

**Information technology —
Telecommunications and
information exchange between
systems —
Intermediate system to
Intermediate system
intra-domain routeing
information exchange protocol
for use in conjunction with the
protocol for providing the
connectionless-mode Network
Service (ISO 8473)**

ICS 33.040.40; 35.100.30

National foreword

| This British Standard reproduces verbatim ISO/IEC 10589 : 1992 (including
| Technical Corrigenda 1 : 1993, 2 : 1996, 3 : 1996 and Amendment 1 : 1996) and
implements it as the UK national standard.

This British Standard is published under the direction of the Information Systems
Technology Standards Policy Committee whose Technical Committee IST/6 has the
responsibility to:

- aid enquirers to understand the text;
- present to the responsible international committee any enquiries on
interpretation, or proposals for change, and keep the UK interests informed;
- monitor related international and European developments and promulgate
them in the UK.

| NOTE. International and European Standards, as well as overseas standards, are available from
Customer Services, BSI, 389 Chiswick High Road, London W4 4AL.

**Compliance with a British Standard does not of itself confer immunity
from legal obligations.**

Summary of pages

The following table identifies the current issue of each page. Issue 1 indicates that a page has been introduced for the first time by amendment. Subsequent issue numbers indicate an updated page. Vertical sidelining on replacement pages indicates the most recent changes (amendment, addition, deletion).

Page	Issue	Page	Issue
Front cover	2	27 to 33	original
Inside front cover	2	34	2
a	1	35	original
b	blank	36	original
ISO/IEC title page	original	37	2
ii	original	38	2
iii	original	39	original
iv	original	40	2
1	2	41 to 46	original
2	2	47 to 50	2
2a	1	51	original
2b	blank	52	2
3	original	52a	1
4	2	52b	blank
5 to 10	original	53 to 57	original
11	2	58	2
11a	1	59 to 67	original
11b	blank	68	2
12	original	69 to 75	original
13	original	76	2
14	original	77 to 82	original
15	original	83	2
16	2	84	original
17	original	85	2
18	2	86 to 99	original
19	2	100	2
19a	1	101	2
19b	blank	102	original
20	original	103	2
21	2	104 to 141	original
22	2	142 to 150	2
23	original	151 to 174	1
24	original	Inside back cover	2
25	original	Back cover	2
26	2		

INTERNATIONAL STANDARD

ISO/IEC
10589

First edition
1992-06-15

Information technology — Telecommunications and information exchange between systems — Intermediate system to Intermediate system intra-domain routing information exchange protocol for use in conjunction with the protocol for providing the connectionless-mode Network Service (ISO 8473)

*Technologies de l'information — Communication de données et échange
d'informations entre systèmes — Protocole intra-domaine de routage
d'un système intermédiaire à un système intermédiaire à utiliser
conjointement avec le protocole fournissant le service de réseau en
mode sans connexion (ISO 8473)*



Reference number
ISO/IEC 10589:1992(E)

AVA Documents
Document Sharing Hub

Contents

1 Scope 1

2 Normative references 1

3 Definitions 2

4 Symbols and abbreviations 3

5 Typographical conventions 4

6 Overview of the protocol 4

7 Subnetwork independent functions 10

8 Subnetwork dependent functions 35

9 Structure and encoding of PDUs 48

10 System environment 66

11 System management 67

12 Conformance 100

Annex A PICS pro forma 105

Annex B Supporting technical material 117

Annex C Implementation guidelines and examples 121

Annex D Congestion control and avoidance 127

Annex E Syntax imported from ISO 10165-5 (SC6 GMI) 129

Annex F Bibliography 141

Index 143

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

International Standard ISO/IEC 10589 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*.

Annexes A and E form an integral part of this International Standard. Annexes B, C, D and F are for information only.

Introduction

This International Standard is one of a set of International Standards produced to facilitate the interconnection of open systems. The set of standards covers the services and protocols required to achieve such interconnection.

The protocol defined in this International Standard is positioned with respect to other related standards by the layers defined in ISO 7498 and by the structure defined in ISO 8648. In particular, it is a protocol of the Network Layer. This protocol permits Intermediate Systems within a routing domain to exchange configuration and routing information to facilitate the operation of the routing and relaying functions of the Network Layer.

The protocol is designed to operate in close conjunction with ISO 9542 and ISO 8473. ISO 9542 is used to establish connectivity and reachability between End Systems and Intermediate systems on individual subnetworks. Data is carried using the protocol specified in ISO 8473. The related algorithms for route calculation and maintenance are also described.

The intra-domain IS-IS routing protocol is intended to support large routing domains consisting of combinations of many types of subnetworks. This includes point-to-point links, multipoint links, X.25 subnetworks, and broadcast subnetworks such as ISO 8802 LANs.

In order to support large routing domains, provision is made for Intra-domain routing to be organised hierarchically. A large domain may be administratively divided into *areas*. Each system resides in exactly one area. Routing within an area is referred to as *Level 1 routing*. Routing between areas is referred to as *Level 2 routing*. Level 2 Intermediate systems keep track of the paths to destination areas. Level 1 Intermediate systems keep track of the routing within their own area. For an NPDU destined to another area, a Level 1 Intermediate system sends the NPDU to the nearest level 2 IS in its own area, regardless of what the destination area is. Then the NPDU travels via level 2 routing to the destination area, where it again travels via level 1 routing to the destination End system.

Information technology — Telecommunications and information exchange between systems — Intermediate system to Intermediate system intra-domain routing information exchange protocol for use in conjunction with the protocol for providing the connectionless-mode Network Service (ISO 8473)

1 Scope

This International Standard specifies a protocol which is used by Network Layer entities operating the protocol specified in ISO 8473 in Intermediate Systems to maintain routing information for the purpose of routing within a single routing domain. The protocol specified in this International Standard relies upon the provision of a connectionless-mode underlying service.¹⁾

This International Standard specifies:

- a) procedures for the transmission of configuration and routing information between network entities residing in Intermediate Systems within a single routing domain;
- b) the encoding of the protocol data units used for the transmission of the configuration and routing information;
- c) procedures for the correct interpretation of protocol control information; and
- d) the functional requirements for implementations claiming conformance to this International Standard.

The procedures are defined in terms of

- e) the interactions between Intermediate system Network entities through the exchange of protocol data units;
- f) the interactions between a Network entity and an underlying service provider through the exchange of subnetwork service primitives; and
- g) the constraints on route determination which must be observed by each Intermediate system when each has a routing information base which is consistent with the others.

Annexes G, H, I and J, which are integral parts of this International Standard provide ICS proformas associated with intra-domain routing protocol management information.

2 Normative references

The following International Standards contain provisions which, through reference in this text, constitute provisions of this International Standard. At the time of publication, the editions indicated were valid. All International Standards are subject to revision, and parties to agreements based on this International Standard are encouraged to investigate the possibility

of applying the most recent editions of the International Standards listed below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO 7498:1984, *Information processing systems — Open Systems Interconnection — Basic Reference Model*.

ISO 7498/Add.1:1987, *Information processing systems — Open Systems Interconnection — Basic Reference Model — Addendum 1: Connectionless-mode Transmission*.

ISO 7498-3:1989, *Information processing systems — Open Systems Interconnection — Basic Reference Model — Part 3: Naming and addressing*.

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

ISO/IEC 8208:1990, *Information technology — Data communications — X.25 packet Layer Protocol for Data Terminal Equipment*.

ISO 8348:1987, *Information processing systems — Data communications — Network service definition*.

ISO 8348/Add.1:1987, *Information processing systems — Data communications — Network Service Definition — Addendum 1: Connectionless-mode transmission*.

ISO 8348/Add.2:1988, *Information processing systems — Data communications — Network Service Definition — Addendum 2: Network layer addressing*.

ISO 8473:1988, *Information processing systems — Data communications — Protocol for providing the connectionless-mode network service*.

ISO/IEC 8473/Add.3:1989, *Information processing systems — Data Communications — Protocol for providing the connectionless-mode network service — Addendum 3: Provision of the underlying service assumed by ISO 8473 over subnetworks which provide the OSI data link service*.

ISO 8648:1990, *Information processing systems — Open Systems Interconnection — Internal organisation of the Network Layer*.

¹⁾ See ISO 8473 and its addendum 3 for the mechanisms necessary to realise this service on subnetworks based on IEC 8208, ISO 8802, and the OSI Data Link Service.

ISO/IEC 8802-1:¹⁾ *Information technology — Telecommunications and information exchange between systems — Local area networks — Part 1: General Introduction.*

ISO 8802-2:1989, *Information processing systems — Local area networks — Part 2: Logical link control.*

ISO/IEC 8802-3:1990, *Information processing systems — Local area networks — Part 3: Carrier sense multiple access with collision detection (CSMA/CD) access method and physical layer specifications.*

ISO/IEC 8802-5:¹⁾ *Information technology — Local area networks — Part 5: Token ring access method and physical layer specifications.*

ISO/IEC 8802-6:¹⁾ *Information technology — Local area networks — Part 6: Distributed Queue Dual Bus (DQDB) access method and physical layer specifications.*

ISO/IEC 8825: 1990, *Information technology — Open systems Interconnection — Specification of Basic Encoding Rules for Abstract Syntax Notation One (ASN.1).*

ISO/IEC 9314:1989, *Information processing systems — Fiber Distributed Data Interface (FDDI).*

ISO 9542:1988, *Information processing systems — Telecommunications and information exchange between systems — End system to Intermediate system Routeing exchange protocol for use in conjunction with the protocol for providing the connectionless -mode network service (ISO 8473).*

ISO/IEC TR 9575:1990, *Information technology — Telecommunications and information exchange between systems — OSI Routeing Framework.*

ISO/IEC TR 9577:1990, *Information technology — Telecommunications and information exchange between systems — Protocol identification in the network layer.*

ISO/IEC 9646-1: 1994, *Information technology — Open Systems Interconnection — Conformance testing methodology and framework — Part 1: General concepts.*

ISO/IEC 9646-2: 1994, *Information technology — Open Systems Interconnection — Conformance testing methodology and framework — Part 2: Abstract Test Suite specification.*

NOTE — ISO/IEC 9646-1:1994 and ISO/IEC 9646-2:1994 supersede ISO/IEC 9646-1:1991 and ISO/IEC 9646-2:1991 respectively. However, when this International Standard was under development, the previous editions were valid and this International Standard is therefore based on these editions, which are listed below.

ISO/IEC 9646-1: 1991, *Information technology — Open Systems Interconnection — Conformance testing methodology and framework — Part 1: General concepts.*

ISO/IEC 9646-2: 1991, *Information technology — Open Systems Interconnection — Conformance testing methodology and framework — Part 2: Abstract test suite specification.*

ISO/IEC 9646-7: 1995, *Information technology — Open systems Interconnection — Conformance testing methodology and framework — Part 7: Implementation Conformance Statements.*

ISO/IEC 10039:1991, *Information technology — Open Systems Interconnection — Local area networks — Medium Access Control (MAC) service definition.*

ISO/IEC 10165-1:¹⁾ *Information technology — Open Systems Interconnection — Structure of Management Information — Part 1: Management Information Model.*

ISO/IEC 10165-4:¹⁾ *Information technology — Open Systems Interconnection — Structure of management information — Part 4: Guidelines for the definition of managed objects.*

ISO/IEC 10165-6: 1994, *Information technology — Open Systems Interconnection — Structure of management information: Requirements and guidelines for implementation conformance statement proformas associated with OSI management.*

ISO/IEC 10733:¹⁾ *Information technology — Telecommunications and information exchange between systems — Elements of management information relating to OSI Network Layer standards.*

3 Definitions

3.1 Reference model definitions

This International Standard makes use of the following terms defined in ISO 7498:

- a) Network Layer
- b) Network Service access point
- c) Network Service access point address
- d) Network entity
- e) Routeing
- f) Network protocol
- g) Network relay
- h) Network protocol data unit

3.2 Network layer architecture definitions

This International Standard makes use of the following terms defined in ISO 8648:

- a) Subnetwork
- b) End system
- c) Intermediate system
- d) Subnetwork service
- e) Subnetwork Access Protocol
- f) Subnetwork Dependent Convergence Protocol
- g) Subnetwork Independent Convergence Protocol

¹⁾ To be published

3.3 Network layer addressing definitions

This International Standard makes use of the following terms defined in ISO 8348/Add.2:

- a) Subnetwork address
- b) Subnetwork point of attachment
- c) Network Entity Title

3.4 Local area network definitions

This International Standard makes use of the following terms defined in ISO 8802:

- a) Multi-destination address
- b) Media access control
- c) Broadcast medium

3.5 Routeing framework definitions

This International Standard makes use of the following terms defined in ISO/IEC TR 9575:

- a) Administrative Domain
- b) Routeing Domain
- c) Hop
- d) Black hole

3.6 Additional definitions

For the purposes of this International Standard, the following definitions apply:

3.6.1 area: A routing subdomain which maintains detailed routing information about its own internal composition, and also maintains routing information which allows it to reach other routing subdomains. It corresponds to the Level 1 subdomain.

3.6.2 neighbour: An adjacent system reachable by traversal of a single subnetwork by a PDU.

3.6.3 adjacency: A portion of the local routing information which pertains to the reachability of a single neighbour ES or IS over a single circuit.

Adjacencies are used as input to the Decision Process for forming paths through the routing domain.

A separate adjacency is created for each neighbour on a circuit, and for each level of routing (i.e. level 1 and level 2) on a broadcast circuit.

3.6.4 circuit: A subset of the local routing information base pertinent to a single local SNPA. The system management view of a circuit is presented in a linkage managed object.

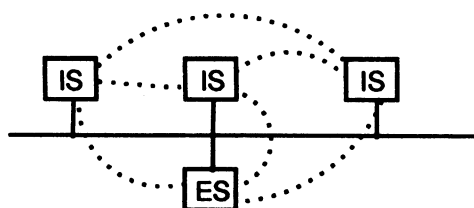
3.6.5 link: The communication path between two neighbours.

A link is "up" when communication is possible between the two SNPAs.

3.6.6 designated IS: The Intermediate system on a LAN which is designated to perform additional duties. In particular it generates Link State PDUs on behalf of the LAN, treating the LAN as a pseudonode.

3.6.7 pseudonode: Where a broadcast subnetwork has n connected Intermediate systems, the broadcast subnetwork itself is considered to be a pseudonode.

The pseudonode has links to each of the n Intermediate and End systems. Each of the ISs has a single link to the pseudonode (rather than $n-1$ links to each of the other Intermediate systems). Link State PDUs are generated on behalf of the pseudonode by the Designated IS. This is depicted below in figure 1.



3.6.8 broadcast subnetwork: A subnetwork which supports an arbitrary number of End systems and Intermediate systems and additionally is capable of transmitting a single SNPDU to a subset of these systems in response to a single SN_UNITDATA request.

3.6.9 general topology subnetwork: A subnetwork which supports an arbitrary number of End systems and Intermediate systems, but does not support a convenient multi-destination connectionless transmission facility, as does a broadcast subnetwork.

3.6.10 routing subdomain: a set of Intermediate systems and End systems located within the same Routing domain.

3.6.11 level 2 subdomain: the set of all Level 2 Intermediate systems in a Routing domain.

3.6.12 jitter: a small random variation introduced into the value of a timer to prevent multiple timer expirations in different systems from becoming synchronised.

4 Symbols and abbreviations

4.1 Data units

PDU	Protocol Data Unit
SNSDU	Subnetwork Service Data Unit
NSDU	Network Service Data Unit
NPDU	Network Protocol Data Unit
SNPDU	Subnetwork Protocol Data Unit

4.2 Protocol data units

ESH PDU	ISO 9542 End System Hello Protocol Data Unit
ISH PDU	ISO 9542 Intermediate System Hello Protocol Data Unit
RD PDU	ISO 9542 Redirect Protocol Data Unit
IIH PDU	Intermediate system to Intermediate system Hello Protocol Data Unit
LSP	Link State Protocol Data Unit
SNP	Sequence Numbers Protocol Data Unit
CSNP	Complete Sequence Numbers Protocol Data Unit
PSNP	Partial Sequence Numbers Protocol Data Unit

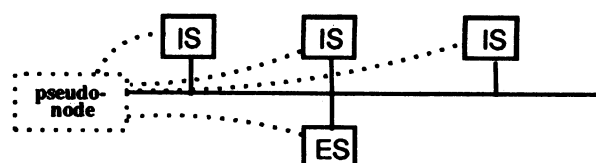


Figure 1 - Use of a pseudonode to collapse a LAN Topology

4.3 Addresses

AFI	Authority and Format Indicator
DSP	Domain Specific Part
IDI	Initial Domain Identifier
IDP	Initial Domain Part
NET	Network Entity Title
NPAI	Network Protocol Addressing Information
NSAP	Network Service Access Point
SNPA	Subnetwork Point of Attachment

4.4 Miscellaneous

DA	Dynamically Assigned
DED	Dynamically Established Data link
DTE	Data Terminal Equipment
ES	End System
IS	Intermediate System
HDLC	High Level Data Link Control
ISDN	Integrated Services Digital Network
FDDI	Fiber Distributed Data Interface
L1	Level 1
L2	Level 2
LAN	Local Area Network
MAC	Media Access Control
MAN	Metropolitan Area Network
MCS	management conformance summary
MICS	management information conformance statement
MOCS	managed object conformance statement
MRCS	managed relationship conformance statement
NLPID	Network Layer Protocol Identifier
PSTN	Public Switched Telephone Network
OSIE	Open Systems Interconnection Environment
PCI	Protocol Control Information
QoS	Quality of Service
SN	Subnetwork
SNACp	Subnetwork Access Protocol

SSN	Send Sequence Numbers
SVC	Switched Virtual Circuit
SNDP	Subnetwork Dependent Convergence Protocol
SNICP	Subnetwork Independent Convergence Protocol
SRM	Send Routing Message

5 Typographical conventions

This International Standard makes use of the following typographical conventions:

- important terms and concepts appear in *italic* type when introduced for the first time;
- protocol constants and management parameters appear in sansSerif type with multiple words run together. The first word is lower case, with the first character of subsequent words capitalised;
- protocol field names appear in San Serif type with each word capitalised; and
- values of constants, parameters, and protocol fields appear enclosed in "double quotes".

6 Overview of the protocol

6.1 System types

For the purposes of this International Standard, systems are classified according to the following types:

End Systems: These systems deliver NPDUs to other systems and receive NPDUs from other systems, but do not relay NPDUs. This International Standard does not specify any additional End system functions beyond those supplied by ISO 8473 and ISO 9542.

Level 1 Intermediate Systems: These systems deliver and receive NPDUs from other systems, and relay NPDUs from other source systems to other destina-

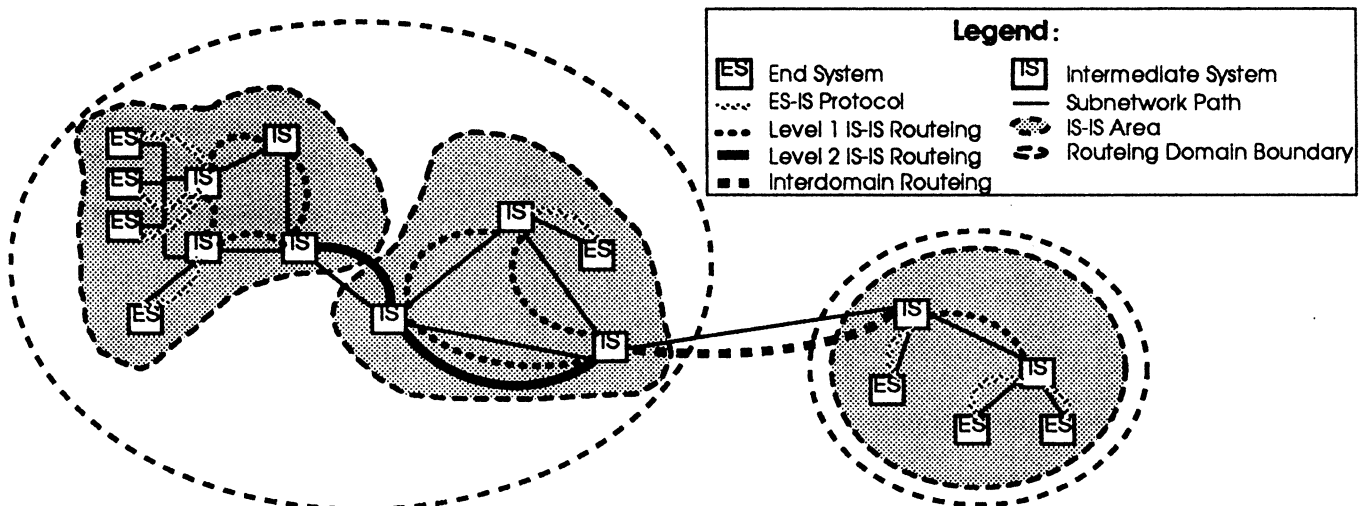


Figure 2 - Topologies and Systems supported by Intradomain Routing

tion systems. They route directly to systems within their own area, and route towards a level 2 Intermediate system when the destination system is in a different area.

Level 2 Intermediate Systems: These systems act as Level 1 Intermediate systems in addition to acting as a system in the subdomain consisting of level 2 ISs. Systems in the level 2 subdomain route towards a destination area, or another routing domain.

These systems and their topological relationship are illustrated in figure 2.

6.2 Subnetwork types

For the purposes of this International Standard, subnetworks are classified according to the following types:

- a) *broadcast subnetworks*: These are multi-access subnetworks that support the capability of addressing a group of attached systems with a single NPDU, for instance ISO 8802-3 LANs.
- b) *general topology subnetworks*: These are modelled as a set of point-to-point links each of which connects exactly two systems.

There are several generic types of general topology subnetworks:

- 1) *multipoint links*: These are links between more than two systems, where one system is a primary system, and the remaining systems are secondary (or slave) systems. The primary is capable of direct communication with any of the secondaries, but the secondaries cannot communicate directly among themselves.
- 2) *permanent point-to-point links*: These are links that stay connected at all times (unless broken, or turned off by system management), for instance leased lines or private links.
- 3) *dynamically established data links (DEDs)*: These are links over connection oriented facilities, for instance X.25, X.21, ISDN, or PSTN networks.

Dynamically established data links can be used in one of two ways:

- i) *static point-to-point (Static)*: The call is established upon system management action and cleared only on system management action (or failure).
- ii) *dynamically assigned (DA)*: The call is established upon receipt of traffic, and brought down on timer expiration when idle. The address to which the call is to be established is determined dynamically from information in the arriving NPDUs. No IS-IS routing PDUs are exchanged between ISs on a DA circuit.

All subnetwork types are treated by the Subnetwork Independent functions as though they were connectionless subnetworks, using the Subnetwork Dependent Convergence functions of ISO 8473 where necessary to provide a connectionless subnetwork service.

The Subnetwork Dependent functions do, however, operate differently on connectionless and connection-oriented subnetworks.

6.3 Topologies

A single organisation may wish to divide its *Administrative Domain* into a number of separate *Routing Domains*. This has certain advantages, as described in ISO/IEC TR 9575. Furthermore, it is desirable for an intra-domain routing protocol to aid in the operation of an inter-domain routing protocol, where such a protocol exists for interconnecting multiple routing domains.

In order to facilitate the construction of such multi-domain topologies, provision is made for the entering of inter-domain routing information. This information is in the form of a set of *Reachable Address Prefixes* which may be entered either by System Management, or provided by an inter-domain routing protocol at the ISs which have links crossing routing domain boundaries. The prefix indicates that any NSAPs whose NSAP address matches the prefix may be reachable via the SNPA with which the prefix is associated. Where this SNPA is connected to a multi-destination subnetwork (e.g., dynamically assigned DED, broadcast), the prefix also has associated with it the required subnetwork addressing information, or an indication that it may be derived from the destination NSAP address (for example, an X.121 DTE address may sometimes be obtained from the IDI of the NSAP address).

The Address Prefixes are handled by the level 2 routing algorithm in the same way as information about a level 1 area within the domain. NPDUs with a destination address matching any of the prefixes present on any Level 2 Intermediate System within the domain can therefore be relayed (using level 2 routing) by that IS and delivered out of the domain. (It is assumed that the routing functions of the other domain will then be able to deliver the NPDUs to its destination.)

Where multiple routing domains are interconnected using this International Standard, the model used is one in which the boundaries between routing domains are on the subnetworks which connect the Intermediate systems. A boundary for a routing domain is constructed by marking the linkage managed object associated with a circuit as being *externalDomain* rather than *internal*.

NOTE 1 This model also permits the construction of routing domains whose scope is not limited by the hierarchical nature of network layer address assignment. For example, it is possible to construct a routing domain, or even a single area, whose area addresses are taken from multiple addressing authorities.

6.4 Addresses

Within a routing domain that conforms to this International Standard, the Network entity titles of Intermediate systems must meet the requirements stated in 7.1.4. It is the routing domain administrative authority's responsibility to ensure that such is the case.

All systems shall be able to generate and forward NPDUs containing NSAP addresses in any of the formats specified by ISO 8348/Add.2. However, the routing domain's administrative authority should ascertain that NSAP addresses of End

systems meet the requirements set forth in 7.1.4 in order to take full advantage the routes derived by this protocol. Within such a domain it is still possible for some End systems to have addresses assigned which do not conform to the rules set forth in 7.1.4 provided that they meet the more general requirements of ISO 8348/Add.2, but these End systems may require additional configuration information to be entered into the Intermediate systems and they may obtain inferior routing performance.

NOTE 2 The procedures whereby the routing domain administrative authority obtains from an appropriate address authority Intermediate system NETs as required by this International Standard, and End system NSAP addresses as recommended by this International Standard are outside its scope.

6.5 Functional organisation

The intra-domain IS-IS routing functions are divided into two groups

- Subnetwork Independent Functions
- Subnetwork Dependent Functions

6.5.1 Subnetwork independent functions

The Subnetwork Independent Functions supply full-duplex NPDU transmission between any pair of neighbour systems. They are independent of the specific subnetwork or data link service operating below them, except for recognising two generic types of subnetworks:

- **General Topology Subnetworks**, which include HDLC point-to-point, HDLC multipoint, and dynamically established data links (such as X.25, X.21, and PSTN links), and
- **Broadcast Subnetworks**, which include ISO 8802 LANs.

NOTE 3 This protocol is intended to operate on any broadcast subnetwork which meets the general requirements listed in 6.7. However, the remainder of this International Standard specifically addresses ISO 8802 LANs. Other LANs, such as FDDI, are believed to be adequately covered by the specification for ISO 8802 LANs. Other broadcast subnetworks, such as ISO 8802-6 MANs, may not be adequately covered at this time.

The following Subnetwork Independent Functions are identified:

- **Routing.** The routing function determines NPDU paths. A path is the sequence of connected systems and links between a source ES and a destination ES.

The combined knowledge of all the Network Layer entities of all the Intermediate systems within a routing domain is used to ascertain the existence of a path, and route the NPDU to its destination. The routing component at an Intermediate system has the following specific functions:

- It extracts and interprets the routing PCI in an NPDU.

- It performs NPDU forwarding based on the destination address.
- It manages the characteristics of the path. If a system or link fails on a path, it finds an alternate route.
- It interfaces with the subnetwork dependent functions to receive reports concerning an SNPA which has become unavailable, a system that has failed, or the subsequent recovery of an SNPA or system.
- It informs the ISO 8473 error reporting function when the forwarding function cannot relay an NPDU, for instance when the destination is unreachable or when the NPDU would have needed to be segmented and the NPDU requested "no segmentation".
- **Congestion control.** Congestion control manages the resources used at each Intermediate system.

6.5.2 Subnetwork dependent functions

The subnetwork dependent functions mask the characteristics of the subnetwork or data link service from the subnetwork independent functions. These include:

- Operation of the Intermediate system functions of ISO 9542 on the particular subnetwork, in order to
 - determine neighbour Network entity title(s) and SNPA address(es);
 - determine the SNPA address(es) of operational Intermediate systems.
- Operation of the requisite Subnetwork Dependent Convergence Function as defined in ISO 8473 and its addendum 3, in order to perform
 - data link initialisation;
 - hop by hop fragmentation over subnetworks with small maximum SNSDU sizes; and
 - call establishment and clearing on dynamically established data links.

6.6 Design goals and non-goals

6.6.1 Goals

This International Standard supports the following design requirements. The correspondence with the goals for OSI routing stated in ISO/IEC TR 9575 are noted.

- **Network Layer Protocol Compatibility:** It is compatible with ISO 8473 and ISO 9542. (See 7.5 of ISO/IEC TR 9575),
- **Simple End systems:** It requires no changes to End systems, nor any functions beyond those supplied by ISO 8473 and ISO 9542. (See 7.2.1 of ISO/IEC TR 9575),
- **Multiple Organisations:** It allows for multiple routing and administrative domains through the provision of static routing information at domain boundaries. (See 7.3 of ISO/IEC TR 9575),

- **Deliverability:** It accepts and delivers NPDUs addressed to reachable destinations and rejects NPDUs addressed to destinations known to be unreachable,
- **Adaptability:** It adapts to topological changes within the routing domain, but not to traffic changes, except potentially as indicated by local queue lengths. It splits traffic load on multiple equivalent paths. (See 7.7 of ISO/IEC TR 9575),
- **Promptness:** The period of adaptation to topological changes in the domain is a reasonable function of the domain diameter (that is, the maximum logical distance between End Systems within the domain) and Data link speeds. (See 7.4 of ISO/IEC TR 9575),
- **Efficiency:** It is both processing and memory efficient. It does not create excessive routing traffic overhead. (See 7.4 of ISO/IEC TR 9575),
- **Robustness:** It recovers from transient errors such as lost or temporarily incorrect routing PDUs. It tolerates imprecise parameter settings. (See 7.7 of ISO/IEC TR 9575),
- **Stability:** It stabilises in finite time to "good routes", provided no continuous topological changes or continuous data base corruptions occur,
- **System Management control:** System Management can control many routing functions via parameter changes, and inspect parameters, counters, and routes. It will not, however, depend on system management action for correct behaviour,
- **Simplicity:** It is sufficiently simple to permit performance tuning and failure isolation,
- **Maintainability:** It provides mechanisms to detect, isolate, and repair most common errors that may affect the routing computation and data bases. (See 7.8 of ISO/IEC TR 9575),
- **Heterogeneity:** It operates over a mixture of network and system types, communication technologies, and topologies. It is capable of running over a wide variety of subnetworks, including, but not limited to: ISO 8802 LANs, ISO/IEC 8208 and X.25 subnetworks, PSTN networks, and the OSI Data Link Service. (See 7.1 of ISO/IEC TR 9575),
- **Extensibility:** It accommodates increased routing functions, leaving earlier functions as a subset,
- **Evolution:** It allows orderly transition from algorithm to algorithm without shutting down an entire domain,
- **Deadlock Prevention:** The congestion control component prevents buffer deadlock,
- **Very Large Domains:** With hierarchical routing, and a very large address space, domains of essentially unlimited size can be supported. (See 7.2 of ISO/IEC TR 9575),
- **Area Partition Repair:** It permits the utilisation of level 2 paths to repair areas which become partitioned due to failing level 1 links or ISs. (See 7.7 of ISO/IEC TR 9575),
- **Determinism:** Routes are a function only of the physical topology, and not of history. In other words, the same topology will always converge to the same set of routes,
- **Protection from Mis-delivery:** The probability of mis-delivering a NPDUs, i.e. delivering it to a Transport entity in the wrong End System, is extremely low,
- **Availability:** For domain topologies with cut set greater than one, no single point of failure will partition the domain. (See 7.7 of ISO/IEC TR 9575),
- **Service Classes:** The service classes of *transit delay, expense*¹⁾, and *residual error probability* of ISO 8473 are supported through the optional inclusion of multiple routing metrics,
- **Authentication:** The protocol is capable of carrying information to be used for the authentication of Intermediate systems in order to increase the security and robustness of a routing domain. The specific mechanism supported in this International Standard however, only supports a weak form of authentication using passwords, and thus is useful only for protection against accidental misconfiguration errors and does not protect against any serious security threat. In the future, the algorithms may be enhanced to provide stronger forms of authentication than can be provided with passwords without needing to change the PDU encoding or the protocol exchange machinery.

6.6.2 Non-goals

The following are not within the design scope of the intra-domain IS-IS routing protocol described in this International Standard:

- **Traffic adaptation:** It does not automatically modify routes based on global traffic load,
- **Source-destination routing:** It does not determine routes by source as well as destination,
- **Guaranteed delivery:** It does not guarantee delivery of all offered NPDUs,
- **Level 2 Subdomain partition repair:** It will not utilise Level 1 paths to repair a level 2 subdomain partition. For full logical connectivity to be available, a connected level 2 subdomain is required,
- **Equal treatment for all ES implementations:** The End system poll function defined in 8.4.5 presumes that End systems have implemented the Suggested ES Configuration Timer option of ISO 9542. An End system which does not implement this option may experience a temporary loss of connectivity following certain types of topology changes on its local subnetwork.

¹⁾ "Expense" is referred to as "cost" in ISO 8473. The latter term is not used here because of possible confusion with the more general usage of the term to indicate path cost according to any routing metric.

6.7 Environmental requirements

For correct operation of the protocol, certain guarantees are required from the local environment and the Data Link Layer.

6.7.1 The required local environment guarantees are:

- a) Resource allocation such that the certain minimum resource guarantees can be met, including
 - 1) memory (for code, data, and buffers)
 - 2) processing;

See 12.2.4 for specific performance levels required for conformance

- b) A quota of buffers sufficient to perform routing functions;
- c) Access to a timer or notification of specific timer expiration; and
- d) A very low probability of corrupting data.

6.7.2 The required subnetwork guarantees for point-to-point links are:

- a) Provision that both source and destination systems complete start-up before PDU exchange can occur;
- b) Detection of remote start-up;
- c) Provision that no old PDUs be received after start-up is complete;
- d) Provision that no PDUs transmitted after a particular startup is complete are delivered out of sequence;
- e) Provision that failure to deliver a specific subnetwork SDU will result in the timely disconnection of the subnetwork connection in both directions and that this failure will be reported to both systems; and
- f) Reporting of other subnetwork failures and degraded subnetwork conditions.
- g) The following events are “very low probability”, which means that performance will be impacted unless they are extremely rare, on the order of less than one event per four years
 - 1) Delivery of NPDUs with undetected data corruption.

6.7.3 The required subnetwork guarantees for broadcast links are:

- a) Multicast capability, i.e., the ability to address a subset of all connected systems with a single PDU;
- b) The following events are “low probability”, which means that they occur sufficiently rarely so as not to impact performance, on the order of once per thousand PDUs

- 1) Routing PDU non-sequentiality,
- 2) Routing PDU loss due to detected corruption; and
- 3) Receiver overrun;
- c) The following events are “very low probability”, which means performance will be impacted unless they are extremely rare, on the order of less than one event per four years
 - 1) Delivery of NPDUs with undetected data corruption; and
 - 2) Non-transitive connectivity, i.e. where system *A* can receive transmissions from systems *B* and *C*, but system *B* cannot receive transmissions from system *C*.

6.7.4 The following services are assumed to be not available from broadcast links:

- a) Reporting of failures and degraded subnetwork conditions that result in NPDU loss, for instance receiver failure. The routing functions are designed to account for these failures.

6.8 Functional organisation of subnetwork independent components

The Subnetwork Independent Functions are broken down into more specific functional components. These are described briefly in this sub-clause and in detail in clause 7. This International Standard uses a functional decomposition adapted from the model of routing presented in subclause 5.1 of ISO/IEC TR 9575. The decomposition is not identical to that in ISO/IEC TR 9575, since that model is more general and not specifically oriented toward a detailed description of intra-domain routing functions such as supplied by this protocol.

The functional decomposition is shown below in figure 3.

The routing processes are:

- *Decision Process*
- *Update Process*

NOTE 4 This comprises both the *Information Collection* and *Information Distribution* components identified in ISO/IEC TR 9575.

- *Forwarding Process*
- *Receive Process*

6.8.1 Decision process

This process calculates routes to each destination in the domain. It is executed separately for level 1 and level 2 routing, and separately within each level for each of the routing metrics supported by the Intermediate system. It uses the *Link State Database*, which consists of information from the latest Link State PDUs from every other Intermediate system in the area, to compute shortest paths from this IS to all other sys-

tems in the area – ⑨ in figure 3. The Link State Data Base is maintained by the Update Process.

Execution of the Decision Process results in the determination of [circuit, neighbour] pairs (known as *adjacencies*), which are stored in the appropriate Forwarding Information base – ⑩ – and used by the Forwarding process as paths along which to forward NPDUs.

Several of the parameters in the routing data base that the Decision Process uses are determined by the implementation. These include:

- maximum number of Intermediate and End systems within the IS's area;
- maximum number of Intermediate and End system neighbours of the IS, etc.,

so that databases can be sized appropriately. Also parameters such as

- routing metrics for each circuit; and
- timers;

can be adjusted for enhanced performance. The complete list of System Management set-able parameters is contained in clause 11.

6.8.2 Update process

This process constructs, receives and propagates Link State PDUs. Each Link State PDU contains information about the identity and routing metric values of the adjacencies of the IS that originated the Link State PDU.

The Update Process receives Link State and Sequence Numbers PDUs from the Receive Process – ④ in figure 3. It places new routing information in the routing information base – ⑥ and propagates routing information to other Intermediate systems – ⑦ and ⑧.

General characteristics of the Update Process are:

- Link State PDUs are generated as a result of topological changes, and also periodically. They may also be generated indirectly as a result of System Management actions (such as changing one of the routing metrics for a circuit).
- Level 1 Link State PDUs are propagated to all Intermediate systems within an area, but are not propagated out of an area.
- Level 2 Link State PDUs are propagated to all Level 2 Intermediate systems in the domain.
- Link State PDUs are not propagated outside of a domain.

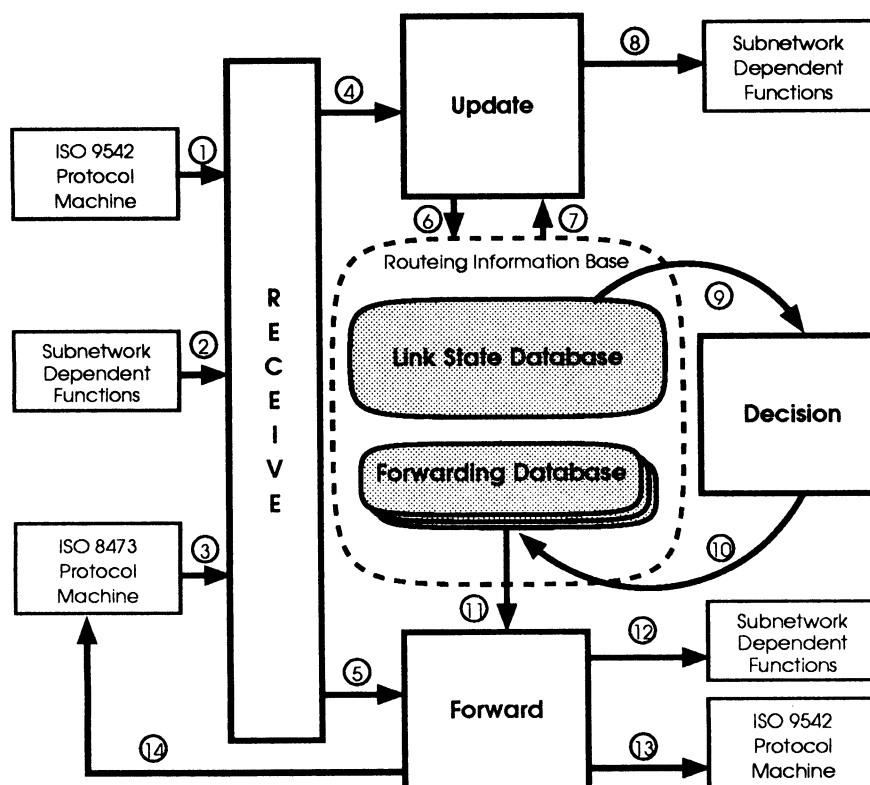


Figure 3 - Decomposition of Subnetwork Independent Functions

- The update process, through a set of System Management parameters, enforces an upper bound on the amount of routing traffic overhead it generates.

6.8.3 Forwarding process

This process supplies and manages the buffers necessary to support NPDU relaying to all destinations.

It receives, via the Receive Process, ISO 8473 PDUs to be forwarded – ⑤ in figure 3.

It performs a lookup in the appropriate¹⁾ Forwarding Database – ① – to determine the possible output adjacencies to use for forwarding to a given destination, chooses one adjacency – ⑫ —, generates error indications to ISO 8473 – ⑭, and signals ISO 9542 to issue Redirect PDUs – ⑬.

6.8.4 Receive process

The Receive Process obtains its inputs from the following sources

- received PDUs with the NLPID of Intra-Domain routing – ② in figure 3,
- routing information derived by the ES-IS protocol from the receipt of ISO 9542 PDUs – ① ; and
- ISO 8473 data PDUs handed to the routing function by the ISO 8473 protocol machine – ③ .

It then performs the appropriate actions, which may involve passing the PDU to some other function (e.g. to the Forwarding Process for forwarding – ⑤).

7 Subnetwork independent functions

This clause describes the algorithms and associated databases used by the routing functions. The managed objects and attributes defined for System Management purposes are described in clause 11.

The following processes and data bases are used internally by the subnetwork independent functions. Following each process or data base title, in parentheses, is the type of systems which must keep the database. The system types are “L2” (level 2 Intermediate system), and “L1” (level 1 Intermediate system). Note that a level 2 Intermediate system is also a level 1 Intermediate system in its home area, so it must keep level 1 databases as well as level 2 databases.

Processes:

- Decision Process (L2, L1)
- Update Process (L2, L1)
- Forwarding Process (L2, L1)
- Receive Process (L2, L1)

Databases:

- Level 1 Link State data base (L2, L1)
- Level 2 Link State data base (L2)
- Adjacency Database (L2, L1)
- Circuit Database (L2, L1)
- Level 1 Shortest Paths Database (L2, L1)
- Level 2 Shortest Paths Database (L2)
- Level 1 Forwarding Databases — one per routing metric (L2, L1)
- Level 2 Forwarding Database — one per routing metric (L2)

7.1 Addresses

The NSAP addresses and NETs of systems are variable length quantities that conform to the requirements of ISO 8348/Add.2. The corresponding NPAI contained in ISO 8473 PDUs and in this protocol's PDUs shall use the preferred binary encoding. Any of the AFIs and their corresponding DSP syntax may be used with this protocol.

7.1.1 Address structure for intradomain IS-IS routing

In order to understand the requirements set under the present clause 7.1, it is necessary to view the encoded NSAPs or NETs as structured according to figure 4, where three fields are distinguished:

- Area Address
- ID
- SEL

¹⁾ The appropriate forwarding database is selected by choosing a routing metric based on fields in the QoS Maintenance option field of ISO 8473.

7.1.2 NPAI — area address field

An area address is a variable length quantity consisting of the entire high-order part of the NPAI, excluding the ID and SEL fields.

7.1.3 NPAI of systems within a routing domain

This clause first defines how the NPAI corresponding to NSAP addresses and Network Entity titles of systems deployed in a routing domain is constructed and second how the NPAI is structured for use by the protocol.

7.1.3.1 Construction of NPAI from network addresses

NPAI is derived from NSAP addresses and NETs according to ISO/IEC 8348/Add.2. The NETs and NSAP addresses are obtained from the appropriate addressing authorities.

For these addresses to be correctly interpreted by the protocol in this International Standard it is mandatory that the routing domain authority ensure that whenever the Network address includes a DSP whose syntax is decimal digits, then:

- if the IDI is in the ISO DCC or E.164 format, the decimal syntax DSP must be an odd number of digits;
- for the other IDI formats the decimal syntax DSP must be an even number of decimal digits.

7.1.3.2 Structure of the NPAI

The structure of the ID and SEL fields of the NPAI are interpreted in the following way by the protocol defined in this International Standard:

ID System identifier — a variable length field from 1 to 8 octets (inclusive). Each routing domain employing this protocol shall select a single size for the ID field and all Intermediate systems in the routing domain shall use this length for the system IDs of all systems in the routing domain.

The set of ID lengths supported by an implementation is an implementation choice, provided that at least one value in the permitted range can be accepted. The routing domain administrator must ensure that all ISs included in a routing domain are able to use the ID length chosen for that domain.

SEL NSAP Selector — a 1-octet field which acts as a selector for the entity which is to receive the PDU (this may be a Transport entity or the Intermediate system Network entity itself). It is the least significant (last) octet of the NPAI.

NOTE – The SEL field is always the last octet of the NPAI, since the rules enforced in 7.1.3.1 guarantee that there will be no pad at the end of the NPAI.

7.1.4 Administration and deployment of systems in a routing domain

It is the responsibility of the routing domain administrative authority to enforce the requirements stated below in this clause and the rules stated above in 7.1.3.1. These requirements place specific constraints on the NSAP addresses and NETs of systems deployed in a routing domain, when these systems operate the protocol defined in this International Standard. The protocol defined in this International Standard assumes that these requirements are met, but has no means to verify compliance with them.

NOTE 5 To interpret correctly the requirements given below, it is necessary to refer to the structure of the NPAI presented in 7.1.1 and 7.1.3.2, and to the concept of manual area addresses defined in 7.1.5.

For correct operation of the routing protocol defined in this International Standard, the following requirements must be met in a routing domain:

- a) For all systems in the routing domain:
 - 1) By definition, all systems in a routing domain that have a given value of area address belong to the same area.

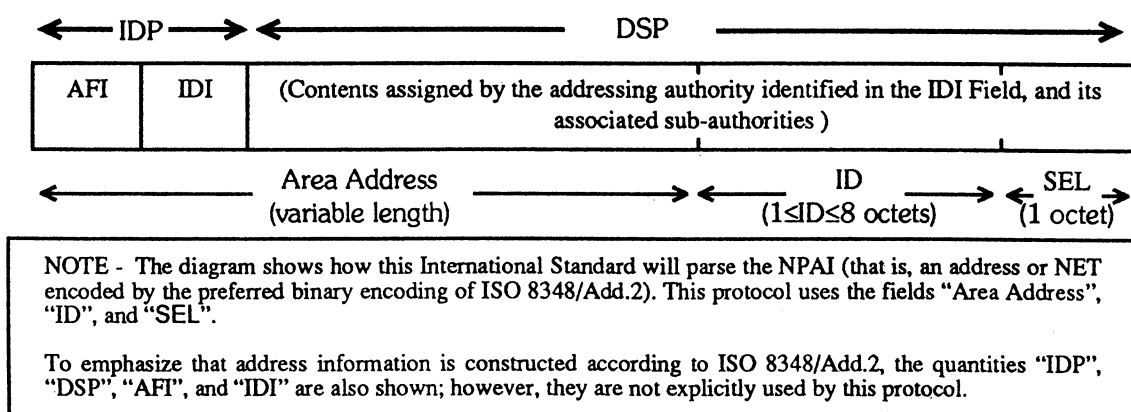


Figure 4 - Address Structure for Intra-domain IS-IS Routing

NOTE 6 A consequence of this requirement is that a reachable address prefix may not match any area address of an area in the routing domain. However, an IS is not required to perform any dynamic check to detect if this property is violated due to system management misconfiguration.

- 2) Each system in an area must have an unambiguous ID; that is, no two systems (IS or ES) in an area may use the same ID value.
 - 3) All systems belonging to a given routing domain must have NETs or NSAP addresses whose ID fields are of equal length.
- b) Additional requirements for Intermediate system addresses:
- 1) Each Level 2 Intermediate system within a routing domain must have an unambiguous value for its ID field; that is, no two level 2 ISs in a routing domain can have the same value in their ID fields.
- c) Additional requirements for area administration:
- 1) No two End systems in an area may have addresses that match in all but the SEL field.
- d) Requirements placed on End systems to be neighbours of a level 1 IS:
- An End system may be a neighbour of a level 1 IS if and only if:

- 1) its area address matches the level 1 IS's area address as contained in its own NET, which is always an entry in the adjacent IS's `manualAreaAddresses` parameter, or
- 2) its area address matches one of the other entries in the adjacent IS's `manualAreaAddresses` parameter.

NOTE 7 The definitions of several entries in the adjacent IS's `manualAreaAddresses` parameter allow for systems with differing area addresses to be merged under one level 1 IS for the purpose of fabricating a single area. This concept is described in detail under 7.1.5 below.

e) Additional requirements for Interdomain routing:

When Interdomain routing between two routing domains is assisted through the use of reachable address prefixes, then

- 1) no address of any area in one of the routing domains should match the address of an area in the other routing domain.

7.1.5 Manual area addresses

Within a routing domain, it is often convenient to associate more than one area address with an area. There are a number of reasons why assigning more than one area address may be useful, including the following.

- a) There may be more than one addressing authority involved in the assignment of addresses in the routing domain, yet it is not efficient to require a separate area for each addressing domain.
- b) At times it may be necessary to reconfigure a routing domain by dividing an area into two or more areas, or combining a number of areas into a single area. These reconfigurations could not be done during normal routing domain operation if only a single area address per area were permitted.

Therefore, this International Standard permits an area to have a number of synonymous area addresses.

The use of several synonymous area addresses by an IS is accommodated through the use of the management parameter `manualAreaAddresses`. This parameter is set locally for each IS by system management; it contains a list of all synonymous area addresses associated with the IS. All of the IS's `manualAreaAddresses`, when combined with the IS's `systemID`, are valid network entity titles for the IS.

Each level 1 IS distributes its `manualAreaAddresses` in its Level 1 LSP's `Area Addresses` field, thus allowing level 2 ISs to create a composite list of all area addresses in use within a given area. Level 2 ISs in turn advertise the composite list throughout the level 2 subdomain by including it in their Level 2 LSP's `Area Addresses` field, thus distributing information on all the area addresses associated with the entire routing domain. The procedures for establishing an adjacency between two level 1 ISs require that there be at least one area address in common between their two `manualAreaAddresses` lists, and the procedures for establishing an adjacency between a level 1 IS and an End system require that the End system's area ad-

dress match an entry in the IS's `manualAreaAddresses` list. Therefore, it is the responsibility of System Management to ensure that each area address associated with an IS is included. In particular, system management must ensure that the area addresses of all ESs and Level 1 ISs adjacent to a given level 1 IS are included in that IS's `manualAreaAddresses` list.

The union of all area addresses of the ISs in an area may exceed the configured capacity of one or more ISs in the area. In order that all ISs agree on the area addresses of an area, each IS has a parameter `maximumAreaAddresses`, established by System Management. The value of `maximumAreaAddresses` is communicated among the ISs in the protocol's PDUs and is checked to ensure that all ISs in an area have the same value of this parameter. All ISs shall support a value of at least 3 for `maximumAreaAddresses`, although the value may be set lower by System Management if desired. Failure to set `maximumAreaAddresses` consistently among the ISs in an area may cause adjacencies to fail to initialise and/or an area to become partitioned.

If the area address field for the destination address of an ISO 8473 PDU — or for the next entry in its source routing field, when present — is not listed in the parameter `areaAddresses` of a level 1 IS receiving the PDU, then the destination system does not reside in the IS's area. Such PDUs will be routed by level 2 routing.

7.1.6 Encoding of addressing information

This International Standard makes use of four types of address information: NETs, NSAP addresses, area addresses, and address prefixes. The encoding rules for each of them are given below.

- a) NETs shall be encoded according to the preferred binary encoding specified in ISO 8348/Add.2.
- b) NSAP addresses shall be encoded according to the preferred binary encoding specified in ISO 8348/Add.2.
- c) The encoded form of an area address shall be obtained by dropping the last $IDLength + 1$ octets of the preferred binary encoding of the corresponding NSAP, where $IDLength$ is equal to the length of the ID field used by the routing domain.
- d) The encoded form of an address prefix shall be obtained by encoding the prefix (expressed in its abstract syntax), according to the preferred binary encoding, unless the end of the prefix falls within the IDP. In this case, each decimal digit in the prefix shall be encoded as the corresponding semi-octet in the range 0000-1001 and no padding characters shall be inserted.

7.1.7 Matching an NSAP address with an area address or an address prefix

A destination NSAP address can be matched against either an area address or an address prefix. An area address or an address prefix which extends into the DSP, shall be compared directly against the encoded NSAP address, including any padding characters that may be present; an address prefix which does not extend into the DSP shall be compared against NSAP', which is obtained from the encoded NSAP address by

removing all padding characters that were inserted by the binary encoding.

The existence of a match shall be determined as follows:

- a) If the encoded NSAP (or NSAP^f) contains fewer semi-octets than the encoded area address (or address prefix), then there is no match.
- b) If the NSAP (or NSAP^f) contains at least as many octets as the area address (or address prefix), and all octets of the encoded area address (or address prefix) are identical to the corresponding leading octets of the encoded NSAP address (or NSAP^f), there is a match. Otherwise, there is no match.

NOTE 8 Any implementation of a matching process that satisfies the requirements listed above may be used. The key point is that matching process must be aware of whether or not the encoded area address or address prefix extends into the DSP, and must then either include or exclude padding characters from the encoded NSAP, as defined above.

7.1.8 Comparison of addresses

Unless otherwise stated, comparison of addresses shall be performed on the addresses encoded as sequences of octets in the form specified in 9.2 or 9.3 as appropriate.

If the encoded addresses are of different lengths and the shorter encoded address is a prefix of the longer, the shorter address is considered to be less than the longer.

Otherwise, the result of the comparison is that obtained by padding the shorter encoded address (if any) with trailing zero octets to the length of the longer, and comparing the two resulting octet sequences as unsigned integers, with the first octet of each sequence considered the most significant.

The addresses to which this procedure applies are NSAP addresses, Network Entity Titles, and SNPA addresses.

7.2 Decision process

This process uses the database of Link State information to calculate the forwarding database(s), from which the forwarding process can know the proper next hop for each NPDU. The Level 1 Link State Database is used for calculating the Level 1 Forwarding Database(s), and the Level 2 Link State Database is used for calculating the Level 2 Forwarding Database(s).

7.2.1 Input and output

INPUT

- Link State Database – This database is a set of information from the latest Link State PDUs from all known Intermediate systems (within this area, for Level 1, or within the level 2 subdomain, for Level 2). This database is received from the Update Process.

- Notification of an Event – This is a signal from the Update Process that a change to a link has occurred somewhere in the domain.

OUTPUT

- Level 1 Forwarding Databases — one per routing metric
- (Level 2 Intermediate systems only) Level 2 Forwarding Databases — one per routing metric
- (Level 2 Intermediate systems only) The Level 1 Decision Process informs the Level 2 Update Process of the ID of the Level 2 Intermediate system within the area with lowest ID reachable with real level 1 links (as opposed to a virtual link consisting of a path through the level 2 subdomain)
- (Level 2 Intermediate systems only) If this Intermediate system is the “Partition Designated Level 2 Intermediate system” in this partition, the Level 2 Decision Process informs the Level 1 Update Process of the values of the default routing metric to and ID of the “partition designated level 2 Intermediate system” in each other partition of this area.

7.2.2 Routing metrics

There are four routing metrics defined, corresponding to the four possible orthogonal qualities of service defined by the QoS Maintenance field of ISO 8473. Each circuit emanating from an Intermediate system shall be assigned a value for one or more of these metrics by System management. The four metrics are as follows:

- a) *Default metric*: This is a metric understood by every Intermediate system in the domain. Each circuit shall have a positive integral value assigned for this metric. The value may be associated with any objective function of the circuit, but by convention is intended to measure the *capacity* of the circuit for handling traffic, for example, its throughput in bits-per-second. Higher values indicate a lower capacity.
- b) *Delay metric*: This metric measures the *transit delay* of the associated circuit. It is an optional metric, which if assigned to a circuit shall have a positive integral value. Higher values indicate a longer transit delay.
- c) *Expense metric*: This metric measures the *monetary cost* of utilising the associated circuit. It is an optional metric, which if assigned to a circuit shall have a positive integral value¹⁾. Higher values indicate a larger monetary expense.
- d) *Error metric*: This metric measures the *residual error probability* of the associated circuit. It is an optional metric, which if assigned to a circuit shall have a non-zero value. Higher values indicate a larger probability of undetected errors on the circuit.

¹⁾ The path computation algorithm utilised in this International Standard requires that all circuits be assigned a positive value for a metric. Therefore, it is not possible to represent a “free” circuit by a zero value of the expense metric. By convention, the value 1 is used to indicate a “free” circuit.

NOTE 9 The decision process combines metric values by simple addition. It is important, therefore, that the values of the metrics be chosen accordingly.

Every Intermediate system shall be capable of calculating routes based on the default metric. Support of any or all of the other metrics is optional. If an Intermediate system supports the calculation of routes based on a metric, its update process may report the metric value in the LSPs for the associated circuit; otherwise, the IS shall not report the metric.

When calculating paths for one of the optional routing metrics, the decision process only utilises LSPs with a value reported for the corresponding metric. If none of an IS's circuits has a value associated with one of the optional metrics, then IS shall not calculate routes based on that metric.

NOTE 10 A consequence of the above is that a system reachable via the default metric may not be reachable by another metric.

See 7.4.2 for a description of how the forwarding process selects one of these metrics based on the contents of the ISO 8473 QoS Maintenance option.

Each of the four metrics described above may be of two types: an *Internal metric* or an *External metric*. Internal metrics are used to describe links/routes to destinations internal to the routing domain. External metrics are used to describe links/routes to destinations outside of the routing domain. These two types of metrics are not directly comparable, except the internal routes are always preferred over external routes. In other words an internal route will always be selected even if an external route with lower total cost exists.

7.2.3 Broadcast subnetworks

Instead of treating a broadcast subnetwork as a fully connected topology, the broadcast subnetwork is treated as a pseudonode, with links to each attached system. Attached systems shall only report their link to the pseudonode. The designated Intermediate system, on behalf of the pseudonode, shall construct Link State PDUs reporting the links to all the systems on the broadcast subnetwork with a zero value for each supported routing metric¹.

The pseudonode shall be identified by the `sourceID` of the Designated Intermediate system, followed by a non-zero `pseudonodeID` assigned by the Designated Intermediate system. The `pseudonodeID` is locally unique to the Designated Intermediate system.

Designated Intermediate systems are determined separately for level 1 and level 2. They are known as the *LAN Level 1 Designated IS* and the *LAN Level 2 Designated IS* respectively. See 8.4.5.

An Intermediate system may resign as Designated Intermediate System on a broadcast circuit either because it (or its SNPA on the broadcast subnetwork) is being shut down or because some other Intermediate system of higher priority has taken over that function. When an Intermediate system resigns as Designated Intermediate System, it shall initiate a network wide purge of its pseudonode Link State PDU(s) by setting

their Remaining Lifetime to zero and performing the actions described in 7.3.16.4. A LAN Level 1 Designated Intermediate System purges Level 1 Link State PDUs and a LAN Level 2 Designated Intermediate System purges Level 2 Link State PDUs. An Intermediate system which has resigned as both Level 1 and Level 2 Designated Intermediate System shall purge both sets of LSPs.

When an Intermediate system declares itself as designated Intermediate system and it is in possession of a Link State PDU of the same level issued by the previous Designated Intermediate System for that circuit (if any), it shall initiate a network wide purge of that (or those) Link State PDU(s) as above.

7.2.4 Links

Two Intermediate systems are not considered neighbours unless each reports the other as directly reachable over one of their SNPAs. On a Connection-oriented subnetwork (either point-to-point or general topology), the two Intermediate systems in question shall ascertain their neighbour relationship when a connection is established and hello PDUs exchanged. A malfunctioning IS might, however, report another IS to be a neighbour when in fact it is not. To detect this class of failure the decision process checks that each link reported as "up" in a LSP is so reported by both Intermediate systems. If an Intermediate system considers a link "down" it shall **not** mention the link in its Link State PDUs.

On broadcast subnetworks, this class of failure shall be detected by the designated IS, which has the responsibility to ascertain the set of Intermediate systems that can all communicate on the subnetwork. The designated IS shall include these Intermediate systems (and no others) in the Link State PDU it generates for the pseudonode representing the broadcast subnetwork.

7.2.5 Multiple LSPs for the same system

The Update process is capable of dividing a single logical LSP into a number of separate PDUs for the purpose of conserving link bandwidth and processing (see 7.3.4). The Decision Process, on the other hand, shall regard the LSP with LSP Number zero in a special way. If the LSP with LSP Number zero and remaining lifetime > 0 is not present for a particular system then the Decision Process shall not process any LSPs with non-zero LSP Number which may be stored for that system.

The following information shall be taken only from the LSP with LSP Number zero. Any values which may be present in other LSPs for that system shall be disregarded by the Decision Process.

- a) The setting of the LSP Database Overload bit.
- b) The value of the IS Type field.
- c) The Area Addresses option field.

7.2.6 Routing algorithm overview

7.2.6.1 The routing algorithm used by the Decision Process is a *shortest path first (SPF)* algorithm. Instances of the algo-

¹ They are set to zero metric values since they have already been assigned metrics by the link to the pseudonode. Assigning a non-zero value in the pseudonode LSP would have the effect of doubling the actual value.

algorithm are run independently and concurrently by all Intermediate systems in a routing domain. Intra-Domain routing of a PDU occurs on a hop-by-hop basis: that is, the algorithm determines only the next hop, not the complete path, that a data PDU will take to reach its destination. To guarantee correct and consistent route computation by every Intermediate system in a routing domain, this International Standard depends on the following properties:

- a) All Intermediate systems in the routing domain converge to using identical topology information; and
- b) Each Intermediate system in the routing domain generates the same set of routes from the same input topology and set of metrics.

The first property is necessary in order to prevent inconsistent, potentially looping paths. The second property is necessary to meet the goal of determinism stated in 6.6.

7.2.6.2 A system executes the SPF algorithm to find a set of legal paths to a destination system in the routing domain. The set may consist of

- a) a single path of minimum metric sum: these are termed *minimum cost paths*;
- b) a set of paths of equal minimum metric sum: these are termed *equal minimum cost paths*; or
- c) a set of paths which will get a PDU closer to its destination than the local system: these are called *downstream paths*.

Paths which do not meet the above conditions are illegal and shall not be used. Paths whose metric sum exceeds the value of the architectural constant **MaxPathMetric** (see table 2) are also illegal and shall not be used.

7.2.6.3 The Decision Process, in determining its paths, also ascertains the identity of the adjacency which lies on the first hop to the destination on each path. These adjacencies are used to form the Forwarding Database, which the forwarding process uses for relaying PDUs.

7.2.6.4 Separate route calculations are made for each pairing of a level in the routing hierarchy (i.e. L1 and L2) with a supported routing metric. Since there are four routing metrics and two levels some systems may execute multiple instances of the SPF algorithm. For example,

- if an IS is a L2 Intermediate system which supports all four metrics and computes minimum cost paths for all metrics, it would execute the SPF calculation eight times.
- if an IS is a L1 Intermediate system which supports all four metrics, and additionally computes downstream paths, it would execute the algorithm $4 \times (\text{number of neighbours} + 1)$ times.

Any implementation of an SPF algorithm meeting both the static and dynamic conformance requirements of clause 12 of this International Standard may be used. Recommended implementations are described in detail in annex C.

7.2.7 Removal of excess paths

When there are more than **maximumPathSplits** legal paths to a destination, this set shall be pruned until only **maximumPathSplits** remain. The Intermediate system shall discriminate based upon:

NOTE 11 The precise precedence among the paths is specified in order to meet the goal of determinism defined in 6.6.

- **adjacency type:** Paths associated with End system or level 2 reachable address prefix adjacencies are retained in preference to other adjacencies
- **metric sum:** Paths having a lesser metric sum are retained in preference to paths having a greater metric sum. By metric sum is understood the sum of the metrics along the path to the destination.
- **neighbour ID:** where two or more paths are associated with adjacencies of the same type, an adjacency with a lower neighbour ID is retained in preference to an adjacency with a higher neighbour id.
- **circuit ID:** where two or more paths are associated with adjacencies of the same type, and same neighbour ID, an adjacency with a lower *circuit ID* is retained in preference to an adjacency with a higher *circuit ID*, where *circuit ID* is the value of
 - **ptPtCircuitID** for non-broadcast circuits,
 - **l1CircuitID** for broadcast circuits when running the Level 1 Decision Process, and
 - **l2CircuitID** for broadcast circuits when running the Level 2 Decision Process.
- **IANAAddress:** where two or more adjacencies are of the same type, same neighbour ID, and same circuit ID (e.g. where a neighbouring system has multiple LAN adapters on the same LAN) an adjacency with a lower IANAAddress is retained in preference to an adjacency with a higher IANAAddress.

7.2.8 Robustness checks

7.2.8.1 Computing routes through overloaded Intermediate systems

The Decision Process shall not utilise a link to an Intermediate system neighbour from an IS whose LSPs have the LSP Database Overload indication set. Such paths may introduce loops since the overloaded IS does not have a complete routing information base. The Decision Process shall, however utilise the link to reach End system neighbours since these paths are guaranteed to be non-looping.

7.2.8.2 Two-way connectivity check

The Decision Process shall not utilise a link between two Intermediate Systems unless both ISs report the link.

NOTE 12 The check is not applicable to links to an End System.

Reporting the link indicates that it has a defined value for at least the default routing metric. It is permissible for two endpoints to report different defined values of the same metric for the same link. In this case, routes may be asymmetric.

7.2.9 Construction of a forwarding database

The information that is needed in the forwarding database for routing metric k is the set of adjacencies for each system ..

7.2.9.1 Identification of nearest level 2 IS by a level 1 IS

Level 1 Intermediate systems need one additional piece of information per routing metric: the next hop to the nearest level 2 Intermediate system according to that routing metric. A level 1 IS shall ascertain the set, R , of "attached" level 2 Intermediate system(s) for metric k such that the total cost to R for metric k is minimal.

If there are more adjacencies in this set than maximumPath-Splits, then the IS shall remove excess adjacencies as described in 7.2.7.

7.2.9.2 Setting the attached flag in level 2 intermediate systems

When executing the level 2 decision process for each supported metric, level 2 IS shall ascertain whether or not it can reach any destinations outside its area using that metric. The IS considers itself attached if either:

- it can reach at least one other area using the corresponding routing metric, or
- it has at least one enabled reachable address prefix with the corresponding metric defined.

Otherwise the IS considers itself not attached.

If the IS discovers that it is not attached and attachedFlag was previously True, it shall:

- set AttachedFlag for metric k to False;
- regenerate its Level 1 LSP with LSP number zero; and
- compute the "nearest level 2 Intermediate system" for metric k for insertion in the appropriate forwarding database, according to the algorithm described in 7.2.9.1 for level 1 Intermediate systems.

NOTE 13 attachedFlag for each metric k is examined by the Update Process, so that it will report the value in the ATT field of its Link State PDUs.

If the IS discovers that it is attached and attachedFlag was previously False, it shall:

- set attachedFlag for metric k to "True";
- regenerate its Level 1 LSP with LSP number zero; and
- set the level 1 forwarding database entry for metric k which corresponds to "nearest level 2 Intermediate system" to "Self".

7.2.10 Information for repairing partitioned areas

An area may become partitioned as a result of failure of one or more links in the area. However, if each of the partitions has a connection to the level 2 subdomain, it is possible to repair the partition via the level 2 subdomain, provided that the level 2 subdomain itself is not partitioned. This is illustrated in figure 5.

All the systems A-I, R and P are in the same area n . When the link between D and E is broken, the area becomes partitioned. Within each of the partitions the *Partition Designated Level 2 Intermediate system* is selected from among the level 2 Intermediate systems in that partition. In the case of partition 1 this is P, and in the case of partition 2 this is R. The level 1 repair path is then established between these two level 2 Intermediate systems. Note that the repaired link is now between P and R, not between D and E.

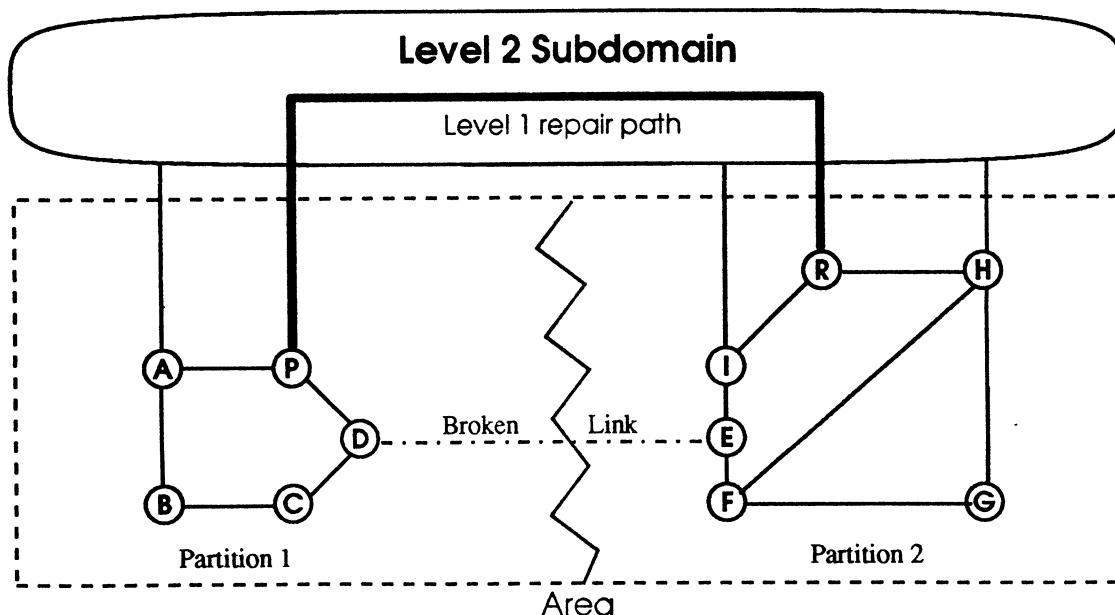


Figure 5 - Repair of partitioned level 1 area

The Partition Designated Level 2 Intermediate Systems repair the partition by forwarding NPDUs destined for other partitions of the area through the level 2 subdomain. They do this by acting in their capacity as Level 1 Intermediate Systems and advertising in their Level 1 LSPs adjacencies to each Partition Designated Level 2 Intermediate System in the area. This adjacency is known as a "Virtual Adjacency" or "Virtual Link". Thus other Level 1 Intermediate Systems in a partition calculate paths to the other partitions through the Partition Designated Level 2 Intermediate System. A Partition Designated Level 2 Intermediate System forwards the Level 1 NPDUs through the level 2 subdomain by encapsulating them in ISO 8473 Data NPDUs with its Virtual Network Entity Title as the source NSAP and the "adjacent" Partition Designated Level 2 Intermediate System's Virtual Network Entity Title as the destination NSAP. The following sub-clauses describe this in more detail.

7.2.10.1 Partition detection and virtual level 1 link creation

Partitions of a Level 1 area are detected by the Level 2 Intermediate System(s) operating within the area. Participation in the partition repair process by a Level 2 Intermediate system is predicated on the fact that all L2 ISs also function as L1 ISs within their own area. A partition of a given area exists whenever two or more Level 2 ISs located in that area are reported in the L2 LSPs as being a "Partition Designated Level 2 IS". Conversely, when only one Level 2 IS in an area is reported as being the "Partition Designated Level 2 IS", then that area is not partitioned. Partition repair is accomplished by the Partition Designated Level 2 IS. The election of the Partition Designated Level 2 IS as described in the next subsection must be done before the detection and repair process can begin.

In order to repair a partition of a Level 1 area, the Partition designated Level 2 IS creates a "Virtual Network Entity" to represent the partition. The Network Entity Title for this virtual network entity shall be constructed from the first listed area address from its Level 2 Link State PDU, and the ID of the Partition Designated Level 2 IS, with a selector value of `IntradomainRoutingSelector`. The IS shall also construct a virtual link (represented by a new Virtual Adjacency managed object) to each Partition Designated Level 2 IS in the area, with the NET of the partition recorded in the Identifier attribute. The virtual links are the repair paths for the partition. They are reported by the Partition Designated Level 2 IS into the entire Level 1 area by adding the ID of each "adjacent" Partition Designated Level 2 IS to the Intermediate System Neighbours field of its Level 1 Link State PDU. The Virtual Flag shall be set "True" for these Intermediate System neighbours. The metric value for this virtual link shall be the default metric value $d(N)$ obtained from this system's Level 2 PATHS database, where N is the "adjacent" Partition Designated Level 2 IS via the Level 2 subdomain. If the computed metric value exceeds the value of `MaxLinkMetric`, the IS shall report the value of `MaxLinkMetric` for the virtual link.

An Intermediate System which operates as the Partition Designated Level 2 Intermediate System shall perform the following steps after completing the Level 2 shortest path computation in order to detect partitions in the Level 1 area and create repair paths:

- a) Examine Level 2 Link State PDUs of all Level 2 Intermediate systems. Search Area Addresses for any ad-

dress that matches any of the addresses in `partitionAreaAddresses`. If a match is found, and the Partition Designated Level 2 Intermediate system's ID does not equal this system's ID, then inform the level 1 update process at this system of the identity of the Partition Designated Level 2 Intermediate system, together with the path cost for the default routing metric to that Intermediate system.

- b) Continue examining Level 2 LSPs until all Partition Designated Level 2 Intermediate systems in other partitions of this area are found, and inform the Level 1 Update Process of all of the other Partition Designated Level 2 Intermediate systems in other partitions of this area, so that
 - 1) Level 1 Link State PDUs can be propagated to all other Partition designated level 2 Intermediate systems for this area (via the level 2 subdomain).
 - 2) All the Partition Designated Level 2 Intermediate systems for other partitions of this area can be reported as adjacencies in this system's Level 1 Link State PDUs.

If a partition has healed, the IS shall destroy the associated virtual network entity and virtual link by deleting the Virtual Adjacency. The Partition Designated Level 2 IS detects a healed partition when another Partition Designated Level 2 IS listed as a virtual link in its Level 1 Link State PDU was not found after running the partition detection and virtual link creation algorithm described above.

If such a Virtual Adjacency is created or destroyed, the IS shall generate a `partitionVirtualLinkChange` event.

7.2.10.2 Election of partition designated level 2 intermediate system

The Partition Designated Level 2 IS is a Level 2 IS which:

- reports itself as "attached" by the default metric in its LSPs;
- reports itself as implementing the partition repair option;
- operates as a Level 1 IS in the area;
- is reachable from the IS performing the calculation via Level 1 routing without traversing any virtual links; and
- has the lowest ID

The election of the Partition Designated Level 2 IS is performed by running the decision process algorithm after the Level 1 decision process has finished, and before the Level 2 decision process to determine Level 2 paths is executed.

In order to guarantee that the correct Partition Designated Level 2 IS is elected, the decision process is run using only the Level 1 LSPs for the area, and by examining only the Intermediate System Neighbours whose Virtual Flag is FALSE. The results of this decision process is a set of all the Level 1 Intermediate Systems in the area that can be reached via Level 1, non-virtual link routing. From this set, the Partition Designated Level 2 IS is selected by choosing the IS for which

- IS Type (as reported in the Level 1 LSP) is "Level 2 Intermediate System";
- ATT indicates "attached" by the default metric;
- P indicates support for the partition repair option; and
- ID is the lowest among the subset of attached Level 2 Intermediate Systems.

7.2.10.3 Computation of partition area addresses

For systems which do not implement partition repair, the value of `partitionAreaAddresses` is identical to the value computed for `areaAddresses` as described in 7.2.11. For systems which do implement partition repair, the value of `partitionAreaAddresses` is determined as follows.

A Level 2 Intermediate System shall compute the set of `partitionAreaAddresses`, the set of `partitionAreaAddresses` is defined as the union of all `manualAreaAddresses` as reported in the Level 1 Link state PDUs of all Level 1 and Level 2 Intermediate Systems reachable in the partition by the traversal of non-virtual links. If more than `maximumAreaAddresses` are present, the Intermediate system shall retain only those areas with numerically lowest area address (as described in 7.1.8). If one of the local system's `manualAreaAddresses` is so rejected the event `manualAddressDroppedFromArea` shall be generated.

7.2.10.4 Encapsulation of NPDUs across the virtual link

All NPDUs sent over virtual links shall be encapsulated as ISO 8473 Data NPDUs. The IS shall encapsulate NPDUs in ISO 8473 Data PDUs as follows:

- a) The Source Address field of the encapsulating Data NPDUs shall contain the Virtual Network Entity Title of the Partition Designated Level 2 IS that is forwarding the NPDUs over the virtual link
- b) The Destination Address field shall contain the Virtual NET of the "adjacent" Partition Designated Level 2 IS
- c) The SEL field in both addresses shall contain the `Intra-domainRouteingSelector` value.
- d) If the PDU to be encapsulated is an ISO 8473 Data or Error PDU the IS shall,
 - 1) Ensure that the header of the encapsulating PDU indicates routing by the default metric. Routing by the default metric may be indicated either by including a QoS parameter with its bits set in accordance with Table 1, or by carrying no QoS parameter at all [see 7.4.2 item a)].
 - 2) copy the Segmentation Permitted, and Error Report flags from the inner PDU.
 - 3) copy the Lifetime field from the inner NPDUs. When the inner NPDUs is decapsulated, its Lifetime field shall be set to the value of the Lifetime field in the outer NPDUs.
- e) If the PDU to be encapsulated is an ISO 8473 Data PDU, the IS shall not segment it before encapsulation, unless

the total length of the Data NPDUs (including header) exceeds 65512 octets. In that case, the original Data NPDUs shall first be segmented, then encapsulated. In all cases, the encapsulated Data NPDUs may need to be segmented by ISO 8473 before transmission in which case it must be reassembled and decapsulated by the destination Partition Designated Level 2 IS.

- f) If the PDU to be encapsulated is an LSP or SNP, the IS shall:
 - 1) set the Segmentation Permitted flag to "True";
 - 2) set the Error Report flag to "False";
 - 3) set the Lifetime field to 255. When an inner LSP is decapsulated, its remaining lifetime shall be decremented by half the difference between 255 and the value of the Lifetime field in the outer NPDUs.

The encapsulation is further described as part of the forwarding process in 7.4.3.2. The decapsulation is described as part of the Receive process in 7.4.4.

7.2.11 Computation of area addresses

A Level 1 or Level 2 Intermediate System shall compute the values of `areaAddresses` (the set of area addresses for this Level 1 area), by forming the union of the sets of `manualAreaAddresses` reported in the Area Addresses field of all Level 1 LSPs with LSP number zero in the local Intermediate system's link state database.

NOTES

- 14 This includes all source systems, whether currently reachable or not. It also includes the local Intermediate system's own Level 1 LSP with LSP number zero.
- 15 There is no requirement for this set to be updated immediately on each change to the database contents. It is permitted to defer the computation until the next running of the Decision Process.

If more than `maximumAreaAddresses` are present, the Intermediate system shall retain only those areas with numerically lowest area address (as described in 7.1.8). If one of the local system's `manualAreaAddresses` is rejected the event `manualAddressDroppedFromArea` shall be generated.

7.2.12 Order of preference of routes

7.2.12.1 If an Intermediate system takes part in level 1 routing, and determines (by looking at the area address) that a given destination is reachable within its area, then that destination will be reached exclusively by use of level 1 routing. In particular:

- a) Level 1 routing is always based on internal metrics.
- b) Amongst routes in the area, routes on which the requested QoS (if any) is supported are always preferred to routes on which the requested QoS is not supported.
- c) Amongst routes in the area of the same QoS, the shortest routes are preferred. For determination of the shortest path, if a route with specific QoS support is available, then the specified QoS metric is used, otherwise the default metric is used.

7.2.12.2 If an Intermediate system takes part in level 1 routing, does not take part in level 2 routing, and determines (by looking at the area address) that a given destination is *not* reachable within its area, and at least one attached level 2 IS is reachable in the area, then that destination will be reached by routing to a level 2 Intermediate system as follows:

- a) Level 1 routing is always based on internal metrics.
- b) Amongst routes in the area to attached level 2 ISs, routes on which the requested QoS (if any) is supported are always preferred to routes on which the requested QoS is not supported.
- c) Amongst routes in the area of the same QoS to attached level 2 ISs, the shortest route is preferred. For determination of the shortest path, if a route on which the specified QoS is available, then the specified QoS metric is used, otherwise the default metric is used.

7.2.12.3 If an Intermediate system takes part in level 2 routing and the IS determines (by looking at the area address) that the destination is reachable within its area, then the destination will be reached either by level 1 or level 2 routing, as follows:

- a) level 1 routing is always based on internal metrics;
- b) a level 2 route, if it exists, shall be preferred over a level 1 route that traverses a virtual link. Otherwise, the level 1 route shall be preferred;
- c) amongst routes in the area, routes on which the requested QoS (if any) is supported are always preferred to routes on which the requested QoS is not supported;
- d) amongst routes in the area of the same QoS, the shortest routes are preferred. For determination of the shortest path, if a route with specific QoS support is available, then the specified QoS metric is used, otherwise the default metric is used.

7.2.12.4 If an Intermediate system takes part in level 2 routing and is attached, and the IS determines (by looking at the area address) that a given destination is not reachable within its area, forwarding for that destination will select routes as follows:

- a) Routes on which the requested QoS (if any) is supported are always preferred to routes on which the requested QoS is not supported.
- b) Amongst routes of the same QoS, routes are prioritised as follows:
 - 1) Highest precedence: routes constructed from the Area Addresses information in an LSP (i.e. the path does not go outside the routing domain).
 - 2) Medium precedence: routes constructed from the Reachable Address Prefix information in an LSP which indicates an internal metric. In the case of multiple prefixes which match a given destination address which all have internal metrics, then the longest prefix shall be preferred.
 - 3) Lowest precedence: routes constructed from the Reachable Address Prefix information in an LSP

which indicates an external metric. In the case of multiple prefixes which match a given destination address which all have external metrics, then the longest prefix shall be preferred.

- c) For routes with equal precedence as specified above, the shortest path shall be preferred. For determination of the shortest path, a route supporting the specified QoS is used if available; otherwise a route using the default metric shall be used. Amongst routes of equal cost, load splitting may be performed.

7.3 Update process

The Update process is responsible for generating and propagating Link State information reliably throughout the routing domain.

The Link State information is used by the Decision process to calculate routes.

7.3.1 Input and output

INPUT

- Adjacency Database – maintained by the Subnetwork Dependent Functions
- Reachable Address managed objects - maintained by System Management
- Notification of Adjacency Database Change – notification by the Subnetwork Dependent Functions that an adjacency has come up, gone down, or changed cost. (Circuit up, Circuit down, Adjacency Up, Adjacency Down, and Cost change events)
- AttachedFlag – (level 2 Intermediate systems only), a flag computed by the Level 2 Decision Process indicating whether this system can reach (via level 2 routing) other areas
- Link State PDUs – The Receive Process passes Link State PDUs to the Update Process, along with an indication of which adjacency it was received on.
- Sequence Numbers PDUs – The Receive Process passes Sequence Numbers PDUs to the Update Process, along with an indication of which adjacency it was received on.
- Other Partitions – The Level 2 Decision Process makes available (to the Level 1 Update Process on a Level 2 Intermediate system) a list of (Partition Designated Level 2 Intermediate system, Level 2 default metric value) pairs, for other partitions of this area.

OUTPUT

- Link State Database
- Signal to the Decision Process of an event, which is either the receipt of a Link State PDU with different information from the stored one, or the purging of a Link State PDU from the database. The reception of a Link State PDU which has a different sequence number or Re-

maining Lifetime from one already stored in the database, but has an identical variable length portion, shall not cause such an event.

NOTE 16 An implementation may compare the checksum of the stored Link State PDU, modified according to the change in sequence number, with the checksum of the received Link State PDU. If they differ, the implementation may assume that the variable length portions are different and it shall signal an event to the Decision Process. However, if the checksums are the same, an octet for octet comparison is needed determine whether or not to signal the event.

7.3.2 Generation of local link state information

The Update Process is responsible for constructing a set of Link State PDUs. The purpose of these Link State PDUs is to inform all the other Intermediate systems (in the area, in the case of Level 1, or in the Level 2 subdomain, in the case of Level 2), of the state of the links between the Intermediate system that generated the PDUs and its neighbours.

The Update Process in an Intermediate system shall generate one or more new Link State PDUs under the following circumstances:

- a) upon timer expiration;
- b) when notified by the Subnetwork Dependent Functions of an Adjacency Database Change;
- c) when a change to some System Management parameter would cause the information in the LSP to change (for example, a change in manualAreaAddresses).

7.3.3 Use of manual routing information

Manual routing information is routing information entered by System Management. It may be specified in two forms.

- a) *Manual Adjacencies*
- b) *Reachable Addresses*

These are described in the following sub-clauses.

7.3.3.1 Manual adjacencies

An End system adjacency may be created by System Management. Such an adjacency is termed a *manual End system adjacency*. In order to create a manual End system adjacency, system management shall specify

- a) the (set of) system IDs reachable over that adjacency; and
- b) the corresponding SNPA Address.

A manual adjacency is considered to be “active” when all the following conditions are true:

- a) The `operationalState` of the linkage managed object associated with the parent circuit is “Enabled”
- b) the `adjacencyState` of the managed object associated with the adjacency is “On” and
- c) the parent circuit is of type broadcast or its underlying data link is operational.

Whenever a manual adjacency changes from being “inactive” to “active” a signal shall be generated to the Update process to cause it to include the IDs of the manual adjacency in the Level 1 LSPs generated by that system as described in 7.3.7.

Whenever a manual adjacency changes from being “active” to “inactive”, a signal shall be generated to the Update process to cause it to cease including the IDs of the manual adjacency in the Level 1 LSPs.

NOTE 17 Manual End system adjacencies are not included in a Level 1 LSPs issued on behalf of a pseudonode, since that would presuppose that all Intermediate systems on a broadcast subnetwork had the same set of manual adjacencies as defined for this circuit.

Metrics assigned to Manual adjacencies shall be internal metrics.

7.3.3.2 Reachable addresses

A Level 2 Intermediate system may have a number of `reachableAddress` managed objects created by System management. When a `reachableAddress` managed object is in `operationalState` “Enabled” and the linkage managed object associated with its parent Circuit is also in `operationalState` “Enabled”, the name and each of its defined routing metrics shall be included in Level 2 LSPs generated by this system.

Metrics assigned to `reachableAddress` managed objects may be either Internal or External.

A reachable address is considered to be “active” when all the following conditions are true:

- a) The `operationalState` of the the linkage managed object associated with the parent circuit is “Enabled”;
- b) the `operationalState` of the `reachableAddress` is “Enabled”; and
- c) the parent circuit is of type broadcast or its underlying data link is operational.

Whenever a reachable address changes from being “inactive” to “active” a signal shall be generated to the Update process to cause it to include the Address Prefix of the reachable address and each of its defined routing metrics in the Level 2 LSPs generated by that system as described in 7.3.9.

Whenever a reachable address changes from being “active” to “inactive”, a signal shall be generated to the Update process to cause it to cease including the Address Prefix of the reachable address in the Level 2 LSPs.

7.3.4 Multiple LSPs

7.3.4.1 Because a Link State PDU is limited in size to `ReceiveLSPBufferSize`, it may not be possible to include information about all of a system’s neighbours in a single LSP. In such cases, a system may use multiple LSPs to convey this information. Each LSP in the set carries the same `sourceID` field (see clause 9), but sets its own `LSP Number` field individually. Each of the several LSPs is handled independently by the Update Process, thus allowing distribution of topology updates to be pipelined. However, the Decision Process recognises that they all pertain to a common originating system because they all use the same `sourceID`.

NOTES

- 18 Even if the amount of information is small enough to fit in a single LSP, a system may optionally choose to use several LSPs to convey it; use of a single LSP in this situation is not mandatory.
- 19 In order to minimise the transmission of redundant information, it is advisable for an IS to group Reachable Address Prefix information by the circuit with which it is associated. Doing so will ensure that the minimum number of LSP fragments need be transmitted if a circuit to another routing domain changes state.

7.3.4.2 The maximum sized Level 1 or Level 2 LSP which may be generated by a system is controlled by the values of the management parameters `originatingL1LSPBufferSize` or `originatingL2LSPBufferSize` respectively.

NOTE 20 These parameters should be set consistently by system management. If this is not done, some adjacencies will fail to initialise.

7.3.4.3 The IS shall treat the LSP with LSP Number zero in a special way, as follows:

- a) The following fields are meaningful to the decision process only when they are present in the LSP with LSP Number zero:
 - 1) The setting of the LSP Database Overload bit.
 - 2) The value of the IS Type field.
 - 3) The Area Addresses option field. The Area Addresses option shall be present only in an LSP with LSP number zero. If there are more area addresses than will fit in a single instance of the Area Addresses option field, then the IS shall place 12 area addresses in each instance of the field except the last.
- b) When the values of any of the above items are changed, an Intermediate System shall reissue the LSP with LSP Number zero, to inform other Intermediate Systems of the change. Other LSPs need not be reissued.

7.3.4.4 Once a particular adjacency has been assigned to a particular LSP Number, it is desirable that it not be moved to another LSP Number. This is because moving an adjacency from one LSP to another can cause temporary loss of connectivity to that system. This can occur if the new version of the LSP which originally contained information about the adjacency (which now does not contain that information) is propagated before the new version of the other LSP (which now contains the information about the adjacency).

NOTE 21 An implementation is recommended to ensure that the number of LSPs generated for a particular system is within approximately 10 % of the optimal number which would be required if all LSPs were densely packed with neighbour options. Where possible this should be accomplished by re-using space in LSPs with a lower LSP Number for new adjacencies.

If it is necessary to move an adjacency from one LSP to another, the SRMflags (see 7.3.15) for the two new LSPs shall be set as an atomic action.¹⁾

7.3.4.5 When some event requires changing the LSP information for a system, the system shall reissue that (or those) LSPs which would have different contents. It is not required to reissue the unchanged LSPs. Thus a single End system adjacency change only requires the reissuing of the LSP containing the End System Neighbours option referring to that adjacency.

7.3.4.6 If an LSP becomes empty because all the adjacencies reported in that LSP no longer exist, an IS may purge that LSP instead of re-issuing it as an empty LSP (i.e. with no options) or not re-issuing it at all allowing the LSP to time out in the other ISs. This saves space in all of the LSP databases and hence is highly recommended.

7.3.5 Periodic LSP generation

The Update Process shall periodically re-generate and propa-

gate on every circuit with an IS adjacency of the appropriate level (by setting SRMflag on each circuit), all the LSPs (Level 1 and/or Level 2) for the local system and any pseudonodes for which it is responsible. The Intermediate system shall re-generate each LSP at intervals of at most maximumLSPGenerationInterval, with jitter applied as described in 10.1.

It is not required to synchronise the regeneration of the individual LSPs.

7.3.6 Event driven LSP Generation

In addition to the periodic generation of LSPs, an Intermediate system shall generate an LSP when an event occurs which would cause the information content to change. The following events may cause such a change.

- an Adjacency or Circuit Up/Down event
- a change in Circuit metric
- a change in Reachable Address metric
- a change in manualAreaAddresses
- a change in systemID
- a change in Designated Intermediate System status
- a change in the waiting status

When such an event occurs the IS shall re-generate changed LSP(s) with a new sequence number. If the event necessitated the generation of an LSP which had not previously been generated (for example, an adjacency "Up" event for an adjacency which could not be accommodated in an existing LSP), the sequence number shall be set to one. The IS shall then propagate the LSP(s) on every circuit by setting SRMflag for each circuit. The timer maximumLSPGenerationInterval shall not be reset.

There is a hold-down timer (minimumLSPGenerationInterval) on the generation of each individual LSP.

7.3.7 Generation of level 1 LSPs (non-pseudonode)

The Level 1 Link State PDU not generated on behalf of a pseudonode contains the following information in its variable length fields.

- In the Area Addresses option — the set of manualAreaAddresses for this Intermediate System.
- In the Intermediate System Neighbours option — the set of Intermediate system IDs of neighbouring Intermediate systems formed from
 - The set of neighbourSystemIDs with an appended zero octet (indicating non-pseudonode) from adjacencies in the state "Up", on circuits of type "Point-to-Point", "In" or "Out", with
 - x neighbourSystemType "L1 Intermediate System"

¹⁾ If the two SRMflags are not set atomically, a race condition will exist in which one of the two LSPs may be propagated quickly, while the other waits for an entire propagation cycle. If this occurs, adjacencies will be falsely eliminated from the topology and routes may become unstable for periods of time potentially as large as maximumLSPGenerationInterval.

- x neighbourSystemType "L2 Intermediate System" and adjacencyUsage "Level 1" or "Level 1 and 2".

The metrics shall be set to the values of Level 1 metric_k of the circuit for each supported routeing metric.

- The set of l1CircuitIDs for all circuits of type "Broadcast" (i.e. the neighbouring pseudonode IDs).

The metrics shall be set to the values of Level 1 metric_k of the circuit for each supported routeing metric, where k = one of (default, delay, error, expense).

- The set of IDs with an appended zero octet derived from the Network Entity Titles of all Virtual Adjacencies of this IS. (Note that the Virtual Flag is set when encoding these entries in the LSP — see 7.2.10.)

The default metric shall be set to the total cost to the virtual NET for the default routeing metric. The remaining metrics shall be set to the value indicating *unsupported*.

- In the End System Neighbours option — the set of IDs of neighbouring End systems formed from
 - The systemID of the Intermediate System itself, with a value of zero for all supported metrics.
 - The set of neighbourSystemIDs from all automatically-created adjacencies, in state "Up", on circuits of type "Point-to-Point", "In" or "Out", with neighbourSystemType "End system".

The metrics shall be set to the values of Level 1 metric_k of the circuit for each supported routeing metric, where k = one of (default, delay, error, expense).

- The set of neighbourSystemIDs from all system management-created adjacencies in state "Up", on all circuits, with neighbourSystemType "End system"

The metrics shall be set to the values of Level 1 metric_k of the circuit for each supported routeing metric, where k = one of (default, delay, error, expense).

- In the Authentication Information field — if the system's areaTransmitPassword is non-null, include the Authentication Information field containing an Authentication Type of "Password", and the value of the areaTransmitPassword.

7.3.8 Generation of level 1 pseudonode LSPs

An IS shall generate a Level 1 pseudonode Link State PDU for each circuit for which this Intermediate System is the Level 1 LAN Designated Intermediate System. The LSP shall specify the following information in its variable length fields. In all cases a value of zero shall be used for all supported routeing metrics

- The Area Addresses option is not present.

NOTE 22 This information is not required since the set of area addresses for the node issuing the pseudonode LSP will already have been made available via its own non-pseudonode LSP.

- In the Intermediate System Neighbours option — the set of Intermediate System IDs of neighbouring Intermediate Systems on the circuit for which this pseudonode LSP is being generated formed from
 - The Designated Intermediate System's own systemID with an appended zero octet (indicating non-pseudonode).
 - The set of neighbourSystemIDs with an appended zero octet (indicating non-pseudonode) from adjacencies on this circuit in the state "Up", with
 - x neighbourSystemType "L1 Intermediate System"
 - x "L2 Intermediate System" and adjacencyUsage "Level 1".

- In the End System Neighbours option — the set of IDs of neighbouring End systems formed from:
 - The set of neighbourSystemIDs from all automatically-created adjacencies, in state "Up", on the circuit for which this pseudonode is being generated, with neighbourSystemType "End system".

- In the Authentication Information field — if the system's areaTransmitPassword is non-null, include the Authentication Information field containing an Authentication Type of "Password", and the value of the areaTransmitPassword.

7.3.9 Generation of level 2 LSPs (non-pseudonode)

The Level 2 Link State PDU not generated on behalf of a pseudonode contains the following information in its variable length fields:

- In the Area Addresses option — the set of partitionAreaAddresses for this intermediate system as described in 7.2.10.3.
- In the Partition Designated Level 2 IS option — the ID of the Partition Designated Level 2 Intermediate System for the partition.
- In the Intermediate System Neighbours option — the set of Intermediate system IDs of neighbouring Intermediate systems formed from:
 - The set of neighbourSystemIDs with an appended zero octet (indicating non-pseudonode) from adjacencies in the state "Up", on circuits of type "Point-to-Point", "In" or "Out", with neighbourSystemType "L2 Intermediate System".
 - The set of l2CircuitIDs for all circuits whose linkage managed object is of type "Broadcast". (i.e. the neighbouring pseudonode IDs)

The metric and metric type shall be set to the values of Level 2 metric_k of the circuit for each supported routeing metric.

- In the Prefix Neighbours option — the set of variable length prefixes formed from:

- The set of prefixAddresses from all reachable-Address managed objects in operationalState “Enabled”, on all circuits in state “Up”.

The metrics shall be set to the values of Level 2 metric_k for the reachable address.

- In the Authentication Information field — if the system’s domainTransmitPassword is non-null, include the Authentication Information field containing an Authentication Type of “Password”, and the value of the domainTransmitPassword.

7.3.10 Generation of level 2 pseudonode LSPs

A Level 2 pseudonode Link State PDU is generated for each circuit for which this Intermediate System is the Level 2 LAN Designated Intermediate System and contains the following information in its variable length fields. In all cases a value of zero shall be used for all supported routing metrics.

- The Area Addresses option is not present.

NOTE 23 This information is not required since the set of area addresses for the node issuing the pseudonode LSP will already have been made available via its own non-pseudonode LSP.

- In the Intermediate System Neighbours option — the set of Intermediate System IDs of neighbouring Intermediate Systems on the circuit for which this pseudonode LSP is being generated formed from:
 - The Designated Intermediate System’s own systemID with an appended zero octet (indicating non-pseudonode).
 - The set of neighbourSystemIDs with an appended zero octet (indicating non-pseudonode) from adjacencies on this circuit in the state “Up” with neighbourSystemType “L2 Intermediate System”.
- The Prefix Neighbours option is not present.
- In the Authentication Information field — if the system’s domainTransmitPassword is non-null, include the Authentication Information field containing an Authentication Type of “Password”, and the value of the domainTransmitPassword.

7.3.11 Generation of the checksum

This International Standard makes use of the checksum function defined in ISO 8473.

The source IS shall compute the LSP Checksum when the LSP is generated. The checksum shall never be modified by any other system. The checksum allows the detection of memory corruptions and thus prevents both the use of incorrect routing information and its further propagation by the Update Process.

The checksum shall be computed over all fields in the LSP which appear after the Remaining Lifetime field. This field (and those appearing before it) are excluded so that the LSP may be aged by systems without requiring re-computation.

As an additional precaution against hardware failure, when the source computes the Checksum, it shall start with the two checksum variables (C0 and C1) initialised to what they would be after computing for the systemID portion of its Source ID. (This value is computed and stored when the Network entity is enabled and whenever systemID changes.) The IS shall then resume Checksum computation on the contents of the PDU after the first ID Length octets of the Source ID field.

NOTE 24 All Checksum calculations on the LSP are performed treating the Source ID field as the first octet. This procedure prevents the source from accidentally sending out Link State PDUs with some other system’s ID as source.

7.3.12 Initiating transmission

The IS shall store the generated Link State PDU in the Link State Database, overwriting any previous Link State PDU with the same LSP Number generated by this system. The IS shall then set all SRMflags for that Link State PDU, indicating it is to be propagated on all circuits with Intermediate System adjacencies.

An Intermediate system shall ensure (by reserving resources, or otherwise) that it will always be able to store and internalise its own non-pseudonode zeroth LSP. In the event that it is not capable of storing and internalising one of its own LSPs it shall enter the overloaded state as described in 7.3.19.1.

NOTE 25 It is recommended that an Intermediate system ensures (by reserving resources, or otherwise) that it will always be able to store and internalise all its own (zero and non-zero, pseudonode and non-pseudonode) LSPs.

7.3.13 Preservation of order

When an existing Link State PDU is re-transmitted (with the same or a different sequence number), but with the same information content (i.e. the variable length part) as a result of there having been no changes in the local topology databases, the order of the information in the variable length part shall be the same as that in the previously transmitted LSP.

NOTE 26 If a sequence of changes results in the state of the database returning to some previous value, there is no requirement to preserve the ordering. It is only required when there have been no changes whatever. This allows the receiver to detect that there has been no change in the information content by performing an octet for octet comparison of the variable length part, and hence not re-run the decision process.

7.3.14 Propagation of LSPs

7.3.14.1 The update process is responsible for propagating Link State PDUs throughout the domain (or in the case of Level 1, throughout the area).

The basic mechanism is flooding, in which each Intermediate system propagates to all its neighbour Intermediate systems except that neighbour from which it received the PDU. Duplicates are detected and dropped.

7.3.14.2 Link state PDUs are received from the Receive Process. The maximum size control PDU (Link State PDU or Sequence Numbers PDU) which a system expects to receive shall

be `ReceiveLSPBufferSize` octets. (i.e. the Update process must provide buffers of at least this size for the reception, storage and forwarding of received Link State PDUs and Sequence Numbers PDUs.) If a control PDU larger than this size is received, it shall be treated as if it had an invalid checksum (i.e. ignored by the Update Process and a `corruptedLSPReceived` event generated).

Upon receipt of a Link State PDU the Update Process shall perform the following functions:

- a) Level 2 Link State PDUs shall be propagated on circuits which have at least one Level 2 adjacency.
- b) Level 1 Link State PDUs shall be propagated on circuits which have at least one Level 1 adjacency or at least one Level 2 adjacency not marked "Level 2 only, and virtual links".
- c) When propagating a Level 1 Link State PDU on a broadcast subnetwork, the IS shall transmit to the multi-destination subnetwork address `AllL1IS`.
- d) When propagating a Level 2 Link State PDU on a broadcast subnetwork, the IS shall transmit to the multi-destination subnetwork address `AllL2IS`.

NOTE 27 When propagating a Link State PDU on a general topology subnetwork the Data Link Address is unambiguous (because Link State PDUs are not propagated across Dynamically Assigned circuits).

- e) An Intermediate system receiving a Link State PDU with an incorrect LSP Checksum or with an invalid PDU syntax shall
 - 1) generate a `corruptedLSPReceived` circuit event,
 - 2) overwrite the Checksum and Remaining Lifetime with zero, and
 - 3) treat the Link State PDU as though its Remaining Lifetime had expired (see 7.3.16.4.)
- f) A Intermediate system receiving a Link State PDU which is new (as identified in 7.3.16) shall
 - 1) store the Link State PDU into Link State database, and
 - 2) mark it as needing to be propagated upon all circuits except that upon which it was received.
- g) When a Intermediate system receives a Link State PDU from source *S*, which it considers older than the one stored in the database for *S*, it shall set the `SRMflag` for *S*'s Link State PDU associated with the circuit from which the older Link State PDU was received. This indicates that the stored Link State PDU needs to be sent on the link from which the older one was received.
- h) When a system receives a Link State PDU which is the same (not newer or older) as the one stored, the Intermediate system shall
 - 1) acknowledge it if necessary, as described in 7.3.17, and

- 2) clear the `SRMflag` for that circuit for that Link State PDU.

- i) A Link State PDU received with a zero checksum shall be treated as if the Remaining Lifetime were zero. The age, if not zero, shall be overwritten with zero.

7.3.14.3 The Update Process scans the Link State Database for Link State PDUs with `SRMflags` set. When one is found, provided the timestamp `lastSent` indicates that it was propagated no more recently than `minimumLSPTransmissionInterval`, the IS shall

- a) transmit it on all circuits with `SRMflags` set, and
- b) update `lastSent`.

7.3.15 Manipulation of SRM and SSN flags

For each Link State PDU, and for each circuit over which routing messages are to be exchanged (i.e. not on DA circuits), there are two flags:

Send Routeing Message (SRMflag) – if set, indicates that Link State PDU should be transmitted on that circuit. On broadcast circuits `SRMflag` is cleared as soon as the LSP has been transmitted, but on non-broadcast circuits `SRMflag` is only cleared on reception of a Link State PDU or Sequence Numbers PDU as described below.

`SRMflag` shall never be set for an LSP with sequence number zero, nor on a circuit whose `externalDomain` attribute is "True" (See 7.3.15.2).

Send Sequence Numbers (SSNflag) – if set, indicates that information about that Link State PDU should be included in a Partial Sequence Numbers PDU transmitted on that circuit with an associated linkage. When the Sequence Numbers PDU has been transmitted `SSNflag` is cleared. Note that the Partial Sequence Numbers PDU serves as an acknowledgement that a Link State PDU was received.

`SSNflag` shall never be set on a circuit whose `externalDomain` attribute is "True".

7.3.15.1 Action on receipt of a link state PDU

When a Link State PDU is received on a circuit *C*, the IS shall perform the following functions

- a) Perform the following PDU acceptance tests:
 - 1) If the LSP was received over a circuit whose `externalDomain` attribute is "True", the IS shall discard the PDU.
 - 2) If the ID Length field of the PDU is not equal to the value of the IS's `routeingDomainIDLength`, the PDU shall be discarded and an `iDFieldLengthMismatch` event generated.
 - 3) If this is a level 1 LSP, and the `Maximum Area Addresses` field of the PDU is not equal to the value of the IS's `maximumAreaAddresses` then the PDU

shall be discarded and a maximumAreaAddressesMismatch event generated, unless the IS only implements a value of three for maximumAreaAddresses, in which case this check may be omitted.

- 4) If the LSP is received on a broadcast circuit and the source subnetwork address of the LSP does not match the neighbourSNPAddress of an adjacency of the same level (e.g. a level 1 LSP with a level 1 adjacency) on the circuit over which the LSP was received, then the IS shall discard the LSP without generating an event.

If the LSP is received on a non-broadcast circuit and there is no adjacency of the same level (e.g. a level 1 LSP with a level 1 or level 1 & 2 adjacency) on the circuit over which the LSP was received, then the IS shall discard the LSP without generating an event.

- 5) If this is a level 1 LSP, and the value of areaTransmitPassword or the set of areaReceivePasswords is non-null, then perform the following tests:
 - i) If the PDU does not contain the Authentication Information field then the PDU shall be discarded and an authenticationFailure event generated.
 - ii) If the PDU contains the Authentication Information field, but the Authentication Type is not equal to "Password", then:
 - (a) If the IS implements the authentication procedure indicated by the Authentication Type whether the IS accepts or ignores the PDU is outside the scope of this International Standard.
 - (b) If the IS does not implement the authentication procedure indicated by the Authentication Type then the IS shall ignore the PDU and generate an authenticationFailure event."
 - iii) Otherwise, the IS shall compare the password in the received PDU with the passwords in the set of areaReceivePasswords, augmented by the value of the areaTransmitPassword. If the value in the PDU matches any of these passwords, the IS shall accept the PDU for further processing. If the value in the PDU does not match any of the above values, then the IS shall ignore the PDU and generate an authenticationFailure event.
- 6) If this is a level 2 LSP, and the value of domainTransmitPassword or the set of domainReceivePasswords is non-null, then perform the following tests:
 - i) If the PDU does not contain the Authentication Information field then the PDU shall be discarded and an authenticationFailure event generated.
 - ii) If the PDU contains the Authentication Information field, but the Authentication Type is not equal to "Password", then:
 - (a) If the IS implements the authentication procedure indicated by the Authentication Type

whether the IS accepts or ignores the PDU is outside the scope of this International Standard.

- (b) If the IS does not implement the authentication procedure indicated by the Authentication Type then the IS shall ignore the PDU and generate an authenticationFailure event."
 - iii) Otherwise, the IS shall compare the password in the received PDU with the passwords in the set of domainReceivePasswords, augmented by the value of the domainTransmitPassword. If the value in the PDU matches any of these passwords, the IS shall accept the PDU for further processing. If the value in the PDU does not match any of the above values, then the IS shall ignore the PDU and generate an authenticationFailure event.
- b) If the LSP has zero Remaining Lifetime, perform the actions described in 7.3.16.4.
 - c) If the source *S* of the LSP is an IS or pseudonode for which all but the last octet are equal to the systemID of the receiving Intermediate System, and the receiving Intermediate System does not have that LSP in its database, or has that LSP, but no longer considers it to be in the set of LSPs generated by this system (e.g. it was generated by a previous incarnation of the system), then initiate a network wide purge of that LSP as described in 7.3.16.4.
 - d) If the source *S* of the LSP is a system (pseudonode or otherwise) for which the first ID Length octets are equal to the systemID of the receiving Intermediate system, and the receiving Intermediate system has an LSP in the set of currently generated LSPs from that source in its database (i.e. it is an LSP generated by this Intermediate system), perform the actions described in 7.3.16.1.
 - e) Otherwise, (the source *S* is some other system),
 - 1) If the LSP is newer than the one in the database, or if an LSP from that source does not yet exist in the database:
 - i) Store the new LSP in the database, overwriting the existing database LSP for that source (if any) with the received LSP.
 - ii) Set SRMflag for that LSP for all circuits other than *C*.
 - iii) Clear SRMflag for *C*.
 - iv) If *C* is a non-broadcast circuit, set SSNflag for that LSP for *C*.
 - v) Clear SSNflag for that LSP for the circuits associated with a linkage other than *C*.
 - 2) If the LSP is equal to the one in the database (same Sequence Number, Remaining Lifetimes both zero or both non-zero, same checksums):
 - i) Clear SRMflag for *C*.
 - ii) If *C* is a non-broadcast circuit, set SSNflag for that LSP for *C*.

3) If the LSP is older than the one in the database:

- i) Set SRMflag for C.
- ii) Clear SSNflag for C.

When storing a new LSP, the Intermediate system shall first ensure that it has sufficient memory resources to both store the LSP and generate whatever internal data structures will be required to process the LSP by the Update Process. If these resources are not available the LSP shall be ignored. It shall neither be stored nor acknowledged. When an LSP is ignored for this reason the IS shall enter the "Waiting State" (See 7.3.19).

When attempting to store a new version of an existing LSP (with the same LSPID), which has a length less than or equal to that of the existing LSP, the existing LSP shall be removed from the routing information base and the new LSP stored as a single atomic action. This ensures that such an LSP (which may be carrying the LSP Database Overload indication from an overloaded IS) will never be ignored as a result of a lack of memory resources.

7.3.15.2 Action on receipt of a sequence numbers PDU

When a Sequence Numbers PDU (Complete or Partial, see 7.3.17) is received on circuit C the IS shall perform the following functions:

a) Perform the following PDU acceptance tests:

- 1) If the SNP was received over a circuit whose externalDomain attribute is "True", the IS shall discard the PDU.
- 2) If circuit C is a broadcast circuit and either
 - i) this is a level 1 PSNP and this IS is not the level 1 designated IS for circuit C, or
 - ii) this is a level 2 PSNP and this IS is not the level 2 designated IS for circuit C, then the IS shall discard the PDU.
- 3) If the ID Length field of the PDU is not equal to the value of the IS's routingDomainIDLength, the PDU shall be discarded and an idFieldLengthMismatch event generated.
- 4) If this is a level 1 SNP, and the Maximum Area Addresses field of the PDU is not equal to the value of the IS's maximumAreaAddresses then the PDU shall be discarded and a maximumAreaAddressesMismatch event generated, unless the IS only implements a value of three for maximumAreaAddresses, in which case this check may be omitted.
- 5) If the SNP is received on a broadcast circuit and the source subnetwork address of the SNP does not match the neighborSNPAAddress of an adjacency of the same level (e.g. a level 1 SNP with a level 1 adjacency) on the circuit over which the SNP was received, then the IS shall discard the SNP without generating an event.

If the SNP is received on a non-broadcast circuit and there is no adjacency of the same level (e.g. a level 1 SNP with a level 1 or level 1 & 2 adjacency) on the circuit over which the SNP was received, then the IS shall discard the SNP without generating an event.
- 6) If this is a level 1 SNP and the value of areaTransmitPassword or the set of areaReceivePasswords is non-null, then perform the following tests:

- i) If the PDU does not contain the Authentication Information field then the PDU shall be discarded and an authenticationFailure event generated.
- ii) If the PDU contains the Authentication Information field, but the Authentication Type is not equal to "Password", then:
 - (a) If the IS implements the authentication procedure indicated by the Authentication Type whether the IS accepts or ignores the PDU is outside the scope of this International Standard.
 - (b) If the IS does not implement the authentication procedure indicated by the Authentication Type then the IS shall ignore the PDU and generate an authenticationFailure event.
- iii) Otherwise, the IS shall compare the password in the received PDU with the passwords in the set of areaReceivePasswords, augmented by the value of the areaTransmitPassword. If the value in the PDU matches any of these passwords, the IS shall accept the PDU for further processing. If the value in the PDU does not match any of the above values, then the IS shall ignore the PDU and generate an authenticationFailure event.

7) If this is a level 2 SNP, and the value of domainTransmitPassword or the set of domainReceivePasswords is non-null, then perform the following tests:

- i) If the PDU does not contain the Authentication Information field then the PDU shall be discarded and an authenticationFailure event generated.
- ii) If the PDU contains the Authentication Information field, but the Authentication Type is not equal to "Password", then:
 - (a) If the IS implements the authentication procedure indicated by the Authentication Type whether the IS accepts or ignores the PDU is outside the scope of this International Standard.
 - (b) If the IS does not implement the authentication procedure indicated by the Authentication Type then the IS shall ignore the PDU and generate an authenticationFailure event.
- iii) Otherwise, the IS shall compare the password in the received PDU with the passwords in the set of domainReceivePasswords, augmented by the value of the domainTransmitPassword. If the value in the PDU matches any of these passwords, the IS shall accept the PDU for further processing. If the value in the PDU does not match any of the above values, then the IS shall ignore the PDU and generate an authenticationFailure event.

b) For each LSP reported in the Sequence Numbers PDU:

- 1) If the reported value equals the database value and C is a non-broadcast circuit, Clear SRMflag for C for that LSP.

- 2) If the reported value is older than the database value, Clear SSNflag, and Set SRMflag.
- 3) If the reported value is newer than the database value, Set SSNflag, and if *C* is a non-broadcast circuit Clear SRMflag.
- 4) If no database entry exists for the LSP, and the reported Remaining Lifetime, Checksum and Sequence Number fields of the LSP are all non-zero, create an entry with sequence number 0 (see 7.3.16.1), and set SSNflag for that entry and circuit *C*. Under no circumstances shall SRMflag be set for such an LSP with zero sequence number.

NOTE 28 This is because possessing a zero sequence number LSP is semantically equivalent to having no information about that LSP. If such LSPs were propagated by setting SRMflag it would result in an unnecessary consumption of both bandwidth and memory resources.

- c) If the Sequence Numbers PDU is a Complete Sequence Numbers PDU, Set SRMflags for *C* for all LSPs in the database (except those with zero sequence number or zero Remaining Lifetime) with LSPIDs within the range specified for the CSNP by the Start LSPID and End LSPID fields, which were not mentioned in the Complete Sequence Numbers PDU (i.e. LSPs this system has, which the neighbour does not claim to have).

7.3.15.3 Action on expiration of complete SNP interval

The IS shall perform the following actions every Complete-SNPInterval for circuit *C*:

- a) If *C* is a broadcast circuit, then
 - 1) If this Intermediate system is a Level 1 Designated Intermediate System on circuit *C*, transmit a complete set of Level 1 Complete Sequence Numbers PDUs on circuit *C*. Ignore the setting of SSNflag on Level 1 Link State PDUs.

If the value of the IS's areaTransmitPassword is non-null, then the IS shall include the Authentication Information field in the transmitted CSNP, indicating an Authentication Type of "Password" and containing the areaTransmitPassword as the authentication value.

- 2) If this Intermediate system is a Level 2 Designated Intermediate System on circuit *C*, transmit a complete set of Level 2 Complete Sequence Numbers PDUs on circuit *C*. Ignore the setting of SSNflag on Level 2 Link State PDUs.

If the value of the IS's domainTransmitPassword is non-null, then the IS shall include the Authentication Information field in the transmitted CSNP, indicating an Authentication Type of "Password" and containing the domainTransmitPassword as the authentication value.

A complete set of CSNPs is a set whose Start LSPID and End LSPID ranges cover the complete possible

range of LSPIDs. (i.e. there is no possible LSPID value which does not appear within the range of one of the CSNPs in the set). Where more than one CSNP is transmitted on a broadcast circuit, they shall be separated by an interval of at least minimumBroadcastLSPTransmissionInterval.

NOTE 29 An IS is permitted to transmit a small number of CSNPs (no more than 10) with a shorter separation interval, (or even "back to back"), provided that no more than $1000 / \text{minimumBroadcastLSPTransmissionInterval}$ CSNPs are transmitted in any one second period.

- b) Otherwise (*C* is a point-to-point circuit, including non-DA DED circuits and virtual links), do nothing. CSNPs are only transmitted on point-to-point circuits at initialisation.

7.3.15.4 Action on expiration of partial SNP interval

The maximum sized Level 1 or Level 2 PSNP which may be generated by a system is controlled by the values of originatingL1LSPBufferSize or originatingL2LSPBufferSize respectively. An Intermediate system shall perform the following actions every partialSNPInterval for circuit *C* with jitter applied as described in 10.1:

- a) If *C* is a broadcast circuit, then

- 1) If this Intermediate system is a Level 1 Intermediate System or a Level 2 Intermediate System with manualL2OnlyMode "False", but is not a Level 1 Designated Intermediate System on circuit *C*, transmit a Level 1 Partial Sequence Numbers PDU on circuit *C*, containing entries for as many Level 1 Link State PDUs with SSNflag set as will fit in the PDU, and then clear SSNflag for these entries. To avoid the possibility of starvation, the scan of the LSP database for those with SSNflag set shall commence with the next LSP which was not included in the previous scan. If there were no Level 1 Link State PDUs with SSNflag set, do not transmit a Level 1 Partial Sequence Numbers PDU.

If the value of the IS's areaTransmitPassword is non-null, then the IS shall include the Authentication Information field in the transmitted PSNP, indicating an Authentication Type of "Password" and containing the areaTransmitPassword as the authentication value.

- 2) If this Intermediate system is a Level 2 Intermediate System, but is not a Level 2 Designated Intermediate System on circuit *C*, transmit a Level 2 Partial Sequence Numbers PDU on circuit *C*, containing entries for as many Level 2 Link State PDUs with SSNflag set as will fit in the PDU, and then clear SSNflag for these entries. To avoid the possibility of starvation, the scan of the LSP database for those with SSNflag set shall commence with the next LSP which was not included in the previous scan. If there were no Level 2 Link State PDUs with SSNflag set, do not transmit a Level 2 Partial Sequence Numbers PDU.

If the value of the IS's domainTransmitPassword is non-null, then the IS shall include the Authentication Information field in the transmitted PSNP, indicating an Authentication Type of "Password" and containing the domainTransmitPassword as the authentication value.

b) Otherwise (C is a point-to-point circuit, including non-DA DED circuits and virtual links)

- 1) If this system is a Level 1 Intermediate system, transmit a Level 1 Partial Sequence Numbers PDU on circuit C, containing entries for as many Level 1 Link State PDUs with SSNflag set as will fit in the PDU, and then clear SSNflag for these entries. To avoid the possibility of starvation, the scan of the LSP database for those with SSNflag set shall commence with the next LSP which was not included in the previous scan. If there were no Level 1 Link State PDUs with SSNflag set, do not transmit a Partial Sequence Numbers PDU.

If the value of the IS's areaTransmitPassword is non-null, then the IS shall include the Authentication Information field in the transmitted PSNP, indicating an Authentication Type of "Password" and containing the areaTransmitPassword as the authentication value.

- 2) If this system is a Level 2 Intermediate system, transmit a Level 2 Partial Sequence Numbers PDU on circuit C, containing entries for as many Level 2 Link State PDUs with SSNflag set as will fit in the PDU, and then clear SSNflag for these entries. To avoid the possibility of starvation, the scan of the LSP database for those with SSNflag set shall commence with the next LSP which was not included in the previous scan. If there were no Level 2 Link State PDUs with SSNflag set, do not transmit a Partial Sequence Numbers PDU.

If the value of the IS's domainTransmitPassword is non-null, then the IS shall include the Authentication Information field in the transmitted PSNP, indicating an Authentication Type of "Password" and containing the domainTransmitPassword as the authentication value.

7.3.15.5 Action on expiration of the minimumLSPTransmissionInterval

An IS shall perform the following action every minimumLSPTransmissionInterval with jitter applied as described in 10.1:

- For all point-to-point circuits C (including non-DA DED circuits and virtual links) transmit all LSPs that have SRMflag set on circuit C, but **do not** clear the SRMflag. The SRMflag will subsequently be cleared by receipt of a Complete or Partial Sequence Numbers PDU.

The interval between two consecutive transmissions of the same LSP shall be at least minimumLSPTransmissionInter-

val. Clearly, this can only be achieved precisely by keeping a separate timer for each LSP. This would be an unwarranted overhead. Any technique which ensures the interval will be between minimumLSPTransmissionInterval and $2 \times \text{minimumLSPTransmissionInterval}$ is acceptable.

7.3.15.6 Controlling the rate of transmission on broadcast circuits

The attribute minimumBroadcastLSPTransmissionInterval indicates the minimum interval between PDU arrivals which can be processed by the slowest Intermediate System on the LAN.

Setting SRMflags on an LSP for a broadcast circuit does not cause the LSP to be transmitted immediately. Instead the Intermediate system shall scan the LSP database every minimumBroadcastLSPTransmissionInterval (with jitter applied as described in 10.1), and from the set of LSPs which have SRMflags set for this circuit, one LSP shall be chosen at random. This LSP shall be multicast on the circuit, and SRMflags cleared.

NOTES

- 30 In practice it would be very inefficient to scan the whole database at this rate, particularly when only a few LSPs had SRMflags set. Implementations may require additional data structures in order to reduce this overhead.
- 31 An IS is permitted to transmit a small number of LSPs (no more than 10) with a shorter separation interval, (or even "back to back"), provided that no more than 1000/minimumBroadcastLSPTransmissionInterval LSPs are transmitted in any one second period.

In addition, the presence of any LSPs which have been received on a particular circuit and are queued awaiting processing shall inhibit transmission of LSPs on that circuit. However, LSPs may be transmitted at a minimum rate of one per second even in the presence of such a queue.

7.3.16 Determining the latest information

The Update Process is responsible for determining, given a received link state PDU, whether that received PDU represents new, old, or duplicate information with respect to what is stored in the database.

It is also responsible for generating the information upon which this determination is based, for assigning a sequence number to its own Link State PDUs upon generation, and for correctly adjusting the Remaining Lifetime field upon broadcast of a link state PDU generated originally by any system in the domain.

7.3.16.1 Sequence numbers

The sequence number is a 4 octet unsigned value. Sequence numbers shall increase from zero to (SequenceModulus - 1). When a system initialises, it shall start with sequence number one for its own Link State PDUs.¹⁾

¹⁾ The IS starts with 1 rather than 0 so that the value 0 can be reserved to be guaranteed to be less than the sequence number of any actually generated Link State PDU. This is a useful property for Sequence Numbers PDUs.

The sequence numbers the Intermediate system generates for its Link State PDUs with different values for LSP number are independent. The algorithm for choosing the numbers is the same, but operationally the numbers will not be synchronised.

If an Intermediate system *R* somewhere in the domain has information that the current sequence number for source *S* is greater than that held by *S*, *R* will return to *S* a Link State PDU for *S* with *R*'s value for the sequence number. When *S* receives this LSP it shall change its sequence number to be the next number greater than the new one received, and shall generate a link state PDU.

If an Intermediate system needs to increment its sequence number, but the sequence number is already equal to `SequenceModulus - 1`, the event `attemptToExceedMaximumSequenceNumber` shall be generated and the IS Network entity shall be disabled for a period of at least `MaxAge + ZeroAgeLifetime`, in order to be sure that any versions of this LSP with the high sequence number have expired. When it is re-enabled the IS shall start again with sequence number 1.

7.3.16.2 LSP confusion

It is possible for an LSP generated by a system in a previous incarnation to be alive in the domain and have the same sequence number as the current LSP.

To ensure database consistency among the Intermediate Systems, it is essential to distinguish two such PDUs. This is done efficiently by comparing the checksum on a received LSP with the one stored in memory.

If the sequence numbers match, but the checksums do not and the LSP is not in the current set of LSPs generated by the local system, then the system that notices the mismatch shall treat the LSP as if its Remaining Lifetime had expired. It shall store one of the copies of the LSP, with zero written as the Remaining Lifetime, and flood the LSP.

If the LSP is in the current set of LSPs generated by the local system then the IS shall change the LSP's sequence number to be the next number greater than that of the received LSP and regenerate the LSP.

7.3.16.3 Remaining lifetime field

When the source generates a link state PDU, it shall set the Remaining Lifetime to `MaxAge`.

When a system holds the information for some time before successfully transmitting it to a neighbour, that system shall decrement the Remaining Lifetime field according to the holding time. Before transmitting a link state PDU to a neighbour, a system shall decrement the Remaining Lifetime in the PDU being transmitted by at least one, or more than one if the transit time to that neighbour is estimated to be greater than one second. When the Remaining Lifetime field reaches zero, the system shall purge that Link State PDU from its database. In order to keep the Intermediate Systems' databases synchronised, the purging of an LSP due to Remaining Lifetime expiration is synchronised by flooding an expired LSP. See 7.3.16.4.

If the RemainingLifetime of the received LSP is zero it shall be processed as described in 7.3.16.4. If the Remaining Life-

time of the received LSP is non-zero, but there is an LSP in the database with the same sequence number and zero Remaining Lifetime, the LSP in the database shall be considered most recent. Otherwise, the PDU with the larger sequence number shall be considered the most recent.

If the value of Remaining Lifetime is greater than `MaxAge`, the LSP shall be processed as if there were a checksum error.

7.3.16.4 LSP expiration synchronisation

When the Remaining Lifetime on an LSP in memory becomes zero, the IS shall

- a) set all SRMflags for that LSP, and
- b) retain only the LSP header.
- c) record the time at which the Remaining Lifetime for this LSP became zero. When `ZeroAgeLifetime` has elapsed since the LSP Remaining Lifetime became zero, the LSP header shall be purged from the database.

NOTE 32 A check of the checksum of a zero Remaining Lifetime LSP succeeds even though the data portion is not present

When a purge of an LSP with non-zero Remaining Lifetime is initiated, the header shall be retained for `MaxAge`.

If an LSP from source *S* with zero Remaining Lifetime is received on circuit *C* :

- a) If no LSP from *S* is in memory, then the IS shall
 - 1) send an acknowledgement of the LSP on circuit *C*, but
 - 2) shall not retain the LSP after the acknowledgement has been sent.
- b) If an LSP from *S* is in the database, then
 - 1) If the received LSP is newer than the one in the database (i.e. received LSP has higher sequence number, or same sequence number and database LSP has non-zero Remaining Lifetime) the IS shall:
 - i) overwrite the database LSP with the received LSP, and note the time at which the zero Remaining Lifetime LSP was received, so that after `ZeroAgeLifetime` has elapsed, that LSP can be purged from the database,
 - ii) set SRMflag for that LSP for all circuits,
 - iii) clear SSNflag for that LSP for all circuits.
 - 2) If the received LSP is equal to the one in the database (i.e. same Sequence Number, Remaining Lifetimes both zero) the IS shall
 - i) clear SRMflag for *C*, and
 - ii) if *C* is a non-broadcast circuit, set SSNflag for that LSP for *C*.

- 3) If the received LSP is older than the one in the database (i.e. received LSP has lower sequence number) the IS shall
 - i) set SRMflag for C, and
 - ii) clear SSNflag for C.
- c) If this system (or pseudonode) is S and there is an un-expired LSP from S (i.e. its own LSP) in memory, then the IS
 - 1) shall not overwrite with the received LSP, but
 - 2) shall change the sequence number of the un-expired LSP from S as described in 7.3.16.1,
 - 3) generate a new LSP; and
 - 4) set SRMflag on all circuits.

7.3.17 Making the update reliable

The update process is responsible for making sure the latest link state PDUs reach every reachable Intermediate System in the domain.

On point-to-point links the Intermediate system shall send an explicit acknowledgement encoded as a Partial Sequence Numbers PDU (PSNP) containing the following information:

- a) source's ID
- b) PDU type (Level 1 or 2)
- c) sequence number
- d) Remaining Lifetime
- e) checksum

This shall be done for all received link state PDUs which are newer than the one in the database, or duplicates of the one in the database. Link state PDUs which are older than that stored in the database are answered instead by a newer link state PDU, as specified in 7.3.14 above.

On broadcast links, instead of explicit acknowledgements for each link state PDU by each Intermediate system, a special PDU known as a Complete Sequence Numbers PDU (CSNP), shall be multicast periodically by the Designated Intermediate System. The PDU shall contain a list of all LSPs in the database, together with enough information so that Intermediate systems receiving the CSNP can compare with their LSP database to determine whether they and the CSNP transmitter have synchronised LSP databases. The maximum sized Level 1 or Level 2 Sequence Numbers PDU which may be generated by a system is controlled by the values of `originatingL1LSPBufferSize` or `originatingL2LSPBufferSize` respectively. In practice, the information required to be transmitted in a single CSNP may be greater than will fit in a single PDU. Therefore each CSNP carries an inclusive range of LSPIDs to which it refers. The complete set of information shall be conveyed by transmitting a series of individual CSNPs, each referring to a subset of the complete range. The ranges of the complete set of CSNPs shall be contiguous (though not necessarily transmitted in order) and shall cover the entire range of possible LSPIDs.

The LAN Level 1 Designated Intermediate System shall periodically multicast complete sets of Level 1 CSNPs to the multi-destination address AllL1ISs. The LAN Level 2 Designated Intermediate System shall periodically multicast com-

plete sets of Level 2 CSNPs to the multi-destination address AllL2ISs.

Absence of an LSPID from a Complete Sequence Numbers PDU whose range includes that LSPID indicates total lack of information about that LSPID.

If an Intermediate system, upon receipt of a Complete Sequence Numbers PDU, detects that the transmitter was out of date, the receiver shall multicast the missing information.

NOTE 33 Receipt of a link state PDU on a link is the same as successfully transmitting the Link State PDU on that link, so once the first Intermediate system responds, no others will, unless they have already transmitted replies.

If an Intermediate system detects that the transmitter had more up to date information, the receiving Intermediate system shall multicast a Partial Sequence Numbers PDU (PSNP), containing information about LSPs for which it has older information. This serves as an implicit request for the missing information. Although the PSNP is multicast, only the Designated Intermediate System of the appropriate level shall respond to the PSNP.

NOTE 34 This is equivalent to the PSNP being transmitted directly to the Designated Intermediate System, in that it avoids each Intermediate System unnecessarily sending the same LSP(s) in response. However, it has the advantage of preserving the property that all routing messages can be received on the multi-destination addresses, and hence by a LAN adapter dedicated to the multi-destination address.

When a point-to-point circuit (including non-DA DED circuits and virtual links) starts (or restarts), the IS shall

- a) set SRMflag for that circuit on all LSPs, and
- b) send a Complete set of Complete Sequence Numbers PDUs on that circuit.

7.3.18 Validation of databases

An Intermediate System shall not continue to operate for an extended period with corrupted routing information. The IS shall therefore operate in a *fail-stop* manner. If a failure is detected, the Intermediate system Network entity shall be disabled until the failure is corrected. In the absence of an implementation-specific method for ensuring this, the IS shall perform the following at least every `maximumLSPGenerationInterval`:

- a) On expiration of this timer the IS shall re-check the checksum of every LSP in the LSP database (except those with a Remaining Lifetime of zero) in order to detect corruption of the LSP while in memory. If the checksum of any LSP is incorrect, the event `corruptedLSP-Detected` shall be logged, and as a minimum the entire Link State Database shall be deleted and action taken to cause it to be re-acquired. One way to achieve this is to disable and re-enable the IS Network entity.

NOTE 35 On point-to-point links, this requires at least that a CSNP be transmitted.

- b) On completion of these checks the decision process shall be notified of an event (even if any newly generated

LSPs have identical contents to the previous ones). This causes the decision process to be run and the forwarding databases re-computed, thus protecting against possible corruption of the forwarding databases in memory, which would not otherwise be detected in a stable topology.

- c) The IS shall reset the timer for a period of maximum-LSPGenerationInterval with jitter applied as described in 10.1.

7.3.19 LSP database overload

As a result of network mis-configuration, or certain transitory conditions, it is possible that there may be insufficient memory resources available to store a received Link State PDU. When this occurs, an IS needs to take certain steps to ensure that if its LSP database becomes inconsistent with the other ISs', that these ISs do not rely on forwarding paths through the overloaded IS.

7.3.19.1 Entering the waiting state

When an LSP cannot be stored, the LSP shall be ignored and Waiting State shall be entered. A timer shall be started for waitingTime, and the Intermediate System shall generate and flood its own LSP with zero LSP number with the LSP Database Overload Bit set. This prevents this Intermediate system from being considered as a forwarding path by other Intermediate Systems.

It is possible that although there are sufficient resources to store an LSP and permit the operation of the Update Process on that LSP, the Decision Process may subsequently require further resources in order to complete. If these resources are not available, the Intermediate system shall then (i.e. during the attempt to run the Decision Process) enter Waiting State until such time as they are available and waitingTime have elapsed since the last LSP was ignored by the Update Process.

An implementation shall partition the available memory resources between the Level 1 and Level 2 databases. An overload condition can therefore exist independently for Level 1 or Level 2 (or both). The status attributes l1State and l2State indicate the condition for the Level 1 and Level 2 databases respectively. On entering Level 1 "Waiting State" the IS shall generate the ISPL1DatabaseOverload event, and on entering Level 2 "Waiting State" the IS shall generate the ISPL2DatabaseOverload event.

7.3.19.2 Actions in level 1 waiting state

While in Level 1 "waiting" state

- a) If a Link State PDU cannot be stored, the IS shall ignore it and restart the timer for waitingTime.
- b) The IS shall continue to run the Decision and Forwarding processes as normal.
- c) When the waitingTime timer expires, the IS shall:
 - 1) Generate an ISPL1DatabaseOverload (recovered) event.
 - 2) Clear the LSP Database Overload bit in its own Level 1 LSP with zero LSP number and re-issue it.

- 3) Set the l1State to "On".
- 4) Resume normal operation.

7.3.19.3 Actions in level 2 waiting state

While in Level 2 "waiting" state

- a) If a Link State PDU cannot be stored, the IS shall ignore it and restart the timer for waitingTime seconds.
- b) The IS shall continue to run the Decision and Forwarding processes as normal.
- c) When the waitingTime timer expires, the IS shall:
 - 1) Generate an ISPL2DatabaseOverload (recovered) event.
 - 2) Clear the LSP Database Overload bit in its own Level 2 LSP with zero LSP number and re-issue it.
 - 3) Set the l2State to "On".
 - 4) Resume normal operation.

7.3.20 Use of the link state database

The only portion of the database relevant to the Decision Process is the data portion of the Link State PDUs.

The Update Process additionally uses the fields Sequence Number, Remaining Lifetime, and variable SRMflag. The Remaining Lifetimes in the stored link state PDUs can either be periodically decremented, or converted upon receipt into an internal timestamp, and converted back into a Remaining Lifetime upon transmission.

7.3.20.1 Synchronisation with the decision process

Since the Update process and the Decision process share the link state database, care must be taken that the Update process does not modify the link state database while the Decision process is running.

There are two approaches to this. In one approach, the Decision process signals when it is running. During this time, the Update process queues incoming Link State PDUs, and does not write them into the link state database. If more Link State PDUs arrive than can fit into the queue allotted while the Decision process is running, the Update process drops them and does not acknowledge them.

Another approach is to have two copies of the link state database — one in which the Decision process is computing, and the other in which the Update process initially copies over the first database, and in which all new Link State PDUs are written. Additionally, depending on the hashing scheme, it is likely that a second copy of the address hash table will be required, so that the Update process can do a rehash occasionally for efficiency.

When the Decision process is ready to run again, it locks the new copy of the link state database, leaving the Update process to copy over the information into the first area, and write new updates while the Decision process runs again.

The advantage of the first approach is that it takes less memory. The advantage of the second approach is that Link State PDUs will never need to be dropped.

NOTE 36 If the decision process is implemented according to the specification in C.2, a finer level of parallelism is possible, as described below.

Arrival of a Link State PDU for a system before that system has been put into TENT is permitted. The new Link State PDU is used when that system is eventually put into TENT. Similarly, arrival of a new Link State PDU for a system after that system has been put into PATHS is permitted. That system has already been completely processed. The arrival of the new Link State PDU is noted and the decision process re-executed when the current execution has completed. An in-progress execution of the decision process shall not be abandoned, since this could prevent the decision process from ever completing.

Arrival of a Link State PDU for a system between that system being put on TENT and being transferred to PATHS can be treated as equivalent to one of the previous two cases (for example, by buffering, or taking some corrective action).

7.3.20.2 Use of buffers and link bandwidth

Implementations shall have a buffer management strategy that does not prevent other clients of the buffering service from acquiring buffers due to excessive use by the Update Process. They shall also ensure that the Update Process does not consume all the available bandwidth of links. In particular no type of traffic should experience starvation for longer than its acceptable latency. Acceptable latencies are approximately as follows:

- Hello traffic – Hello timer $\times 0,5$
- Data Traffic – 10 s.

NOTE 37 The first of these requirements can be met by restricting the Update process to the use of a single buffer on each circuit for transmission. This may also cause the second requirement to be met, depending on the processor speed.

7.3.21 Parameters

MaxAge – This is the amount of time that may elapse since the estimated origination of the stored Link State PDU by the source before the LSP is considered expired. The expired LSP can be deleted from the database after a further ZeroAgeLifetime has expired. MaxAge shall be larger than maximumLSPGenerationInterval, so that a system is not purged merely because of lack of events for reporting Link State PDUs.

MaxAge is an architectural constant equal to 20 min.

ZeroAgeLifetime — This is the minimum amount of time for which the header of an expired LSP shall be retained after it has been flooded with zero Remaining Lifetime. A very safe value for this would be $2 \times \text{MaxAge}$. However all that is required is that the header be retained until the zero Remaining Lifetime LSP has been safely propagated to all the neighbours.

ZeroAgeLifetime is an architectural constant with a value of 1 min.

maximumLSPGenerationInterval – This is the maximum amount of time allowed to elapse between generation of Link State PDUs by a source. It shall be less than MaxAge.

Setting this parameter too fast adds overhead to the algorithms (a lot of Link State PDUs). Setting this parameter too slow (and not violating constraints) causes the algorithm to wait a long time to recover in the unlikely event that incorrect Link State information exists somewhere in the domain about the system.

A reasonable setting is 15 min.

minimumLSPGenerationInterval – This is the minimum time interval between generation of Link State PDUs. A source Intermediate system shall wait at least this long before re-generating one of its own Link State PDUs.

Setting this too large causes a delay in reporting new information. Setting this too small allows too much overhead.

A reasonable setting is 30 s.

minimumLSPTransmissionInterval – This is the amount of time an Intermediate system shall wait before further propagating another Link State PDU from the same source system.

Setting this too large causes a delay in propagation of routing information and stabilisation of the routing algorithm. Setting this too small allows the possibility that the routing algorithm, under low probability circumstances, will use too many resources (CPU and bandwidth).

Setting minimumLSPTransmissionInterval greater than minimumLSPGenerationInterval makes no sense, because the source would be allowed to generate LSPs more quickly than they would be allowed to be broadcast. Setting minimumLSPTransmissionInterval smaller than minimumLSPGenerationInterval is desirable to recover from lost LSPs.

A reasonable value is 5 s.

CompleteSNPInterval – This is the amount of time between periodic transmissions of a complete set of Sequence Number PDUs by the Designated Intermediate system on a broadcast link. Setting this too low slows down the convergence of the routing algorithm when Link State PDUs are lost due to the datagram environment of the Data Link layer on the broadcast link.

Setting this too high results in extra control traffic overhead.

A reasonable value is 10 s.

7.4 Forwarding process

The Forwarding process is responsible both for transmitting NPDUs originated by this system, and for forwarding NPDUs originated by other systems

7.4.1 Input and output

INPUT

- NPDUs from the ISO 8473 protocol machine
- PDUs from Update Process
- PDUs from Receive Process
- Forwarding Databases (Level 1 and 2) — one for each routing metric

OUTPUT

- PDUs to Data Link Layer

7.4.2 Routing metric selection

The Forwarding process selects a forwarding database for each NPDu to be relayed based on:

- the level at which the forwarding is to occur: level 1 or level 2; and
- a mapping of the ISO 8473 QoS Maintenance field onto one of the Intermediate system's supported routing metrics.

The former selection is made by examining the Destination Address field of the NPDu.

The latter selection is made as follows:

- a) If the QoS Maintenance field is not present in the NPDu, then the IS shall select the forwarding database calculated for the *default metric*.
- b) If the QoS Maintenance field is present, the IS shall examine bits 7 and 8 of the parameter value octet. If these two bits specify any combination other than "1 1" (meaning globally unique QoS), then the IS shall select the forwarding database calculated for the *default metric*, otherwise
- c) The IS shall select a forwarding database by mapping the values of bits 3, 2 and 1 of the parameter value as shown below in table 1 and shall proceed as follows:
 - 1) If the IS does not support the selected routing metric, the IS shall forward based upon the *default metric*;
 - 2) If the forwarding database for one of the optional routing metrics is selected and the database either does not contain an entry for the Destination Address in the NPDu being relayed, or contains an entry indicating that the destination is unreachable using that metric, then the IS shall attempt to forward based upon the *default metric*;

- 3) Otherwise, forward based on the selected optional metric.

Table 1 - QoS Maintenance bits to routing metric mappings

bit 1	bit 2	bit 3	Selected Routing Metric
0	0	0	expense metric
1	0	0	default metric
0	1	1	default metric
0	1	0	expense metric
0	0	1	delay metric
1	1	0	error metric
1	0	1	delay metric
1	1	1	error metric

7.4.3 Forwarding decision

7.4.3.1 Basic operation

Let DEST = the Network Layer destination address of the PDU to be forwarded, or the next entry in the source routing field, if present. It consists of sub-fields Area Address, ID, and SEL.

NOTE 38 The SEL field in the destination address is not examined by Intermediate Systems. It is used by End Systems to select the proper Transport entity to which to deliver NSDUs.

This system's (the one examining this PDU for proper forwarding decision) address consists of sub-fields area address and ID.

- a) If the local system type is a level 1 Intermediate system, or the local system type is a level 2 Intermediate system and AttachedFlag_k = False, then:
 - 1) If the Area Address in the PDU to be forwarded matches any one of the area addresses of this IS, then consult the level 1 forwarding database to determine the adjacency which is the next hop on the path to the NPDu's destination. Forward the NPDu on this adjacency.
 - 2) Otherwise, consult the level 1 forwarding database to determine the adjacency which is the next hop on the path to the nearest level 2 IS in the area, and forward the NPDu on this adjacency.
- b) If the local system type is Level 2, and AttachedFlag_k = "True" then:
 - 1) If the Area Address in the PDU to be forwarded matches any one of the area addresses of this IS, then consult the level 1 forwarding database to determine the adjacency which is the next hop on the path to the NPDu's destination. Forward the NPDu on this adjacency.

- 2) Otherwise, consult the level 2 forwarding database to determine the adjacency which is the next hop on the path to the destination area, and forward the NPDU on this adjacency.

7.4.3.2 Encapsulation for partition repair

If this Intermediate system is the Partition Designated Level 2 IS for this partition, and the PDU is being forwarded onto the special adjacency to a Partition Designated Level 2 Intermediate system in a different partition of this area, encapsulate the complete PDU as the data field of a data NPDU (i.e., with an additional layer of header), making this system the Source address and the other Partition Designated Level 2 Intermediate system (obtained from the identifier attribute of the Virtual Adjacency managed object) the Destination Address field in the outer PDU header. Set the header of the outer PDU to indicate the forwarding via the default routing metric [see 7.2.10.4, item d)].

Then forward the encapsulated PDU as described in 7.4.3.3 below.

7.4.3.3 The procedure forward

Choose either the level 1 or level 2 forwarding database, depending on the destination Network Address in the NPDU. From that database, select the adjacency for the next hop to that destination. If forwarding at level 1 for a destination which is not in the area, choose the adjacency for the nearest Level 2 IS computed as described in 7.2.9.1.

If there are multiple possible adjacencies, as a result of multiple minimum cost paths, then one of those adjacencies shall be chosen. An implementation may choose the adjacency at random, or may use the possible adjacencies in "round robin" fashion.

If there is no entry in the selected forwarding database for the destination Network Address, and the NPDU originated from the local Transport entity and the system has one or more Intermediate System adjacencies, then one of those is chosen at random (or in "round robin" fashion).¹⁾

NOTES

39 Since the local adjacency database is pre-loaded into the decision process, there will always be an entry in the forwarding database for destinations to which an adjacency exists.

40 The PDU to be forwarded may require fragmentation, depending on which circuit it is to be forwarded over.

7.4.3.4 Generating redirect PDUs

In addition to forwarding an NPDU, the IS shall inform the local ISO 9542 protocol machine to generate a *Redirect PDU* if the PDU is being forwarded onto the same circuit from which it came, and if the source SNPA address of the NPDU indicates that the NPDU was received from an End System.

7.4.4 Receive process

The Receive process is passed information from any of the following sources.

- received PDUs with the NLPID of IntraDomain-RoutingPD,
- configuration information from the ISO 9542 protocol machine,
- ISO 8473 data PDUs handed to the routing function by the ISO 8473 protocol machine.

When an area is partitioned, a level 2 path is used as a level 1 link to repair the partitioned area. When this occurs, all PDUs (between the neighbours which must utilise a multi-hop path for communication) shall be encapsulated in a data NPDU, addressed to the IntraDomainRoutingSelector. Control traffic (LSPs, Sequence Numbers PDUs) shall also be encapsulated, as well as data NPDUs that are to be passed between the "neighbours".

NOTE 41 It is not necessary to transmit encapsulated IIH PDUs over a virtual link, since virtual adjacencies are established and monitored by the operation of the Decision Process and not the Subnetwork Dependent functions.

7.4.4.1 Basic operation

The Receive Process shall perform the following functions on each received PDU:

- If it is a Link State PDU, pass it to the Update Process
- If it is a Sequence Numbers PDU, pass it to the Update Process
- If it is an IIH PDU, pass it to the appropriate Subnetwork Dependent Function
- If it is a data NPDU or Error Report for another destination, pass it to the Forwarding Process
- Otherwise, ignore the PDU

7.4.4.2 Decapsulation

If an ISO 8473 data NPDU, addressed to *this system*, is received, and the SEL field of the address is equal to IntraDomainRoutingSelector, then the IS shall

- decapsulate the NPDU (remove the outer NPDU header).
- If the decapsulated PDU is a data NPDU, move the "congestion" indications to the decapsulated NPDU, and pass it to the ISO 8473 protocol machine.
- Otherwise, if the decapsulated PDU is not an ISO 8473 PDU, perform the following steps on the decapsulated PDU:

¹⁾ This is done so that a system in the overloaded state will still be able to originate or forward NPDUs. If a system with a partial routing information base were prohibited from attempting to forward to an unknown destination, system management would be unable to either communicate with this system, or route through it, for the purpose of diagnosing and/or correcting the underlying fault.

- If it is a link state PDU or Sequence Numbers PDU, pass it to the Update process;
- Otherwise, ignore the PDU.

7.5 Routeing constants and parameters

The architectural constants are described in table 2.

The routing parameters settable by System Management are listed for each managed object in clause 11.

8 Subnetwork dependent functions

The *Subnetwork Dependent Functions* mask the characteristics of the different kinds of Subnetworks from the *Subnetwork Independent Routeing Functions*. The only two types of circuits the Subnetwork Independent Functions recognise are *broadcast* and *general topology*.

The Subnetwork Dependent Functions include:

- The use of the ISO 8473 *Subnetwork Dependent Convergence Functions (SND CF)* so that this protocol may transmit and receive PDUs over the same subnetwork types, using the same techniques, as does ISO 8473.
- Co-ordination with the operation of the ES-IS protocol (ISO 9542) in order to determine the Network layer addresses (and on Broadcast subnetworks, the subnetwork point of attachment address) and identities (End System

or Intermediate System) of all adjacent neighbours. This information is held in the *Adjacency* data base. It is used to construct Link State PDUs.

- The exchange of IIH PDUs. While it is possible for an Intermediate System to identify that it has an Intermediate System neighbour by the receipt of an ISO 9542 ISH PDU, there is no provision within ISO 9542 to indicate whether the neighbour is a Level 1 or a Level 2 Intermediate System. Specific PDUs (*LAN Level 1*, *LAN Level 2* and *Point-to-point IIH PDUs*) are defined to convey this information.

8.1 Multi-destination circuits on ISs at a domain boundary

Routeing information (e.g. Link State PDUs) is not exchanged across a routing domain boundary. All routeing information relating to a circuit connected to another routing domain is therefore entered via the Reachable Address managed objects. This information is disseminated to the rest of the routing domain via Link State PDUs as described in 7.3.3.2. This has the effect of causing NPDUs destined for NSAPs which are included in the addressPrefix of the Reachable Addresses to be relayed to that Intermediate System at the domain boundary. On receipt of such an NPDU the Intermediate system shall forward it onto the appropriate circuit, based on its own Link State information. However in the case of multi-destination subnetworks (such as an ISO 8208 subnetwork using Dynamic Assignment, a broadcast subnetwork, or a connectionless subnetwork) it is necessary to ascertain additional subnetwork dependent addressing information in order to forward the NPDU to a suitable SNPA. (This may be the target End system or an Intermediate system within the other domain.)

Table 2 - Routeing architectural constants

Name	Value	Description
MaxLinkMetric	63	Maximum value of a routeing metric assignable to a circuit
MaxPathMetric	1023	Maximum total metric value for a complete path
ISO-SAP	FE	The SAP for ISO Network Layer on ISO 8802-2 LANs. This SAP value is used for transmission and reception of all PDUs of this protocol.
IntradomainRouteingPD	10000011	The Network Layer Protocol Identifier assigned to this protocol, as recorded in ISO/TR 9577
IntradomainRouteingSelector	0	The NSAP selector for the Intermediate System Network entity
SequenceModulus	2^{32}	Size of the sequence number space used by the Update Process
ReceiveLSPBufferSize	1492	The size of LSP which all Intermediate systems must be capable of receiving.
MaxAge	1200	Number of seconds before LSP considered expired.
ZeroAgeLifetime	60	Number of seconds that an LSP with zero Remaining Lifetime shall be retained after propagating a purge.
ISISHoldingMultiplier	10	The number by which to multiply iSiSHelloTimer to obtain Holding Timer for Level 1 and Level 2 IIH PDUs.
Jitter	25%	The percentage of jitter which is applied to the generation of periodic PDUs. See 10.1 for further information on generating jitter on timers.

In general the SNPA address to which an NPDU is to be forwarded can be derived from the destination NSAP of the NPDU. It may be possible to perform some algorithmic manipulation of the NSAP address in order to derive the SNPA address. However there may be some NSAPs where this is not possible. In these cases it is necessary to have pre-configured information relating an address prefix to a particular SNPA address.

This is achieved by additional information contained in the reachable address managed object. The mappingType attribute specifies the means by which next hop subnetwork addressing information can be derived for NPDUs forwarded based upon a given address prefix. The mappingType attribute may be specified as:

explicit — The SNPA address or set of SNPA addresses is manually pre-configured as an attribute of the reachable address managed object.

extractIDI — The SNPA is embedded in the IDI of the destination NSAP address according to the format and encoding rules of ISO 8348/Add.2. This SNPA extraction algorithm can be used in conjunction with destination addresses from the X.121, F.69, E.163, and E.164 addressing subdomains.

extractDSP — All or a suffix of the SNPA is embedded in the DSP of the destination address. This SNPA extraction algorithm requires manual pre-configuration of sNPAMask and sNPAPrefix attributes of the reachable address managed object. The sNPAMask attribute is a bit mask with 1s indicating the location of the SNPA (suffix) within the destination NSAP DSP. The part of the SNPA extracted from the NSAP is appended to the sNPAPrefix to form the next hop subnetwork addressing information.

An example of a set of Reachable Addresses is shown in table 3.

The table is interpreted as follows:

- a) For the ISO DCC prefix 39 123, use the SNPA address X.

- b) For the X.121 IDI address prefix 37 aaaaaa, do not use aaaaaa, but use B instead.
- c) For all IDPs based on SNPAs with DNIC D (i.e. with address prefix 37 D), use the address Y (which would probably be a gateway to a subnetwork with DNIC D).
- d) For any other X.121 IDI (i.e. address prefix 37) – use the SNPA whose address is used as the IDI.
- e) For the ISO ICD prefix 47 0005 C0 use the SNPA address formed by concatenating Z with next 6 octets of the DSP following the 47 0005 C0 prefix.
- f) Anything else (“*” in table 3) – use one of the SNPA addresses R, S or T. These would typically be the SNPA addresses of Level 2 Intermediate Systems through which any other addresses could potentially be reached.

8.2 Point-to-point subnetworks

This clause describes the identification of neighbours on both point-to-point links and Static circuits.

The IS shall operate the ISO 9542 protocol, shall be able to receive ISO 9542 ISH PDUs from other ISs, and shall store the information so obtained in the adjacency database.

8.2.1 Receipt of ESH PDUs — database of end systems

An IS shall enter an End system into the adjacency database when an ESH PDU is received on a circuit. If an ESH PDU is received on the same circuit, but with a different NSAP address, the new address shall be added to the adjacency, with a separate timer. A single ESH PDU may contain more than one NSAP address. When a new data link address or NSAP address is added to the adjacency database, the IS shall generate an adjacencyStateChange (Up) event on that adjacency.

The IS shall set a timer for the value of Holding Time in the received ESH PDU. If another ESH PDU is not received from the ES before that timer expires, the ES shall be purged from the database, provided that the Subnetwork Independent Functions associated with initialising the adjacency have been completed. Otherwise the IS shall clear the adjacency as soon as those functions are completed.

Table 3 - Example of reachable address information

Address Prefix	Mapping Type	SNPA Address
39 123	explicit	X
37 aaaaaa	explicit	B
37 D	explicit	Y
37	extractIDI	Extract X.121 SNPA address from NSAP IDI
47 0005 C0	extractDSP	sNPAPrefix=Z sNPAMask=00000000FFFFFFFFFFFF
*	explicit	R, S, T