with time constraints.

3.1.17
time-critical data transmission
TCDT
the transmission of a PDU within a given time window

3.1.18
time-critical multipeer data transmission
TCMDT
the transmission of a protocol data unit to more than one destination within a defined time window or time windows

3.1.19
time window
TW
a bounded time interval which is characterized by starting time and delay or starting time and end time which are application dependent

NOTE The degree of resolution of time and delay is implementation dependent.

3.2 OSI Basic Reference Model terms

This Technical Report uses the following terms drawn from ISO/IEC 7498-1:1994: (N)-layer, (N)-protocol, (N)-protocol-data-unit, (N)-service, (N)-service-access-point, (N)-subsystem, open system, and OSI environment.

3.3 Service convention terms

This Technical Report uses the following terms drawn from ISO/IEC 10731:1994: (N)-service-facility, (N)-service-provider, (N)-service-user, requester, and acceptor.

3.4 Basic Reference Model of Open Distributed Processing terms

3.4.1
quality of service
QoS
a set of qualities related to the collective behaviour of one or more objects

NOTE This definition was drawn from ISO/IEC/DIS 10746-2.1.

3.5 QoS Basic Framework terms

NOTE The terms in this subclause are drawn from ISO/IEC/JTC 1/SC 21 N 9309.

3.5.1
policy control function
PCF
the function that determines policy that is to apply to the operation of the subsystem

NOTE The PCF determines the constraints under which all other decisions of the subsystem are made. Any action is performed to control the remaining operation of the subsystem.

3.5.2
protocol entity
PE
the entity responsible for operating the protocol in order to provide the service to the service user and also responsible for operating with its peer protocol-entity, its peer service user, and the inferior service provider

3.5.3
quality control function
QCF
the function that takes account of QoS requirements in selecting the entities that will participate in communications

NOTE The QCF may represent a point of interaction at which knowledge of QoS may be used to influence the choice and at which the consideration of the choice may affect the treatment of QoS.

3.5.4
QoS alert
the use of QoS mechanisms to signal to an entity that some limit has been reached or threshold crossed

3.5.5
QoS attribute
an attribute of a managed object relating to QoS

3.5.6
QoS characteristic
a quantifiable aspect of QoS, which is defined independently of the means by which it is represented or controlled

3.5.7
QoS context
QoS information that is retained, interpolated or extrapolated by one or more entities and used in managing QoS

NOTE QoS context is further classified into requirement context and data context.

3.5.8
QoS data
QoS information other than QoS requirements, e.g. warning, QoS measures, and information used in QoS inquiries

3.5.9
QoS enquiry
the use of QoS mechanisms to determine properties of the environment relating to QoS in which communications take place

3.5.10
QoS establishment
the use of QoS mechanisms to create the conditions for some systems activity before that activity occurs, so that a desired set of QoS characteristics is attained

3.5.11
QoS information
information related to QoS classified into QoS context (when retained in an entity), QoS parameters (when conveyed between entities), QoS requirements (if it expresses requirements for QoS) and QoS data (if it does not express requirements for QoS)

3.5.12
QoS maintenance
the use of QoS mechanisms to maintain a set of QoS characteristics at required values for some systems activity, while the activity is in progress

3.5.13
QoS management
any set of activities performed by a system or communications service to support QoS monitoring, control, and administration

3.5.14
QoS management function
a function specifically intended to meet a user or application requirement for QoS, provided by one or more QoS mechanisms

3.5.15
QoS measure
one or more observed values relating to a QoS characteristic

3.5.16
QoS mechanism
a specific mechanism that may use protocol elements, QoS parameters, or QoS context, possibly in conjunction with other QoS mechanisms, in order to support establishment, monitoring, maintenance, control, or enquiry of QoS

3.5.17
QoS monitoring
the use of QoS measures to estimate the values of a set of QoS characteristics actually achieved for some systems activity

3.5.18
QoS parameter
QoS information that is conveyed between entities as part of a QoS mechanism

NOTE Parameters are classified into requirement parameters and data parameters. The information conveyed may relate to one or more QoS characteristics.

3.5.19
QoS requirement
QoS information that expresses part or all of a requirements to manage one or more QoS characteristics, e.g. a maximum value, a target, or a threshold

NOTE When conveyed between entities, a QoS requirement is expressed in terms of QoS parameters.

3.5.20
QoS operating target
QoS information that represents the target values of a set of QoS characteristics, derived from QoS requirements

3.6 TCNM terms

3.6.1
network management_management entity
(NM)-ME
a management service entity or management agent responsible for operating management actions, e.g. state control and report, alarm report, logging, violation report, access control, monitor metrics, test control, scan report, scheduling, resource control, knowledge management, etc., and for operating the (NM)-protocol in order to co-operate with other peer (NM)-subsystem in the sense of system management in OSI management

NOTE The (NM)-ME combines two functions classes that represent both a system-wide and a locally restricted capability for tuning and coordinate the behaviour of the various QoS entities in operation in distributed TCNM subsystems.

3.6.2
network management policy control function
(NM)-PCF
the function that determines the TCNM policy that is to apply to the operation of the TCCN

NOTE The (NM)-PCF determines the constraints under which all other decisions of the (NM)-subsystem are made, and the constraints affects the (N)-PCF in the (N)-subsystem in order to fulfill the QoS requirements imposed on TCCN. Any action is performed to control the remaining operation of the (NM)-subsystem.

3.6.3
network management quality control function
(NM)-QCF
the function that takes account of QoS requirements in selecting the entities that will participate in the network management

NOTE The (NM)-QCF represents a point of interaction at which knowledge of QoS will be used to influence the choice and at which the consideration of the choice will affect the treatment of QoS.

3.6.4
TCNM-management information base
TCNM-MIB
a conceptual repository in which common knowledge of the behaviour of the various QoS entities in operation will be used to influence the policy control, the QoS control, the protocol processing and the management agent processing

3.6.5
time-critical communications network
TCCN
a network of an implementation of a time-critical communications architecture

3.6.6
time-critical network management
TCNM
network management which supports network management functionality to be identified in this work

NOTE TCNM primarily covers a network management within a TCCN.

4 User requirements for time-critical communications systems

4.1 Introduction

User requirements for time-critical communications systems (TCCS) can be grouped into these categories: user control, integration issues of openness, timing-related issues, and architecture considerations including topology, management, and sovereignty aspects. These requirements are discussed separately in detail in 4.3 to 4.6.

4.1.1 Summary of user control requirements

User control requirements can be summarized as follows:

a) differentiation of messages is essential;

b) user control of error recovery;

c) user specification of behaviour.

4.1.2 Summary of integration issues of openness

Integration issues of openness can be summarized as follows:

a) the TCCA shall be defined within the context of OSI;

b) TCCAs shall support ISO/IEC 9506 functionality in a timely fashion;

c) the TCCA requires standardized user interfaces.

4.1.3 Summary of timing-related issues

Timing-related issues can be summarized as follows:

a) agreed probability of message transfer time or transaction completion time is necessary;

b) Temporal coherence is required and spatial coherence may be required in TCCA, i.e.

   1) temporal coherence,

   2) time-critical communications transaction,

   3) spatial coherence, and

   4) synchronization.

4.1.4 Summary of architecture considerations

Architecture considerations can be summarized as follows:

a) scheduling functions, i.e.

   1) differentiation of PDUs,

   2) non-time-critical PDUs shall not prejudice the delivery of time-critical PDUs, and

   3) static and dynamic scheduling;

b) dependability, i.e.

   1) message transfer time,

   2) mechanisms to assist latent failure detection are desirable in TCCSs, and

   3) availability;

c) network management, i.e.

   1) membership of TCCS shall be controlled,

   2) reconfigurability of the application and the use of the TCCS,

   3) off-segment communication, and

   4) management and sovereignty;

d) some protocols implementation aspects, i.e.

   1) applications requiring time-critical traffic shall not commit time-critical PDUs to the transmitting stack faster than it or the network or the receiving stack can handle the traffic, and

   2) class of conformance to individual requirement level.

4.2 Transaction models

A TCC architecture needs to provides for a number of different communication scenarios. In general the user's overview is perceived from the application; an AE may map directly to the TCC entity or party to the TCC entity because it also has non-TCC-entity elements in it (see Figure 3).

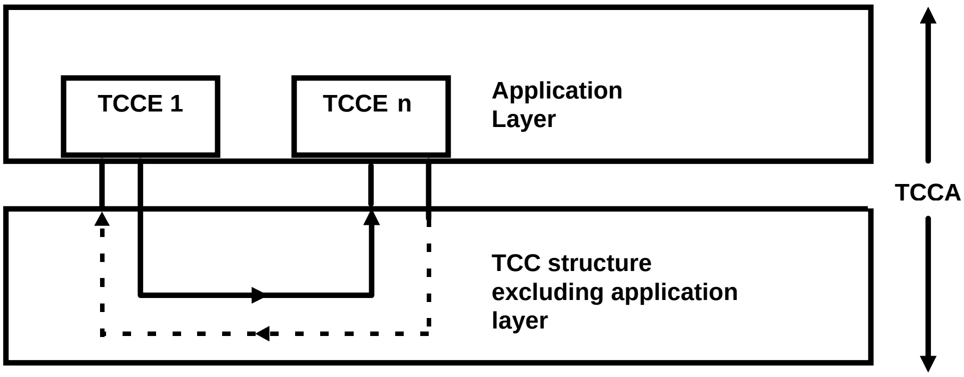

Figure 3 — Relationship of TCCEs and TCCA

A variety of application models have been identified and are referred to as client-server, initiator-responder, producer-consumer, producer-distributor-consumer, etc. The following types of transfers have been identified as essentials for TCC, but the technical implications for TCC systems are not discussed in detail as these will depend on the particular TCC system chosen and its implementation:

— point-to-point, i.e. an exchange from one sender to one receiver;

— multicast, i.e. an exchange from one sender to many (n) receivers which form a known group of TCC entities.

NOTE Where an exchange is from one sender to all receivers without negotiation of the communications context it may be referred to as broadcast.

In applications there will be the following interactions:

— peer-to-peer interaction, where the application which initiates the interaction or requests the service (the initiator) is the sender and there may or not be a requirement for a response (see Figures 4a and 4b);

— multipeer interaction, where the initiator may make one multipoint transfer or n point to transfers and again there may, or may not, be a requirement for responses, which may be confirmations from each member of the group (see Figures 5a, 5b and 5c);

— peer-to-distributing peer is sometimes recognized as a third category, but this can be equated to Figures 4a or 4b and 5a, 5b or 5c depending on what level of confirmation is required (see Figure 6).

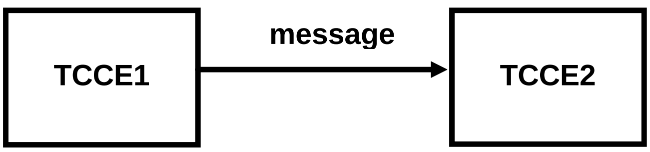

Figure 4a — Peer-to-peer without confirmation

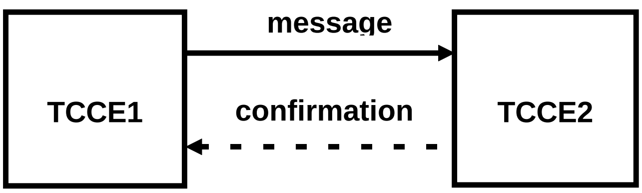

Figure 4b — Peer-to-peer with confirmation

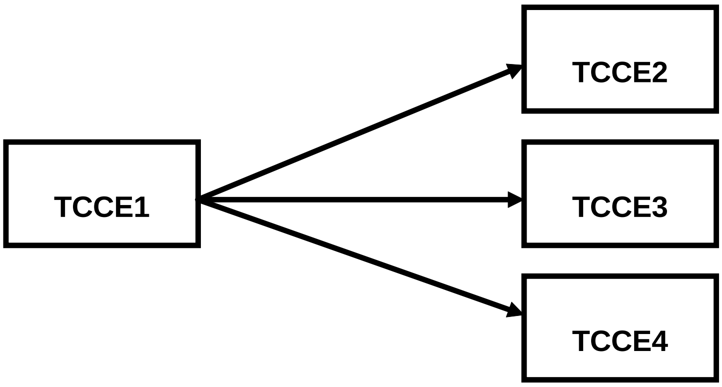

Figure 5a — Multicast of members of group without confirmation

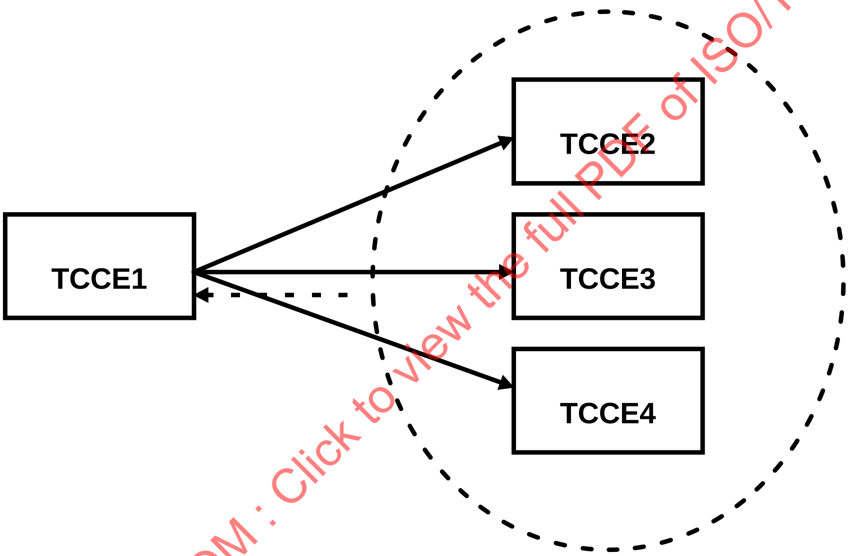

Figure 5b — Multicast to members of a group with confirmation from group of receipt of that message

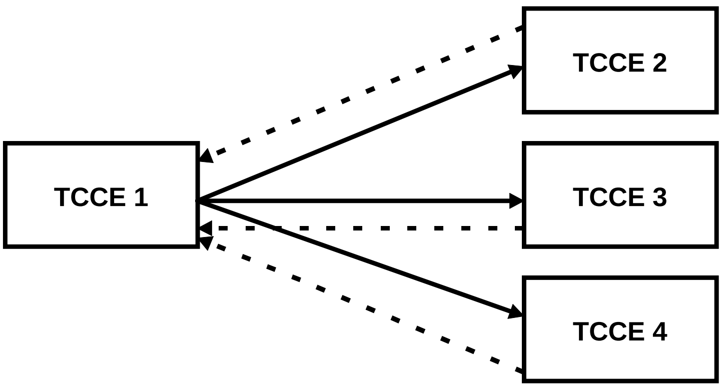

Figure 5c — Multicast to members of a group with individual confirmation to that message transfer from each member

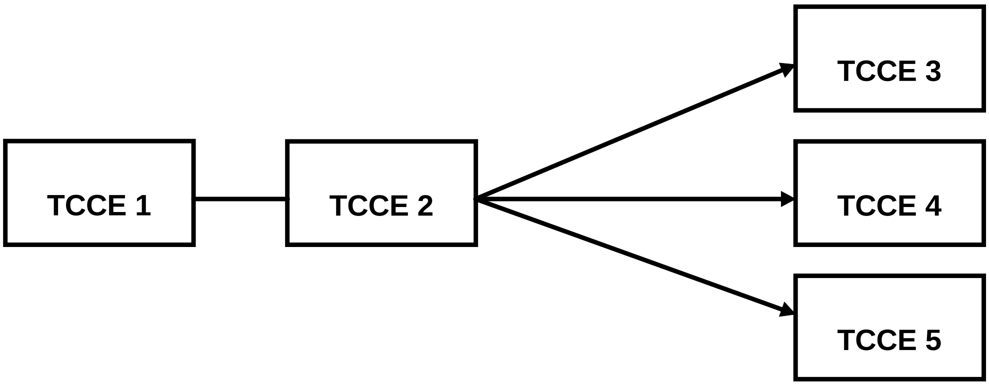

Figure 6 — Peer-to-distributing peer

A TCC system clearly requires point-to-point communication. There should be a facility which allows an application to invoke procedures or request services, or information, or both, on an ad-hoc basis. In many situations a TCC system needs the capability to conduct communications with several applications at the same instant, such as the synchronization of real-time clocks at multiple locations. Apparently that TCC systems require multicast capability for spatial resources and temporal coherence and for efficient use of limited resources, e.g. bandwidth. Finally, there is a requirement to minimize the communications requirements on simple devices, while allowing open access to the most the current data.

Distributed real-time applications are expected to have a varied mix of messages including periodic traffic (e.g. remote data samples), aperiodic regular traffic (e.g. file transfers, user requests), and sporadic traffic (e.g. alarms, reconfiguration requests). Therefore a TCC system has to support both periodic and aperiodic traffic, i.e. both state-driven and event-driven traffic. Since many time-critical applications will be event-driven (i.e. there will be variable loading on the network in terms of traffic, TCC entities, etc.), time-critical traffic has to be able to be sent within prescribed time windows with an acceptable probability of success in all circumstances.

4.3 User control

Some parameters and characteristics which define or contribute to the behaviour of the traffic communication system must be specified by the user. The communication system does not have knowledge regarding

- — membership behaviour,
- — error recovery, and
- — defining the urgency of messages.

4.3.1 Differentiation of messages is essential

4.3.1.1 Summary

The application can differentiate between time-critical and non-time-critical messages. In addition it is necessary for the communications system to distinguish between time-critical and non-time-critical messages and provide appropriate service.

4.3.1.2 Explanation

The differentiation of messages should be dynamically allocated when the application commits the message to the TCC entity. Once the messages have been differentiated the communication system should carry the messages from source to destination without altering their designation.

Differentiation could be done by using attributes attached to the PDUs, and/or multiple (dual) stacks. The former is the most elegant solution but involves changing many existing standards whereas the latter could be introduced more immediately. The coexistence of time-critical and non-time-critical PDUs in a single stack architecture may lead to processing difficulties so multiple stacks may be required in addition to PDU tagging. Other alternative mechanisms may be devised to allow the differentiation of messages in TCC systems. It may be a requirement that time-critical PDUs can pre-empt non-time-critical PDUs.

4.3.2 User control of error recovery

4.3.2.1 Summary

The detection and reporting of failure to complete the delivery of a message, or to receive an acknowledgment within a nominated time frame, is necessary.

4.3.2.2 Explanation

After notification of failure, retransmissions will typically cease unless the user requests a new attempt to send the message. A message can be nominated to belong to a particular category or nominated to use a given time frame, and perhaps be allowed to be retransmitted a specific number of times (possibly zero).

A TCC architecture shall be able to handle any type and combination of types of data traffic such as periodic and aperiodic time-critical traffic and aperiodic non-time-critical traffic.

Nomination of message could be regarded as a dynamic (QoS) parameter. Dynamic QoS parameters are those that can be set for a particular invocation of the service and, in a TCC system at the application layer service interface, should include an item such as nomination.

4.3.3 User specification behaviour

4.3.3.1 Summary

Before modifying the associations between APs or adding new ones, it is necessary to verify that the behaviour of the system is not damaged by such modifications.

4.3.3.2 Explanation

The control of membership of off-segment communication is made by some elements of the TCC architecture — the network management, for example. But to be able to do that, it is mandatory that the user gives the element's parameters and characteristics of behaviour to the system. This may be possible at a configuration stage or upon opening an association.

4.4 Integration issues of openness

The TCCA has to be integrated with the CIM architecture. This is possible by trying to conform to present CIM standards and by introducing necessary changes in evolving standards.

This subclause is concerned with the place of a TCCA in the OSI world.

4.4.1 The TCCA shall be defined within the context of OSI

4.4.1.1 Summary

A TCC architecture shall conform to the OSI layer standards wherever possible while also meeting the other requirements described in this Technical Report.

4.4.1.2 Explanation

An open systems orientation is necessary to allow integration with other real open-end systems. A TCC architecture should enable communications in the manufacturing environment to benefit from the same flexibility as other implementation developed in accordance with ISO/IEC 7498-1.

It is known that in their present form the OSI standards cannot readily accommodate time-critical applications. Technological improvements, however, may allow OSI to be used in an increasing number of time-critical applications in which time windows are of relatively short duration. The introduction of yet another technology to meet time-critical requirements should be avoided. However, changes to the layer standards may be necessary to meet the requirements described in this Technical Report.

4.4.2 TCC systems shall support ISO/IEC 9506 functionality in a timely fashion

4.4.2.1 Summary

A requirement of a TCC system in discrete parts manufacturing shall be a set of services aligned with ISO/IEC 9506, the Manufacturing Message Specification (MMS).

4.4.2.2 Explanation

MMS services provide peer-to-peer application interactions in accordance with standard syntax and semantics. Standardized semantics are necessary to reduce the programming burden for application developers, and to promote interoperability between different implementations of the TCC architecture.

Clearly not all of the services provided by MMS will be required by all applications using the architecture. Where the functionality provided by MMS services is implemented in a real end-system, the corresponding MMS services shall also be implemented in applications which do not use the client-server model so that the service will “look and feel” like MMS.

Modifications or extensions to the content of MMS may be required to support multicast and broadcast message transfers on a multipoint association.

The services of TCC systems should be an enhanced set of MMS services (connectionless or connection-oriented) which include time constraints. The TCC system should be able to co-operate with MMS compatible components.

4.4.3 The TCCA requires standardized user interfaces

4.4.3.1 Summary

Standardized interfaces to the application communication services shall be required.

4.4.3.2 Explanation

These standardized interfaces shall be in the form of function and/or procedure calls providing for synchronous operation, or asynchronous operation, or both. Interfaces shall be required for service invocation at the client side. Where the service functionality is implemented at a server by a “user” application, there shall also be standardized interfaces from the service provider to the user at the server side. Additional service interface, which are not standardized and add value to an implementation, are not precluded by this architecture.

Standardized interfaces are necessary to ensure application program transportability across different implementations of the architecture. Standard service specifications are also necessary; these functional definitions would include time constraints relating to the use of the TCC system.

A standardized environment, such as an operating system, that supports real-time requirements is a prerequisite for defining a standardized user interface. Specific standardized application program interfaces for TCC systems are also necessary under a standardized environment.

4.5 Timing-related issues

It is impossible to be sure that any action or any transaction may be completed in a given decay or at a given time. The delivery time is invariably a statistical quantity; only probabilities may be known. For time-critical communication, the delivery time cannot be known a priori as the time constraints of the application imply that the communication system itself must meet time constraints.

Two kinds of user requirements may be taken into account:

— time constraints attached to a message or a time-critical PDU;

— time constraints attached to a complex transaction.

This subclause deals with time aspects analysis.

4.5.1 Agreed probability of message transfer time or transaction completion time is necessary

4.5.1.1 Summary

The TCC system should be capable of transferring messages within a predetermined interval with a specified probability. In order to assess the suitability of a particular communication system for a specific application, it is essential to be able to establish that the communication system will be able to meet the application's timing constraints. In practice this will be an assessment of the probability that the timing constraints will be met. The application should include "fail-safe" procedures for use in the event of the network failing to deliver (or the receiving AE failing to act on) a time-critical message within the required time. It is necessary to be able to calculate the probability of failure per message.

4.5.1.2 Explanation

With a reservation media access control (MAC) strategy, an upper bound of delivery time can be determined on the assumption that no faults occur. With most contention MAC strategies there is no upper bound, although it may be possible to determine the probability that a message will be delivered within a certain time.

Even with a reservation strategy, there is a non-zero probability of a fault (e.g. a broken cable or a lost token) causing the delivery time to exceed the calculated upper bound. Therefore, whatever the MAC strategy, the performance shall be measured in terms of the probability of a message failing to be delivered within a certain time (see 4.6.2.1).

QoS characteristics need to be handled coherently through the OSI stack so that applications can specify parameters, such as the following:

— required message transfer time, allowing the network to distinguish between urgent and less urgent traffic;

— maximum tolerable probability of failing to meet this requirement, allowing the network to identify whether there is a requirement for retransmission, renegotiation of QoS, etc.;

— maximum throughput across a connection, allowing the lower layers to reserve a sufficient part of the network's resource, e.g. bandwidth for the traffic that the connection may be required to carry (the network may allow any of this "reserved" resource that is not required for time-critical messages to be used for other data flows such as file transfer and the downloading of programs).

Provision needs to be made for service providers to renegotiate the QoS after the connection has been made (because more needs have been added to the LAN segment or because the original route is no longer available). The values of those QoS parameters need to be related to the metrics derived from standard benchmarks or it will not be possible to assess the performance of a network with benchmark tests and then translate into QoS parameter values.

It may be appropriate that the differentiation of PDUs or the priority assigned to a PDU is ultimately handled as a QoS matter.

4.5.2 Temporal coherence is required and spatial coherence may be required in TCCA

4.5.2.1 Summary

A TCC system shall support timeliness and synchronization of information. The concept of temporal coherence is required in all TCC applications; the concept of spatial coherence is required in many TCC applications. Those important characteristics set TCC systems apart from other communication systems.

4.5.2.2 Explanation

4.5.2.2.1 Temporal coherence and timeliness

Temporal coherence is a property of a list of variables. It indicates whether or not the value of each variable of the list has been produced and transmitted and/or received in a given time window.

The TCC system shall support explicit timeliness attributes of information, for example, indications of temporal properties or values of different objects. This will then provide the receiving TCC entity with an indication of the time elapsed since the information was originally sampled from a process, product or network, or since the information was requested by a TCC entity. In many applications it is important to know whether or not values have been:

— produced at the same time, or

— transmitted and/or received at the same time.

Thus there are two distinct types of temporal coherence: a temporal coherence regarding the production of information, and a temporal coherence regarding the transmission of information.

Temporal coherence is not only important in information transfers but may also be important when there is a delivery deadline for a message. Timeliness attributes transfer the time of the system to meet delivery requirements. Solutions for temporal coherence could be absolute timestamping or relative timestamping (e.g. using refreshment and punctuality parameters).

4.5.2.2.2 A time-critical communication transaction

A time-critical communications transaction is a set of message transfers (or exchange) coming from difficult AEs which all have to be completed within a given time window.

In many applications it is important to be able to complete a number of message transfers within a given time window. For example, the MMS service “Read a list of variables” takes no account of time. Ideally it should be possible to provide a time-critical service which could specify some parameters so that the user could be certain that the variables are temporally coherent, i.e. that all the values of the variables were taken in a given time window. In addition the variables may be located in different VMDs and with MMS it will be necessary to use several different associations, and it will be important to be able to complete all the transfers within a certain time window.

A time-critical communications transaction is coherent with respected to production time when all the application’s messages or information units have been produced within a given time window.

A time-critical communications transaction is coherent with respect to transmission time when all the applications have been received within a given time window.

A time-critical communications transaction is globally coherent when it is coherent with respect to both production and transmission time.

4.5.2.2.3 Spatial coherence

Spatial coherence is a property of a duplicated list or of multiple copies of a list of variables. It indicates whether or not all the copies are identical at a given time or within a given time window.

Spatial coherence is useful when broadcast or multicast exchanges are necessary. It is necessary to provide such a facility if the user needs to know that systems support redundancy spatial coherence is necessary so that multiple controllers can have an identical view of the distributed control system and take appropriate action. Broadcasting or multicasting may allow TCC systems to offer point-to-point connections which require acknowledgments, but in such circumstances spatial coherence will be essential.

A time-critical communications transaction is spatially coherent when copies of messages, or list of variables are identical within a given time or time window.

4.5.2.2.4 Synchronization

The TCC system may frequently be required to support the synchronization of actions such as event triggering or information retrieval. It may be necessary for a TCC system to have a mechanism by which all remote TCC entities have an agreement of absolute or relative time, within some maximum jitter time under normal (non-fault) conditions. When this is required, maximum jitter time is generally much less than maximum transport time.

With this synchronization mechanism, a system designer can be sampled at the same instant in time (within the jitter specified). Given this mechanism, it is the responsibility of the other elements in the system to synchronize the sampling of information subsequently presented to the communication system.

The feasibility of this approach is demonstrated by existing networks. Some networks use a broadcast synchronization mechanism while other networks use timestamping. Accurate time stamping requires remote clocks to be periodically adjusted. This can be done by periodic broadcast messages, or by a reliable reference (e.g. a satellite clock) on each remote node. Synchronization is also important in TCC systems which have “sleepy devices”.

Applications using the TCC system may require access a common reference clock and a TCC entity may need to be able to adjust its clock to allow synchronization to the resolution required by the application.

4.6 Architecture considerations

In a flexible workshop, many events can occur, e.g. decisions from production management or unexpected failures. To adapt to these events, the TCC system must endorse this flexibility allowing for

— reconfiguring of the application when necessary, and

— authorizing configuration, maintenance, diagnostics and set-up messages, while controlling the membership of a TCCS to avoid, as much as possible, an overload which would jeopardize a time-critical message to be delivered within its prescribed time window.

4.6.1 Scheduling functions

The traffic on a TCCS is either time-critical or not time-critical. The time-critical traffic may be classified into different urgency classes. A scheduling function is then necessary. Some properties or characteristics of such scheduling functions are described below.

4.6.1.1 Differentiation of PDUs

4.6.1.1.1 Summary

In networks supporting time-critical communication, it is necessary to distinguish time-critical PDUs and non-time-critical PDUs. The time-critical PDU can be stamped with different relative urgencies.

4.6.1.1.2 Explanation

There are many possibilities for the most appropriate number of urgency levels. For example, two levels of priority might correspond directly with time-critical and non-time-critical PDUs, while three priority levels could distinguish between periodic time-critical PDUs, aperiodic time-critical PDUs and all other PDUs. Once a priority mechanism has been established it could be argued that more levels are desirable but this would probably greatly complicate the architecture and its management. The priority of the PDU needs to be clearly distinguished throughout its traverse of both stacks, since the time-critical PDU should be propagated by all layers in preference to less time-critical PDUs. Such pre-emption should occur

— at the data link layer, where preference is given to the highest priority PDU,

— at layers concerned with segmenting; time-critical PDUs should be propagated in preference to discrete frames of segmented large non-time-critical PDUs, and

— in other layers, where any opportunity for the pre-emption should be considered if this leads to better performance.

This strategy offers the virtual channel, in turn, to the hierarchy of priorities. Escalating priorities with time is undesirable since this will rapidly remove the differentiation of the traffic, e.g. under fault conditions. When there are a number of priorities only the highest level will get best determinism and this will be especially critical under fault conditions. If more than two priorities are supported by a TCC system, a mechanism should exist to ensure that all implementations of the architecture are able to treat identical priority levels in identical ways.

Mechanisms such as deadline scheduling may assist the management of priorities in TCC system or may be used as other ways of distinguishing the relative urgency of messages. If such mechanisms can only be used at the application layer it is essential that the lower layers can be relied upon to transmit the message within a given time window.

4.6.1.2 Non-time-critical PDUs shall not prejudice the delivery of time-critical PDUs

4.6.1.2.1 Summary

In practical applications time-critical and non-time-critical traffic will be required to share a single network for economic reasons. Development of tuning parameters, prioritization, etc., may allow this to work.

4.6.1.2.2 Explanation

An application's timing requirements for time-critical traffic should still be met if the offered load is at its maximum. However, a single non-time-critical transmission such as downloading a file on an otherwise idle network should be able to use a large proportion of the network bandwidth, which implies that non-time-critical traffic will be slowed down when the network is busy. At the moment, OSI stacks cannot take appropriate action if there is a conflict, i.e. where the delivery requirements of the time-critical PDU are incompatible with the baseload, PDU size, etc. of the non-time-critical PDU.

4.6.1.3 Static and dynamic scheduling

4.6.1.3.1 Summary

The sequence of the PDUs in a TCC system is determined by the application and circumstances, and is not completely prescribed.

4.6.1.3.2 Explanation

In a TCC system there will be periodic time-critical transfers, aperiodic time-critical transfers and non-time-critical transfers. The system shall operate with any combination of such transfers. The different applications using the TCC system will dictate the sequence of the PDUs.

4.6.2 Dependability

4.6.2.1 Message transfer time

The following factors will affect the message transfer time (depending on the architecture of the TCC system).

— The incoming application layer PDUs might well add latency to an outgoing PDU (since most stacks share one processor). While the TCC entity is always in control of the amount of outgoing traffic, in general it is not in control of the amount of incoming traffic and thus the stack can still be overwhelmed. Therefore, incoming time-critical PDUs should be removed such that flow control is not invoked.

— Changing the frame size will modify protocol efficiency. In a stack, large PDUs impede the flow of time-critical PDUs. If application PDU size exceeds frame size then segmenting, etc., is invoked with consequent time delays.

— Number of nodes (application entities) in the TCC system.

— Dynamic allocation of bandwidth is desirable since users will then get better service. Message transfer times can be measured experimentally (but it will often be impossible to find a representative system to measure or to vary the traffic enough to ensure that the "worst case" transfer has been found), heuristically by gathering statistics while operating, and by simulating temporal and procedural aspects to calculate performance. Standard metrics will be required for this, otherwise meaningful comparisons are impossible.

Figures 7 and 8 show how the proportion of messages received in a given time and the probability of a frame being received as a function of time can be measured.

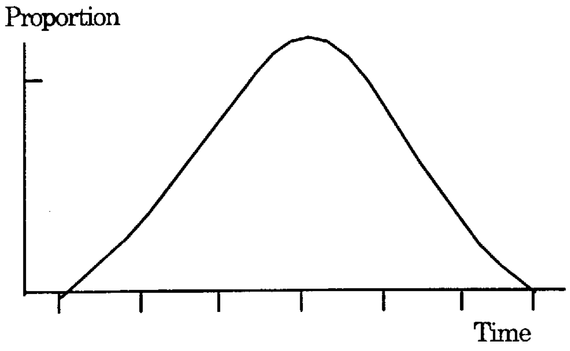

Figure 7 — Proportion of messages received in a given time

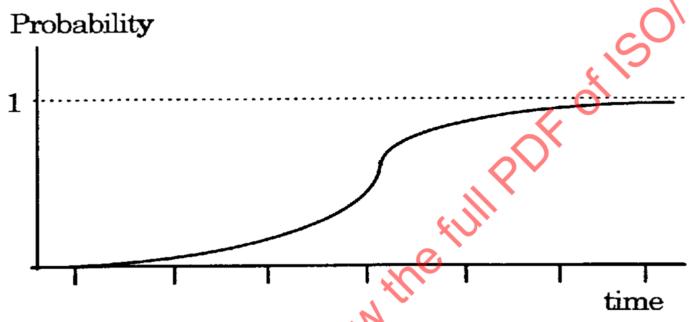

Figure 8 — Probability of frame being received as a function of time

4.6.2.2 Mechanisms to assist latent failure detection are desirable in TCCSs

4.6.2.2.1 Summary

Wherever possible latent communication failures, i.e, failures not yet apparent to participants, should be detectable to allow the highest probability of successful time-critical communication to be offered to the TCC entity.

4.6.2.2.2 Explanation

In a TCC system it is desirable to have monitoring mechanisms to allow the detection of failures or partial failures in components of the TCC system before these failures have affected the time-critical traffic on the TCC system. AEs in a TCC system may perform this role themselves or there may be special monitoring entities. Management functions will be necessary to administer such error detection processes and coordinate reporting of errors, etc.

4.6.2.3 Availability

4.6.2.3.1 Summary

Availability is defined as the probability of the communication system being in the function state. This can be expressed by the equation

$$A = \frac{MTBF}{(MTBF + MTTR)}$$

where

A	is the availability
MTBF	is the Mean Time Between Failures
MTTR	is the Mean Time To Repair

4.6.2.3.2 Explanation

It is essential that a TCC system offer a high degree of availability because the execution time of a service is lengthened by the recovery time (time to repair) if a failure occurs during execution.

4.6.2.3.3 Factors affecting availability

In a communication system two different failures can occur:

— permanent failure due to line or component failure, and

— temporary failures due to high noise on the line or other temporary cause such as lost or duplicated tokens.

It is desirable that the impact of a single fault does not modify the availability of the system. Three methods to improve availability are as follows.

— The “perfectionistic” approach. With this concept the mean time between failure will be lengthened by selecting components with lower failure rates or by choosing environments with lower noise level, or by choosing media with less susceptibility to environmental noise.

— The “fault tolerant” approach. With this concept the mean time to repair (to recover) will be reduced. The most promising way is to use redundancy. One example is code redundancy, which is used to recover from transmission errors. Another example is the use of duplicated transmission lines, where a failure of one line can be tolerated.

— The “fault containment” approach. With this concept the system continues to operate in the presence of any single fault. The system becomes unavailable only if a second fault occurs before the first is repaired.

The overall optimal solution may combine all three approaches.

4.6.3 Network management

4.6.3.1 Membership of a TCCS shall be controlled

4.6.3.1.1 Summary

It may be necessary to limit number of associations set up by AEs and/or to limit the load offered on them so that the network can carry time-critical traffic within prescribed time windows with the defined acceptable probability of success.

4.6.3.1.2 Explanation

Time-critical associations may be pre-arranged so that they can be available when required. This avoids the overhead of setting up an application association at the time a time-critical message has to be sent, and also allows resources to be reserved for such messages. Because resources should always be available for time-critical communications it is essential to control the number of TCC entities generating such traffic.

4.6.3.2 Reconfigurability of the application and the use of the TCCS

4.6.3.2.1 Summary

A TCC system may be require a preset time for applications to be activated or deactivated or modified.

4.6.3.2.2 Explanation

Many factory automation users require that an application be modified or initiated and/or terminated within a preset time window. Initiation may include, but not be limited to, network access, connection establishment, association negotiation, and application program download. Initiation is complete when the application program is fully supported by the TCC entity. Termination is complete when the application program is no longer supported by, and no longer affects, the TCC entity. Modification is complete when the application is operating in the changed state according to the modification request. It may be important that reconfigurability occurs in such a way that the performance of any stable regime is not degraded or that the requirement for dynamic reconfiguration occurs without disruption to the network. Metrics should enable the user to decide how quickly an application can be modified.

4.6.3.3 Off-segment communication

4.6.3.3.1 Summary

Off-segment communication can be allowed only if the application time constraints are not broken. There is clearly a requirement for off-segment communication for non-time-critical communication, e.g. configuration, maintenance, diagnostics, setup.

4.6.3.3.2 Explanation

Determinism on any time-critical segment ends when the propagation of the frame passes to another segment. In a bridge there is an implied application (the cloud) sitting above the two data link interfaces. The latency within the cloud may be indeterminate. In a router and a gateway the indeterminate latency is simply higher up the stack. A relay (bridge, router or gateway) in a TCC system segment will have a stack that is a normal conforming instance of the protocol in that segment. The relay can receive or dispatch time-critical traffic although no specific requirements for time-critical traffic through a relay seem to be needed in exactly the same manner as any other stack.

Inter-segment traffic cannot be allowed to enter a time-critical segment if this will impact on already promised performance or reserved resource. Control of such traffic would have to be a local issue.

4.6.3.4 Management and sovereignty

4.6.3.4.1 General

Detailed consideration of the management required for time-critical communications and discussion of sovereignty issues are not covered in this Technical Report. However, the following general observations are relevant to this review of user requirement for time-critical communications.

4.6.3.4.2 Management issues

Many of the identified requirements for time-critical communications will require network management to allow their implementation in a real system. Management of TCC entities in TCC groups and TCC systems will be necessary, and management of resources will be required. It is probable that automatic management functions will be required in TCC systems.

4.6.3.4.3 Sovereignty issues

TCC systems are groups of co-operating AEs that interact to form a regulated and stable system. Thus there is an overriding requirement for control in a TCC system to ensure that individual applications do not perturb the system with their requirements for time-critical communications, and to ensure that all members of a TCC group get the network performance and resilience that they need. It may be necessary to provide messages with temporal attributes and differentiation at all layers, and to have a controlling AE that identifies and qualifies the TCC entities in a TCC system and monitors and manages the network.

4.6.4 Some protocol implementation aspects

4.6.4.1 Applications requiring time-critical traffic shall not commit time-critical PDUs to the transmitting stack faster than it or the network or the receiving stack can handle the traffic

4.6.4.1.1 Summary

It is apparent that communication on a TCC system ceases to be determinate if its virtual channel capacity is overloaded.

4.6.4.1.2 Explanation

The problem of ensuring there is an upper bound on the demands made of a finite resource (communications) is not the responsibility of the communications system. In any TCC system it is a requirement that resources should always be available for time-critical communications. The systems designer should provide break rules for a network that will reflect the network constraints, e.g. bandwidth or throughput.

4.6.4.2 Classes of conformance to individual requirement level

4.6.4.2.1 Summary

Classes of conformance allow the TCC architecture to match individual application requirements.

4.6.4.2.2 Explanation

In some time-critical applications it is not essential for the TCC architecture to meet all the requirements. Therefore there will be identifiable classes of TCC architecture which meet specific subsets of the requirements. The identification of these classes is not within the scope of this Technical Report. Specific benchmarks may need to be developed to test TCC systems concerning conformance classes.

4.7 TCCS and OSI

In general, current OSI standards have concentrated on functional aspects to ensure intercommunication between end systems rather than performance issues. The performance agreement between OSI entities in current OSI standards are best-effort; that is, there is no assurance that a level of service will be provided, no undertaking to monitor it, or any remedial action taken if the desired level is not achieved. Under the current OSI environment, for a connection-less type of communication, QoS characteristics related to performance are a set of constant values given as prior knowledge. For connection-oriented communications, values are negotiated at the establishment stage of a context, or a connection, or an association, and these negotiated values are sustained until the end of the context, the connection, or the association between communicating peer entities.

On the contrary, a TCCA may require the capability to dynamically tune the values of QoS characteristics related to performance. The QoS requirements are represented, for instance, as a set of time windows and throughputs of message transfers over the network system, which are derived from time constraints specified by end users. Another prerequisite for a TCCA is the ability to support co-operation among multiple parties, i.e. time-critical communications entities (TCCEs), distributed over the whole TCC network system. This cooperation is on a multi-peer communications basis rather than a peer-to-peer basis. Further requirements concern monitoring and controlling the network behaviour and resources. The network management aspects in TCCA are focused on these issues.

Therefore, user requirements for TCCSs require additional network management functions. These functions are required to manage the time-window constraints specified by the user. The translation of the user requirements into functional requirements as network management functions is necessary to support time critical communications. TCCAs are distinguished from standard OSI architectures by their support for these additional network management functions.

4.8 TCCA network management

The behaviour of a time-critical communications network system is affected by a number of factors which are the combinations of unpredictable or unstable factors and predictable or stable factors, e.g. sporadic traffic, collisions, temporary resource congestion, group membership change, and so on.

The performance of a time-critical communication network system, for instance, a transfer completion time, may be assessed in the terms of comparisons between the predicted, the requested and the actual completion time. These differences exist in OSI because the requested completion time is negotiated between users and providers, and the operational environment is always probabilistic and not completely deterministic. In the case of failure, the TCCS must support notification to the TCCE within specified time windows. Such examples show that existing systems management are insufficient to accommodate every event having specific time constraints.

Ultimately, TCCS functionality should be developed to provide the capability to meet the time-window requirements specified by the user with a high degree of probability even in a worst-case scenario. Additionally historical information can be used to predict that performance in a similar environment is acceptable to the user. Network management of a TCCS provides the capability to meet QoS requirements derived from time-window-related constraints.

4.9 Network management in TCCN

In future manufacturing environments a wide variety of real-time applications groups will be linked to a communications network system that will be installed as a highly reliable communication infrastructure to integrate these applications groups. In such an operational environment it is extremely difficult to predict precisely the behaviour or the characteristics of the whole network system. Where a number of cooperating applications groups are distributed over the network system, the loading condition over the network system will vary dynamically due to configuration changes. The network load will change as the application groups are activated or deactivated dynamically according to the needs of different applications, and as participating nodes are added or removed.

Therefore, at every stage in design, operation, maintenance, fault repair, etc., there will be a growing demand for some facilities to assist decision making to cope with the dynamic evolution of a network system in order to meet application-dependent time constraints. Such facilities can allow network system designers to construct new network systems to meet their customers' requirements and to add or change applications processes over an existing network system without disrupting the system. This will allow system managers to change configuration during operation and to maintain highly reliable network systems. Such facilities will benefit network system users in a wide spectrum of applications.

Network management mechanisms in time-critical communications will ensure exchange of time-critical messages within the defined time windows. Evaluation of information provided by these mechanisms should enable users to adapt their systems to meet future application requirements.

5 QoS requirements for network management in TCCA

5.1 Introduction

The QoS requirements on network management in a TCCA express a set of qualities related to the provision of network management services as perceived by users. These QoS requirements on network management, which are extracted and translated from the user requirements stated in clause 4, are especially concerned with the automatic control of the characteristics of a time-critical communications network system. These requirements are essential in order to meet the time constraints specified by users under circumstances in which the behaviour of the time-critical communications network system varies dynamically.

The QoS requirements, in a QoS Basic Framework sense, express part or all of the requirements to manage one or more QoS characteristics, and express the trade-off between the QoS characteristics. Further, the requirements are expressed in terms of QoS parameters as part of a mechanism when a QoS requirement is conveyed between entities.

These requirements form a policy that is to apply to the operation of a time-critical communications network system. The topics related to policy, which will determine the QoS characteristics and the QoS management functions, are also contained in the network management requirements. The characteristics and the management functions will be identified in clauses 7 and 8.

In some time-critical applications all the user requirements identified in clause 4 need not always be satisfied. It is therefore not essential for a TCCA or the network management in a TCCA to meet all the network management requirements identified in this clause. This means that there will be allowable classes which meet specific subsets of this requirements. Identification of these classes, however, is out of the scope of this Technical Report.

5.2 Network management requirements

5.2.1 Adaptability to events requires parameters attached to PDUs to express an absolute degree of urgency

As there is a duality of the state-driven and event-driven in the control paradigm of existing real-time systems, this same duality should be supported in a time-critical communications network system. This duality will be used in the majority of future network systems to be applied to real-time systems, e.g. aperiodic time-critical communications coping with events which are sporadic in nature, such as an alarm, warning, state-change, processing completion, request processing and so on, and periodic time-critical communications which are pre-scheduled.

Even in the case of definite loading of periodic time-critical communications in which the constraints are to deliver to its destinations a number of PDUs within each cycle period, a time-critical communications network system as a control network system should not compromise its capability to adapt to aperiodic time-critical communications which are generated at the instance of events that have occurred in applications processes distributed over the whole network system.

The policy of adapting to a time-critical PDU, which has been generated in response to an event that has occurred over the entire time-critical communications network system, is that the precedence of the PDU transfer can be decided on the basis of the absolute degree of urgency of the PDU. This policy is to be applied under circumstances in which periodic and aperiodic time-critical PDUs coexist. The absolute degree of urgency of the PDU is closely related not only to the required completion time specified by the user, but also to a time window for transfer nominated by communication entities. The time window to be nominated is dependent on the condition of the time-critical communications network system in operation, and then is to be varied and to be re-nominated at each communication entity in progress.

This policy is to be applied equally to both periodic and aperiodic time-critical PDUs in transfer processing, and the differentiation between them is to be done in accordance with the difference in the degree of urgency between them.

5.2.2 An expected completion time can be offered for comparison with the required completion time contained in a PDU

There is never a completely deterministic process over a communications network system. There is always a non-zero probability of a fault causing an expected completion time to be exceeded or an expected action not to be completed. Used as a control network system, a time-critical communications network system will have the capability to offer an expected completion time for transfer to the user. In an OSI sense the user can determine whether delivery of a PDU to its destination can be achieved within the required completion time for transfer. The user can then decide whether to give up the transfer request and take alternative actions. Where an offered expected completion time for transfer happens to be much longer than the required completion time, a persistent request for transfer can potentially cause a crucial perturbation to the existing operation.

A certain parameter that signifies the required completion time specified by the user should be contained in a message or a PDU and conveyed between communication entities in order to manage the transfer processing.

5.2.3 A time-critical PDU is allocated by the user and can be pre-empted against non-time-critical PDUs in transfer processing

An application process determines from its needs (which are dependent on operating circumstances) whether or not a message is to be assigned as a time-critical one. This implies that it is not always true that a message, which completely contains the same information as that in a time-critical message already sent, is also a time-critical one.

It is therefore necessary for a message or a PDU to be explicitly indicated as a time-critical one. It is then required that a certain parameter attached to the PDU indicates whether the PDU is a time-critical one or not.

Once a message has been assigned as a time-critical one, the message is differentiated over the whole time-critical communications network system and is delivered to its destination so as to meet the required completion time for transfer. This is based on a policy of differentiation that pre-empts this time-critical PDU against non-time-critical PDUs in transfer processing.

5.2.4 A specific time frame can be nominated and given as a tentative time window for starting transmission of a PDU by a scheduling function

It is necessary to receive an acknowledgment from a destination within a given time window. With temporal and spatial coherence as a fundamental property in a time-critical communications network system, it is essential to have timely receipt of PDUs not only at a single destination, but also at multiple destinations. Especially in the delivery of PDUs to multiple destinations the capacity of transfer processing and the distances between the destinations are always different. This implies that timeliness of starting the transmission of PDUs at the sending side is essential for the timeliness of receipt of the PDUs at the receiving sides.

Some kind of scheduling function should be needed to estimate the expected transfer time(s), to determine the transfer order of PDUs, to nominate a specific time frame to be used to deliver the PDUs to destinations, and then to manage the to start of the transmission of PDUs.

The scheduling policy is to be based on fulfillment of transmission timeliness. If the transmission timeliness cannot be guaranteed, the PDUs are to be discarded because they have no value when there is no receipt timeliness.

Figure 9 shows the relationship between the time window and the nominated time frame.

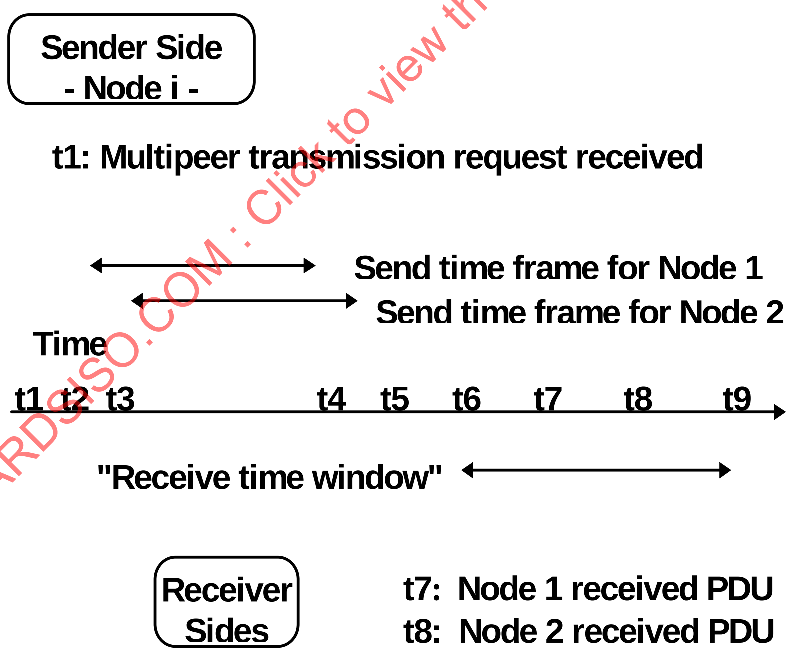

Figure 9 — Relationship between time window and nominated time frame

5.2.5 Best use of whatever possible opportunity there is to fulfill the requirements

The disposal of PDUs with a same degree of urgency is a matter of policy concerning communications scheduling of a time-critical communications network system, e.g. first in first service, last in first service, service at random, and so on. The degree of urgency of a PDU stems from the combination of the specified time constraints by the user and the remaining time to its deadline for the transfer of the PDU to its destination. The deadline time is dependent on the operational conditions of a time-critical communications network system. The degree of urgency of a PDU therefore naturally changes according to the operational conditions of a time-critical communications network system.

In the case of delivering a PDU to its destination in which multiple routes or communication channels for delivery are existing and available, one of the most effective methods is to choose the best route or channel for completing the delivery of the PDU to its destination within the time window. If there happens to be a collision of two PDUs with the same degree of urgency, it is required to give a higher tentative degree of urgency to one of them in order to achieve the delivery of both PDUs to their destinations within their time windows.

This policy is to be applied at every stage in communications scheduling or processing in order to increase the possibility of succeeding in the delivery of a PDU to its destination within a time constraint specified by the user.

Figure 10 shows an example of the possible choices of routes from node 1 to node 3. Figure 11 shows an example of relationship between a change in the time window and the degree of urgency of a PDU.

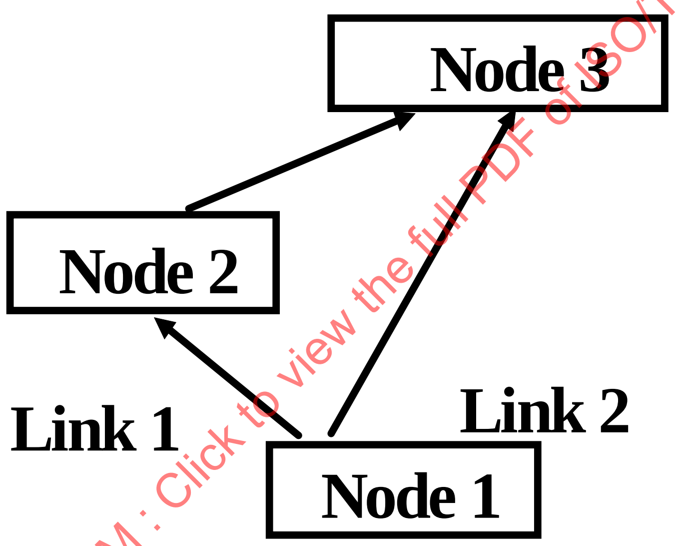

Figure 10 — Possible choice in multiple routes

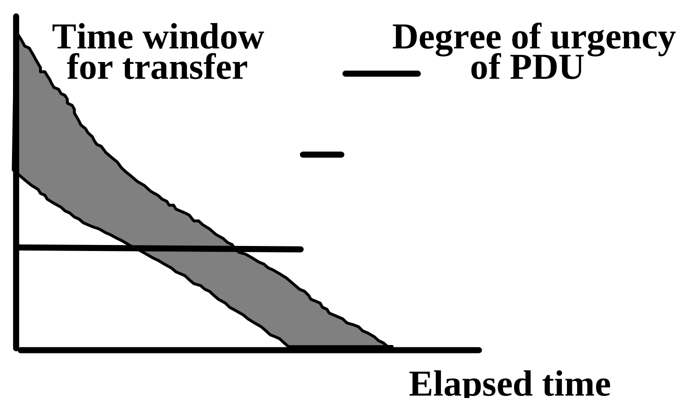

Figure 11 — Change of time window and degree of urgency

5.2.6 Use of a common time clock for synchronization between entities over a whole time-critical communications network system

There is a requirement to provide local time facilities in a time-critical communications network system. The time clocks of these local time facilities are synchronized with some maximum difference and jitter. One of the time clocks is accommodated as a common time clock to communication entities over a whole time-critical communications network system, and then the communication entities can use this common time clock to synchronize their actions, such as event triggering or information retrieval.

5.2.7 A time-critical communications request should be issued without a retransmission request

Whenever there is a failure to complete transfer to a destination within the expected time it is important for the sender of a PDU to receive a failure notification PDU. The user, in an OSI Basic Reference Model sense, can then determine whether a new attempt to retransmit the PDU can be undertaken or not. However, it will be very complicated for an application user to obtain appropriate information on transfer time, to calculate the expected completion time to deliver the PDU to its destination under a variety of conditions, to consider some sequences of applications processes to be executed within an expected time, and then to allocate time frames to a number of re-transmission attempts.

It is simpler and more effective for an applications user to always initiate a transfer request without retransmission in a time-critical communications network system. If the applications user wants to retransmit a same PDU upon receipt of a failure notification PDU, the applications user should attempt to reinitiate the transfer request according to the applications needs.

5.2.8 Ability to make decisions by using information which is common to all participants over the whole time-critical communications network system

The behaviour of a time-critical communications network system is a collection of communications interactions between all participating entities over the whole network system. It is essential that the participating entities co-operate to share common knowledge among them. This common knowledge consists of historical information on events, variables, variable changes, parameters, parameter changes, state changes, statistics, configuration changes and so on, that have happened over the whole network system.

Network management in a time-critical communications network system make decisions using this common knowledge at every necessary stage in order to meet user requirements. In the case of the set-up stage of an application association when a time-critical PDU has to be sent, this common knowledge can be used to limit the overhead of setting up a number of associations by communication entities. Further, some load shedding can be achieved by regulating a volume of the traffic at every service boundary in a time-critical communications network system using this common knowledge in order to avoid the traffic congestion. Thus, a certain portion of the total resources over a whole network system can be always available to a number of time-critical PDUs.

Figure 12 shows the relationship between a globally common knowledge and the entities within participating nodes over a time-critical communications network system.

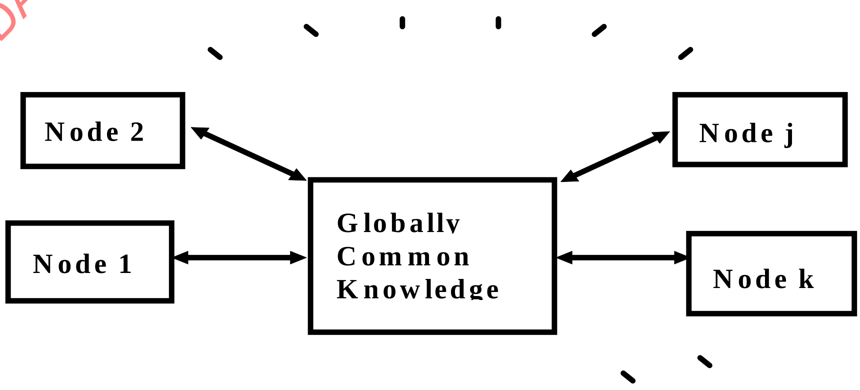

Figure 12 — Relationship between globally common knowledge and the entities within nodes

The historical information is used to calculate an expected completion time to deliver a PDU to its destination. For this calculation, the historical information contains the following time values concerning a PDU delivery to its destination over a time-critical communications network system: the inbound upper layer stack time(s) (t6, t7 of Figure 13); the outbound upper layer stack time(s) (t1, t2 of Figure 13); the inbound datalink time(s) (t5 of Figure 13); the outbound datalink time(s) (t3 of Figure 13); the signal propagation time(s) in both direction between nodes including the pass-through time within intermediate system on its route (t4, t9 of Figure 13). The outbound datalink time contains the time to media access, and the collection of the signal propagation times contains the time of every alternative route between a sender node and a destination node.

These values are changing dynamically according to the operational conditions in a time-critical communications network system. These values are therefore stochastic rather than determinate.

Figure 13 shows the time variables related to a PDU delivery to it destination.

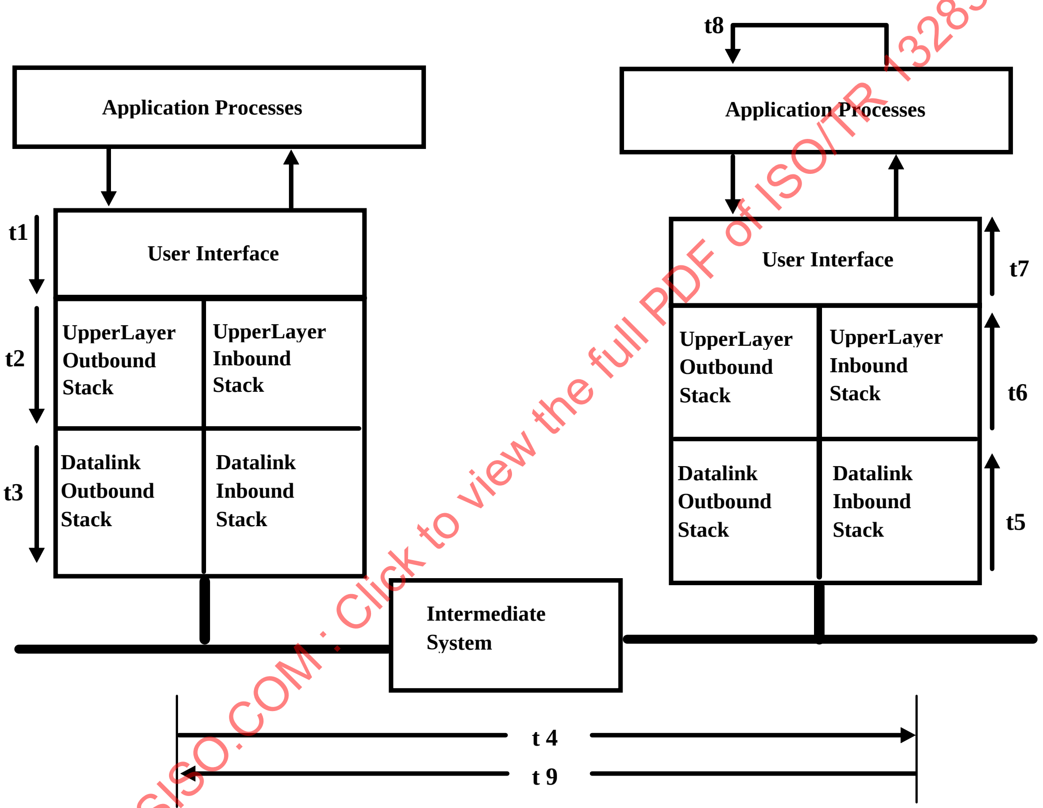

Figure 13 — Time variables related to a PDU delivery to its destination

5.2.9 Ability to perceive latent failures in progress

Facilities are required to provide detection of latent failures. This implies the ability to perceive some symptoms, estimate their potential, infer its influence and, when necessary, support the taking of alternative actions or fail-safe actions in advance to avoid completion failures of time-critical PDU transfers.

6 QoS model of TCNM

6.1 Introduction

The QoS model in TCNM defines the architectural concept, principles and structures of the network management in a TCCA. The model provides a common view of the global system work of the network management in a TCCA. The model itself does not specify any of the QoS parameters or QoS information that are exchanged during system operation.

6.2 Architectural principles

The model considers two classes of entities that take part in the network management of TCCN or the management of QoS in TCNM in open systems:

— system TCNM QoS entities which have a system-wide role;

— local TCNM QoS entities associated with the operation of a particular subsystem.

The local TCNM QoS entities are those that coordinate the response to the requirements imposed on the TCCN or the TCCS. The system TCNM QoS entities interact with the local TCNM QoS entities to monitor and control the performance of the TCCN system. The system TCNM QoS entities implement managed objects as a means by which the system TCNM QoS entities interact with the provision of QoS in the TCCN system.

The local TCNM QoS entities implement direct control of protocol entities that are necessary for the support of QoS requirements on the TCCN system. In doing so, they respond to the control imposed on them by the system TCNM QoS entities and interact with the distributed local TCNM QoS entities in the TCCN system.

Figure 14 shows the relationship between the system TCNM QoS entities and the local TCNM QoS entities. In addition, in an OSI Basic Reference Model sense, the system TCNM QoS entities have a relationship with the TCCS user and the local TCNM QoS entities have a relationship with the (N)-service-user and (N)-service-provider.

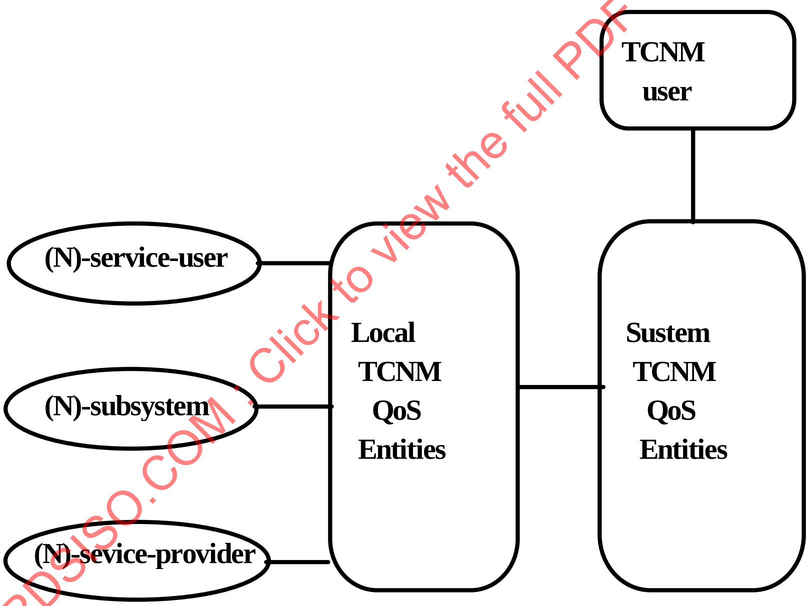

Figure 14 — The relationship of system to TCNM QoS Entities

Another view of the model of QoS in TCNM is represented by the analogy to a closed loop control system in automatic control theory. A system in which provision of services may respond to the requirements from a superior service user, implements measures for specifying, controlling, and monitoring the QoS on the connections with the requirements. The TCCN is considered as a system being controlled by the provision of services in TCNM in order to fulfill the requirements. These requirements represent the system policies that establish the constraints under which all operations of the system may perform.

Figure 15 shows the overall model of the way in which the provision of services in TCNM responds to the QoS requirements of the superior service user. There are two overall inputs to TCNM:

— the TCNM user’s QoS requirements which establish initial conditions for the provision of services;

— the disturbances which counteract a specified QoS to be maintained.

The QoS management inside the TCNM represent the control entities in which the QoS parameters of the QoS characteristics are regulated and outputted in order to fulfill the specified QoS requirements. The QoS management performs monitoring of all the possible measures of the QoS characteristics related to the working of the TCCN. The QoS measures are the observed values relating to the QoS characteristics providing feedback which causes the QoS management to regulate or, as necessary, re-establish the QoS parameters under which it operates.

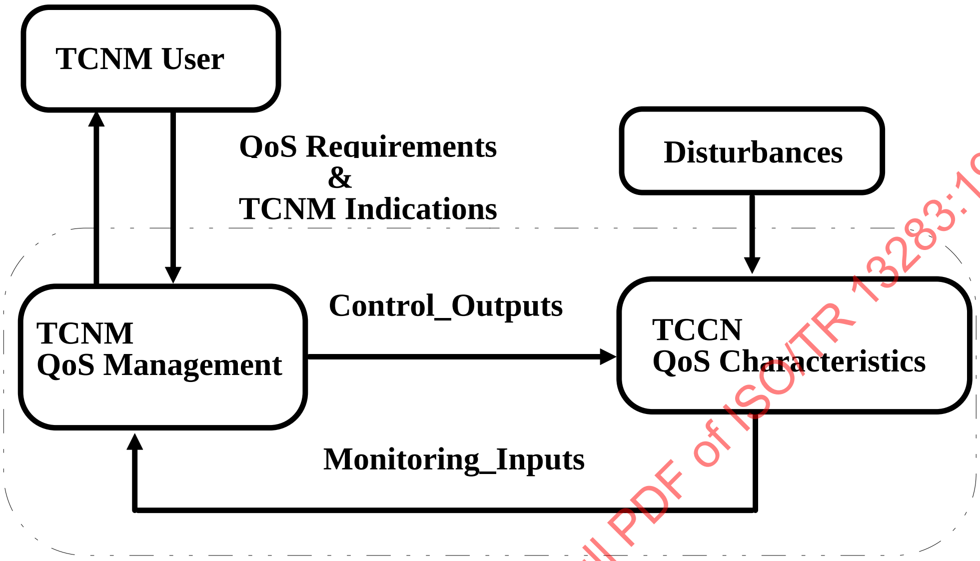

Figure 15 — Closed loop control of QoS in TCNM

6.3 Information exchanged

The TCNM QoS-related information is classified into the following kinds:

— the TCNM QoS context which is retained in an TCNM entity;

— the TCNM QoS parameters which is conveyed between TCNM entities.

The TCNM QoS-related information which is conveyed between TCNM QoS entities, i.e. the TCNM QoS parameters, is classified into the following kinds of elements:

— the TCNM QoS requirements parameters which are statements of requirements on one or more TCNM QoS characteristics;

— the TCNM QoS data parameters which are the TCNM QoS information other than the TCNM QoS requirements parameters, and includes the measures, the indications of status, the notifications of warnings, the requests for information, etc.

The QoS requirements are conveyed between the related QoS entities and results in them being accepted, rejected, or modified to a more suitable form. The results are reported back to the source of the requirements in the course of operation.

6.4 QoS requirement flow

6.4.1 QoS requirement flow in an (N)-subsystem

In general the QoS-related information is exchanged at an (N)-service boundary between (N)-service-user and (N)-service-provider for unconfirmed and confirmed (N)-service-facilities respectively, and also is exchanged between and within (N)-subsystems as both the outgoing and the incoming flow.

The outgoing flow of the QoS requirements is to be passed to the (N-1)-subsystem across the (N-1)-service boundary or to be conveyed to the peer (N)-subsystem in (N)-protocol. On the contrary the incoming flow of the QoS requirements is to be received from the (N-1)-subsystem across the (N-1)-service boundary or to be received from the peer (N)-subsystem in (N)-protocol. Each of the flows has a reverse flow to respond respectively when the entity receives QoS requirements.

The (N)-subsystem is modeled as a composition of (N)-PCF, (N)-QCF and (N)-PE, which are described below.

— The (N)-PCF determines policy that is to apply to the operation of the (N)-subsystem. It determines the constraints under which all other decisions of the (N)-subsystem are made. Any action is performed to control the remaining operation of the (N)-subsystem.

— The (N)-QCF takes account of QoS requirements in selecting the entities that will participate in communications. It may represent a point of interaction at which knowledge of QoS may be used to influence the choice and at which the consideration of the choice may affect the treatment of QoS.

— The (N)-PE is responsible for operating the (N)-protocol in order to provide the (N)-service to the (N)-service user and is also responsible for operating with its peer (N)-protocol-entity, its peer (N)-service user and the (N-1)-service provider.

Figure 16 shows the flow of the QoS requirements between and within (N)-subsystems. The flow between (N)-subsystem and (N-1)-subsystem is regarded as being conveyed in (N-1)-service primitives whether it is modeled as passing between (N)-PE and (N-1)-PCF or between (N-1)-PE and (N)-PCF.

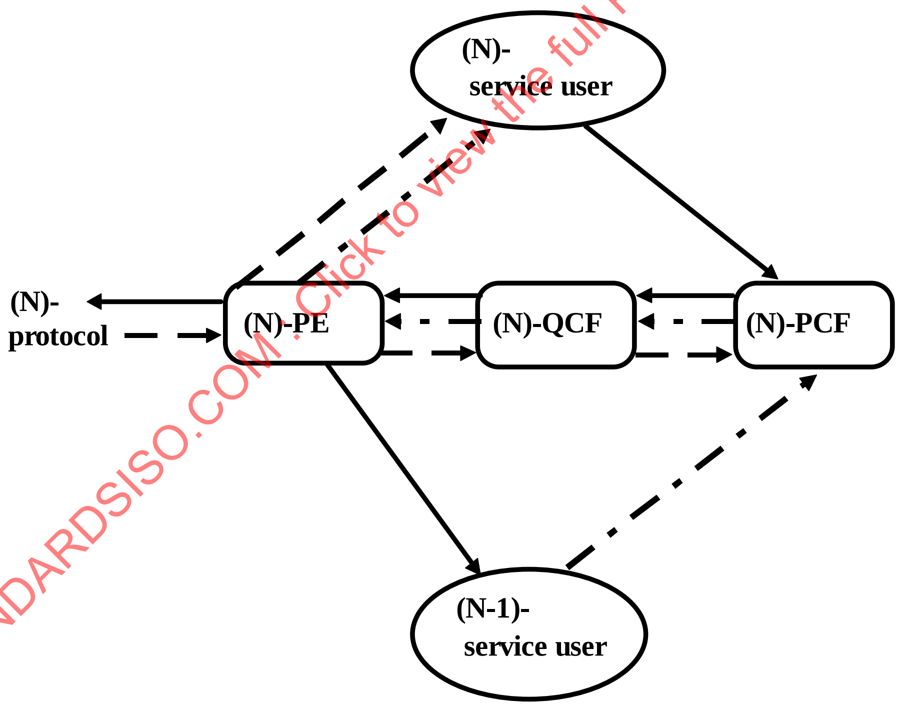

Figure 16 — The flow of QoS requirements within and between (N)-subsystems

6.4.2 QoS requirements flow between (N)-subsystem and TCNM

The TCNM QoS related information is exchanged at the (NM)-service boundary between an (NM)-service user and the TCNM-subsystem for unconfirmed and confirmed service facilities respectively, and also is exchanged between the TCNM-subsystems as both the outgoing and the incoming flow. Further, the outgoing and the incoming flow of the QoS information implies interactions to be exchanged between the (N)-subsystem and the TCNM-subsystem in order to fulfill the QoS requirements imposed on the TCCN.

The operation of the TCNM-subsystem will perform autonomously, though cooperating with other TCNM-subsystems in managing the whole network system, and the TCNM-subsystem may or may not be decomposed into sub-layers or partitions according to the organization of the open system as an aggregation of (N)-subsystems to be managed by itself.

The TCNM-subsystem is modeled as a composition of (NM)-PCF, (NM)-QCF, and (NM)-ME, which are described below.

— The (NM)-PCF determines the TCNM policy that is to apply to the operation of the TCCN. It determines the constraints under which all other decisions of the TCNM-subsystem are made, and the constraints affect the (N)-PCF in the (N)-subsystem in order to fulfill the QoS requirements imposed on TCCN. Any action is performed to control the remaining operation of the TCNM-subsystem.

— The (NM)-QCF takes account of QoS requirements in selecting the entities that will participate in the network management. It represents a point of interaction at which knowledge of QoS will be used to influence the choice and at which the consideration of the choice will affect the treatment of QoS.

— The (NM)-ME is the management agent entity responsible for operating management actions, e.g. state control and reporting, alarm reporting, logging, violation reporting, access control, monitoring metrics, test control, scan reporting, traffic scheduling, resource control, knowledge management. In an OSI System Management sense, the (NM)-ME uses the (NM)-protocol in order to cooperate with other peer TCNM-subsystems. The (NM)-ME combines two function classes that represent both a system-wide and a locally restricted capability for tuning and coordinating the behaviour of the various QoS entities in operation in distributed TCNM subsystems.

The outgoing flow of the QoS requirements is to be passed across the (NM)-service boundary down to the (NM)-PCF, (NM)-QCF and (NM)-ME, and conveyed to the peer TCNM subsystem in (NM)-protocol. On the contrary the incoming flow of the QoS requirements is to be received from the peer TCNM subsystem in (NM)-protocol and to be passed up to the (NM)-PCF. The QoS requirements modified within the TCNM subsystem are to be passed to the (N)-PCF within the (N)-subsystem by the (NM)-PCF. Figure 17 shows the flow of the QoS requirements between the (N)-subsystem and TCNM subsystem, and also the flow from the TCNM subsystem to (N)-subsystem. Each of the flows has a reverse flow to respond respectively when the entity receives QoS requirements.

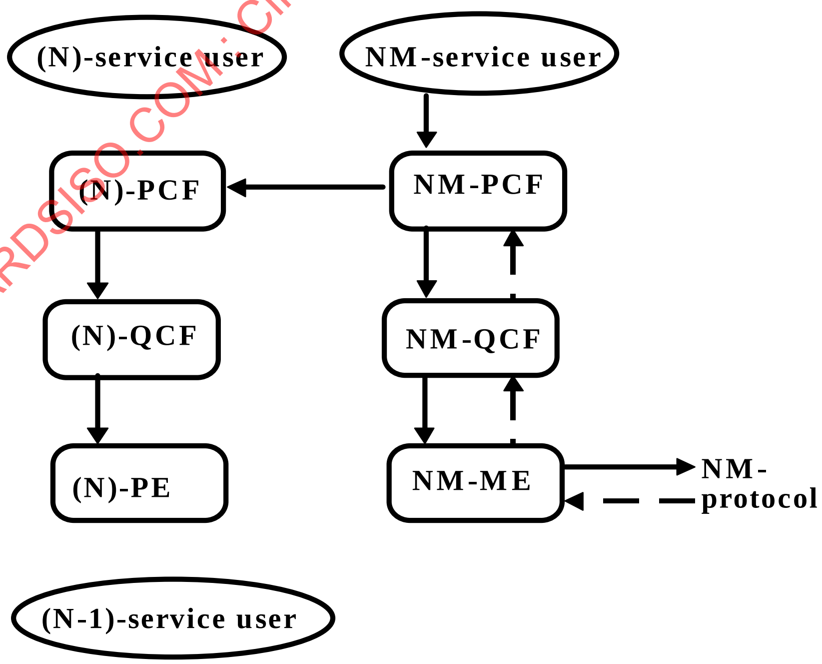

Figure 17 — The flow of QoS requirements between the (N)-subsystem and TCNM

6.5 QoS data flow between (N)-subsystem and TCNM

The data flow between the (N)-subsystem and the TCNM-subsystem does not interact with them directly, but instead interacts via the TCNM-MIB. The TCNM-MIB represents a point of interaction at which common knowledge of the behaviour of the various QoS entities in operation will be used to influence the policy control, the QoS control, the protocol processing, and the management agent processing inside both the (N)-subsystem and the TCNM-subsystem. The interactions via the TCNM-MIB will occur at any time.

In an OSI System Management sense per ISO/IEC 10040 and ISO/IEC 10165, the management information model of the TCNM-MIB is basically similar to, or may be included in, that of the MIB. The composition of the TCNM-MIB is a set of the structures of the objects to be managed which includes the QoS data parameters to be exchanged both inside and between the TCNM-subsystems and also between the (N)-subsystem and the TCNM-subsystem, which retains the QoS context and access information.

The managed objects in the TCNM-MIB are used for state control and reporting, alarm reporting, logging, violation reporting, access control, monitoring metrics, test control, scan reporting, traffic scheduling, resource control, etc. The TCNM QoS mechanism will monitor and even control the remote behaviour of these operating management actions using the managed objects.

Figure 18 shows the relationship in the QoS data flow between the (N)-subsystem and the TCNM subsystem and also the QoS data flow both inside and between the TCNM subsystems.

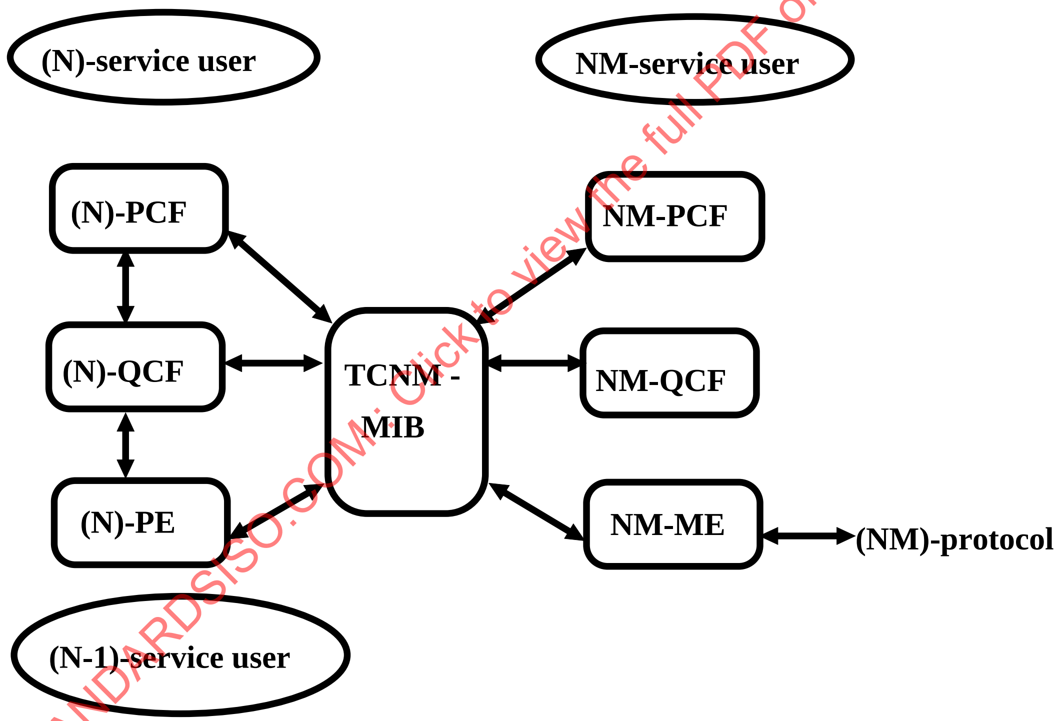

Figure 18 — The relationship in QoS data flow between the (N)-subsystem and TCNM

7 QoS characteristics of TCCA

7.1 Introduction

QoS characteristics of a TCCA correspond to a fundamental aspect of QoS that can be quantitative and that is to be managed. These characteristics also represent the true underlying state of affairs or are used as a description of the actual behaviour of a whole network system under a time-critical communications environment.

A TCCS will be designed, procured or configured to operate with some kind of policy for QoS requirements. The policy will determine the QoS characteristics and the QoS management functions to fulfill the QoS requirements that are placed on one or more QoS characteristics and to express the trade-off between them.

7.2 QoS characteristics

A QoS characteristic is attached to a service request or a set of requests by a given user. This subclause defines QoS characteristics related to elapsed time or integrity of data associated or concerned by the requests.

7.2.1 Date/time characteristic

7.2.1.1 Definition

This characteristic is defined as the absolute time when an event occurs.

7.2.1.2 Quantification

This characteristic is quantified as any time unit such as minute, second, millisecond, etc. with reference to a known time-origin.

7.2.1.3 Specializations

Characteristics are specialized from date/time by specifying the particular event, the unit/origin of measurement, and/or any statistics that are to be applied.

7.2.2 Time delay characteristic

7.2.2.1 Definition

This characteristic is defined as the elapsed time (T2 – T1) between two general events, E1 and E2, which occur at times T1 and T2.

7.2.2.2 Quantification

This characteristic is quantified as any time unit such as minute, second, millisecond, etc.

7.2.2.3 Specializations

Characteristics are specialized from time delay by specifying the particular event, the measurement, and/or any statistics that are to be applied.

7.2.2.4 Statistical derivations

Any statistical derivations may be applied to the time delay characteristic defined above. For example:

— for a transit delay, E1 is the passage of the first item past date point P1 and E2 is the passage of the last item past date point P2;

— for a request/reply delay, E1 is the passage of the first item past date point P1 and E2 is the passage of the last item of the related reply past point P1;

— for a request/confirm delay, E1 is the passage of the first item past date point P1 and E2 is the passage of the last item of the related (intermediate) confirmation past point P1.

NOTE Jitter is of particular importance, and is defined as the range (minimum to maximum) of a specific time delay over the duration of the communication. For example, jitter is important for data streams in which only small variations in transit delays can be tolerated without significant loss of service to the end user.

7.2.3 Life time validity characteristic

7.2.3.1 Definition

This characteristic is defined as the duration that the date is valid.

7.2.3.2 Quantification

This characteristic is quantified as a time interval.

7.2.4 Remaining lifetime characteristic

7.2.4.1 Definition

This characteristic is defined as the time remaining before the date ceases to be valid.

7.2.4.2 Quantification

This characteristic is quantified as a time interval.

7.2.5 Temporal coherence characteristic

7.2.5.1 Definition

This characteristic is defined as the indication whether the same action has been performed on each entity, data value etc. in a list within a given time window.

7.2.5.2 Quantification

This characteristic is quantified as a Boolean value, either “true” or “false”.

7.2.5.3 Specializations

There is a range of possible further specializations for the temporal coherence characteristic including temporal data production coherence, temporal data transmission coherence and temporal data consumption coherence.

7.2.6 Temporal data production coherence characteristic

7.2.6.1 Definition

This characteristic is defined as the indication whether the value of each variable in a list has been produced within a given time.

7.2.6.2 Quantification

This characteristic is quantified as a Boolean value, either “true” or “false”.

7.2.7 Temporal data transmission coherence characteristic

7.2.7.1 Definition

This characteristic is defined as the indication whether the value of each variable in a list has been transmitted within a given time window.

7.2.7.2 Quantification

This characteristic is quantified as a Boolean value, either “true” or “false”.

7.2.8 Temporal data consumption coherence characteristic

7.2.8.1 Definition

This characteristic is defined as the indication whether the value of each variable in a list has been consumed within a given time window.

7.2.8.2 Quantification

This characteristic is quantified as a Boolean value, either “true” or “false”.

7.2.9 Spatial consistency characteristic

7.2.9.1 Definition

This characteristic indicates whether or not all of a duplicated list or multiple copies of a list of variables are identical at a specific time or within a given time window.

7.2.9.2 Quantification

This characteristic is quantified as a Boolean value, either “true” or “false”.

7.2.9.3 Specializations

There is a range of possible further specializations for the spatial consistency characteristic, including timeless spatial consistency, temporal spatial consistency, etc.

7.2.10 Timeless spatial consistency characteristic

7.2.10.1 Definition

This characteristic is defined as the spatial consistency where time is not an issue.

7.2.10.2 Quantification

This characteristic is quantified as a Boolean value, either “true” or “false”.

7.2.11 Temporal spatial consistency characteristic

7.2.11.1 Definition

This characteristic is defined as the spatial consistency within a given time window or at a specific time.

7.2.11.2 Quantification

This characteristic is quantified as a Boolean value, either “true” or “false”.

7.2.12 Capacity

7.2.12.1 Definition

This characteristic is defined as the time amount of service that can be provided in a specified period of time.

7.2.12.2 Quantification

Since the capacity characteristic can be applied to different types of resources, it is quantified using various units. The quantification of capacity also depends on the unit of time used for the measurement.

7.2.12.3 Specializations

Throughput is a specialization of the capacity characteristic.

7.2.13 Throughput (communications capacity) characteristic

7.2.13.1 Definition

This characteristic is defined as an amount of information transferred through a channel averaged over a specified period of time t.

7.2.13.2 Quantification

This characteristic is quantified as a rate such as bits per second or bytes per second.

7.2.13.3 Specializations

There is a very wide range of possible further specializations for the communications throughput characteristic, including user information throughput, application information throughput, subsystem throughput, processing capacity, and instructions per second.

7.2.13.4 Statistical derivations

There are two types of statistical derivations for the communications throughput characteristics. These are:

— out-bound throughput, i.e. the amount of outgoing information transferred through a channel averaged over a specified period between (N) and (N-1) layers and/or over (N)-associations among end-to-end systems;

— in-bound throughput, i.e. the amount of incoming information transferred through a channel averaged over a specified period between (N) and (N-1) layers and/or over (N)-associations among end-to-end systems.

NOTE Throughput must be a rate. “User data” should be made explicit in order to differentiate it from control data used to manage the channel (e.g. flow control window parameters). The point at which the rate is defined must be identified, and it is necessary to identify the time interval over which the rate is defined. By varying t one can deal with cases where throughput must be maintained essentially constant (t small) as well as those where wide variation is permitted so long as a long-term average is met (t large), e.g. video streams versus typical packet-switched data. Some of the “specifiers” defined for time delay can also be applied to throughput. “Process” can be used to distinguish between bits/bytes or packets/frames/cells. “Location” can define the channel, e.g. OSI (N)-SAP to peer (N)-SAP. “Operation” can be the type of data in question. If it is important to ensure that queuing or other indeterminate delays do not occur when user information is transferred, then users need to be able to request their maximum user information throughput for each class of traffic (i.e. degree of urgency) so that they can ensure that determinism is not degraded. Where throughput is affected by load in the communications system, or parts of the system such as the communications stacks or in the network, it is essential for some users to be able to quantify the effects of loading on the throughput.

7.2.14 User information throughput characteristic

7.2.14.1 Definition

This characteristic is defined as an amount of user information transferred in a specified period of time.

7.2.14.2 Quantification

The characteristic is quantified as a rate such as bits per second or bytes per second.

7.2.14.3 Statistical derivations

Statistical derivations for this characteristic include out-bound and in-bound user information throughput characteristics specified at the boundaries between (N) and (N-1) layer and/or over (N)-associations among end-to-end systems.

7.2.15 Application information throughput characteristic

7.2.15.1 Definition

This characteristic is defined as an amount of information transferred between applications in a specific period of time.

7.2.15.2 Quantification

This characteristic is quantified as a rate such as bits per second or bytes per second.

7.2.15.3 Statistical derivations

Statistical derivations for this characteristic include out-bound and in-bound application information throughput characteristics specified at the boundaries over application associations among end-to-end systems.

7.2.16 Subsystem throughput characteristic

7.2.16.1 Definition

This characteristic is defined as an amount of information transferred over all (N)-associations in a specified period of time.

7.2.16.2 Quantification

This characteristic is quantified as a rate such as bits per second or bytes per second.

7.2.16.3 Statistical derivations

Statistical derivations for this characteristic include out-bound and in-bound information throughput characteristics specified at the boundaries among subsystems.

7.2.17 Processing capacity characteristic

7.2.17.1 Definition

This characteristic is defined as an amount of processing performed in a specified period of time.

7.2.17.2 Quantification

This characteristic is quantified as instructions per second.

7.2.17.3 Specializations

There is a very wide range of possible further specializations for the processing capacity characteristic including system throughput, loading, etc.

7.2.18 System throughput characteristic

7.2.18.1 Definition

This characteristic is defined as an amount of processing performed in a specified period of time at the end-system.

7.2.18.2 Quantification

This characteristic is quantified as instructions per second.

7.2.19 Operation loading characteristic

7.2.19.1 Definition

This characteristic is defined as the proportional capacity being used.

7.2.19.2 Quantification

This characteristic is quantified as the ratio of capacity used to capacity available.

7.2.19.3 Specializations

There is a very wide range of possible further specializations for the loading characteristic including association loading, subsystem loading, etc.

7.2.20 Association loading characteristic

7.2.20.1 Definition

This characteristic is defined as the (N)-association capacity being used over the (N)-association.

7.2.20.2 Quantification

This characteristic is quantified as the ratio of capacity used to capacity available.

7.2.21 Subsystem loading characteristic

7.2.21.1 Definition

This characteristic is defined as the (N)-subsystem capacity being used in subsystem.

7.2.21.2 Quantification

This characteristic is quantified as a ratio of capacity used to capacity available.

7.2.22 Accuracy characteristic

7.2.22.1 Definition

This characteristic is defined as the correctness of an event, a set of events or a condition. Accuracy is a QoS characteristic of concern to the user. It refers to the integrity of the user information only (the integrity of headers and similar protocol control information may be the subject of other characteristics).

7.2.22.2 Quantification

This characteristic is quantified as a probability.

7.2.22.3 Specializations

The accuracy characteristic is specialized in many ways, including addressing error, delivery error, residual error, etc.

7.2.23 Addressing error characteristic

7.2.23.1 Definition

This characteristic is defined as the incorrect choice of address(es) used for delivery of data.

7.2.23.2 Quantification

This characteristic is quantified as a probability.

7.2.24 Delivery error characteristic

7.2.24.1 Definition

This characteristic is defined as the delivery of data to an incorrect address.

7.2.24.2 Quantification

This characteristic is quantified as a probability.

7.2.25 Residual error characteristic

7.2.25.1 Definition

This characteristic is defined as the delivery of duplicated and/or excess data.

7.2.25.2 Quantification

This characteristic is quantified as a probability.

7.2.26 Transfer error characteristic

7.2.26.1 Definition

This characteristic is defined as the incorrect transmission of an amount of data.

7.2.26.2 Quantification

This characteristic is quantified as a probability.

7.2.27 Allowable error characteristic

7.2.27.1 Definition

This characteristic is defined as the amount of errors that may be considered to be acceptable.

7.2.27.2 Quantification

This characteristic is quantified as a probability.

7.2.28 Resilience characteristic

7.2.28.1 Definition

This characteristic is defined as the ability to recover from errors.

7.2.28.2 Quantification

This characteristic is quantified as the probability.

7.2.29 Transfer integrity characteristic

7.2.29.1 Definition

This characteristic is defined as the amount of data transferred in a time interval without error.

7.2.29.2 Quantification

This characteristic is quantified as a probability.

7.2.30 Establishment error characteristic

7.2.30.1 Definition

This characteristic is defined as the inability to establish, within a specified time window, a connection or association that was requested.

7.2.30.2 Quantification

This characteristic is quantified as a probability.

7.2.31 Transfer error (loss) characteristic

7.2.31.1 Definition

This characteristic is defined as the inability to transfer all or part of an amount of data that was submitted for delivery.

7.2.31.2 Quantification

This characteristic is quantified as a probability.

7.2.32 Recovery error characteristic

7.2.32.1 Definition

This characteristic is defined as the inability to recover from an error condition.

7.2.32.2 Quantification

This characteristic is quantified as a probability.

7.2.33 Release error characteristic

7.2.33.1 Definition

This characteristic is defined as the inability to release, within a specified time window, a connection or association.

7.2.33.2 Quantification

This characteristic is quantified as a probability.

7.2.34 Safety characteristic

7.2.34.1 Definition

This characteristic is defined as the level of safety of an event, an action, or a resource.

7.2.34.2 Quantification

This characteristic is quantified as a probability of failure of the protection.

7.2.35 Protection characteristic

7.2.35.1 Definition

This characteristic is defined as the security afforded to a resource or to information.

NOTE Protection QoS is the degree to which a service provider attempts to counter security threats using security services. The handling of protection QoS service parameters is a local matter controlled according to the security policy in force. Protection QoS is not negotiated between the service users. For an instance of communication, a service user may indicate its protection QoS requirements to the service provider. A service provider may indicate the protection QoS provided on an instance of communication to the service user. The protection QoS provided by the service provider need not be the same as that requested by the service user. In OSI, any lower layer protocol exchanges between open systems (referred to as “in band” protocol exchanges) to convey information on the security services to be selected are carried in a security association protocol which is independent of an instance of communication. This may be carried implicitly be a security label or explicitly by other means. For further information on the provision of security in the lower layers and the handling of protection QoS see ITU-T Recommendation X.802 | ISO/IEC TR 13594 (Lower Layer Security Guidelines); for the upper layers see ITU-T Recommendation X.803 | ISO/IEC 10745 (Upper Layers Security Model).

7.2.35.2 Quantification

This characteristic is quantified as the probability of failure of the protection.

7.2.36 Access control characteristic

7.2.36.1 Definition

This characteristic is defined as the protection against unauthorized access to a resource.

7.2.36.2 Quantification

This characteristic is quantified as the probability of failure of the protection.

7.2.37 Data protection characteristic

7.2.37.1 Definition

This characteristic is defined as the protection against unauthorized access to data.

7.2.37.2 Quantification

This characteristic is quantified as the probability of failure of the protection.

7.2.38 Confidentiality characteristic

7.2.38.1 Definition

This characteristic is defined as the protection against unauthorized viewing of data.

7.2.38.2 Quantification

This characteristic is quantified as the probability of failure of the protection.

7.2.39 Authenticity characteristic

7.2.39.1 Definition

This characteristic is defined as the authenticity afforded to a resource and to information.

7.2.39.2 Quantification

This characteristic is quantified as the probability of failure of the protection according to ranks of authenticity.

7.2.40 Availability characteristic

7.2.40.1 Definition

This characteristic is defined as the percentage of time that satisfactory service is available.

Availability is generally used as a requirement on the "engineering quality" of the QoS requirement. For a simple system which has no fault tolerance (or no redundancy), availability is simple function of reliability and maintainability, i.e

$$A = \frac{MTBF}{(MTBF + MTTR)} \times 100\ \%$$

where

A	is the availability percentage
MTBF	is the Mean Time Between Failures
MTTR	is the Mean Time To Repair

For some applications, the availability requirement is specified over a finite time interval, e.g. 99 % over a 30-day period. This formulation permits a design where repair or maintenance is not possible during the operational period. The availability is then a function solely of the system reliability.

In more complex systems, the availability requirement can be met even though some elements of the system have failed by providing some degree of redundancy, e.g. a communications service may use an alternate routing in a communications network. In some cases it will be sufficient to specify availability, and not reliability or maintainability. For some systems it may however be necessary to define a limit to the length of time for which a system can be out of service, in which case the "downtime" or maintainability characteristics must be specified as well as the availability. Since availability, reliability and maintainability are related, it should never be necessary to specify all three.