Project Number: IST-1999-11387
Project Title: Layers Interworking in Optical Networks (LION)
Deliverable Type: (P/R/L/I)* P

CEC Deliverable Number: IST-1999-11387/CSELT/LION_D6
Contractual Date of Delivery to the CEC: Month 9
Actual Date of Delivery to the CEC: 30/10/00
Title of Deliverable: Network Scenarios and Requirements
Workpackage contributing to the Deliverable: WP1
Nature of the Deliverable: (P/R/S/T/O)** R
Author(s): LION

Abstract:

The goal of this deliverable is to provide a preliminary description of scenarios (in terms of services and architectures) and requirements for Next Generation Networks particularly reffering to IP-based over Optical Transport Networks (OTN).

The deliverable starts with a description of the main business drivers moving the evolution of current transport networks. Then an overview of current and future application and transport services introduces a preliminary description of the network requirements envisaged to support such services. The document presents also the current status of the key enabling technologies and protocols which are likely to be used in the Next Generation Networks.. The deliverable contains also a brief survey of currently deployed transport networks. This allows to set the starting point of the project roadmap. Innovative functionality (e.g. interworking) are preliminary described to be studied and experimented in the test-bed during the Project life

Keyword list: network requirements, network scenarios, layers interworking, interconnection Optical Transport Netwoks, DWDM, IP, MPLS, G-MPLS, DiffServ, IntServ, SDH, network architecture, network survivability, optical internetworking test-bed, ASON.

* Type: P-public, R-restricted, L-limited, I-internal

** Nature: P-prototype, R-report, S-specification, T-tool, O-other

Status and Version:	Version Final	
Date of issue:	27.10.2000	
Distribution:	Project internal – call for comments	
Authors:	**Name:**	**Partner:**
	Artur Lason	AGH
	Jacek Roman	AGH
	Jacek Rząsa	AGH
	Jan Derkacz	AGH
	Janusz Gozdecki	AGH
	Mateusz Kosciuszko	AGH
	Piotr Pacyna	AGH
	Robert Chodorek	AGH
	Rafal Stankiewicz	AGH
	Antonio Manzalini	CSELT
	Giuseppe Marone	CSELT
	Chris Develder	IMEC
	Didier Colle	IMEC
	Mario Pickavet	IMEC
	Piet Demeester	IMEC
	Pim Vanheuven	IMEC
	Sophie Demaesschalck	IMEC
	Steven Van den Berghe	IMEC
	Katsuhiro Shimano	NTT
	Naohide Nagatsu	NTT
	Giorgos Chatzilias	NTUA
	Lampros Raptis	NTUA
	Panagiotis Papagiannis	NTUA
	Yannis Manolessos	NTUA
	Stefano Brunazzi	SICN
	Andreas Gladisch	T-NOVA
	Frank Tetzlaff	T-NOVA
	Fritz-Joachim Westphal	T-NOVA
	Jochen Knapp	T-NOVA
	Josef Röse	T-NOVA
	Monika Jäger	T-NOVA
	Ralph-Peter Braun	T-NOVA
	Regine Liebenstein	T-NOVA
	Sabine Szuppa	T-NOVA
	Adam Matzke	TPSA
	Janusz Maliszewski	TPSA
	Albert Rafel	UPC
	Carlos Veciana Nogues	UPC
	Davide Careglio	UPC
	Jaume Comellas	UPC
	Josep Prat	UPC
	Josep Sole-Pareta	UPC
	Julio Moyano	UPC
	Salvatore Spadaro	UPC
Checked by:		

1 Introduction

The goal of this Deliverable D6 is to define preliminary network scenarios and requirements for the next generation transport network based on the OTN.

The Deliverable is based on the results of WP1 activities reported in the two milestones: WP1M1 "Network Requirements" and WP1M2 "Network Scenarios and Guidelines for the Testbed Configuration".

Particularly D6 starts from a description of the main business drivers moving the transport network evolution. Then an overview of application and transport services for the next generation networks introduces emerging network requirements. Innovative functionality (e.g. interworking) are preliminary described to be studied and experimented in the test-bed during the Project life. The state of art of enabling technologies and a general survey of currently deployed transport networks allows to set the starting point of the project roadmap.

As a main conclusion a preliminary roadmap is given. As data (mainly Internet) is the fastest growing segment of network traffic, transport network models are likely to evolve to data-centric solutions, primarly ASON (overlay vision) and then even G-MPLS based (peer-to-peer vision). As the OTN can provide the large amount of raw bandwidth supporting the increasing data traffic, in the short term a client-independent OTN (overlay vision) is likely to be the missing link between legacy (e.g. mainly SDH) and data centric(e.g. mainly IP) networks.

1.1 Reference Material

1.1.1 Reference Documents

[ARCH]	E.C.Rosen, A.Viswanathan, R.Callon, "Multiprotocol Label Switching Architecture", draft-ietf-mpls-arch-06.txt, Internet Draft, August 1999.
[Awduche]	Awduche/Rekhter et al., "Multi-Protocol Lambda Switching: Combining MPLS Traffic Engineering Control With Optical Crossconnects", draft-awduche-mpls-te-optical-02.txt
[BGPLAB]	Y. Rekhter, E. Rosen, "Carrying Label Information in BGP-4", draft-ietf-mpls-bgp4-mpls-04, work in progress.
[BLAC]	Uyless Black, "TCP/IP and related Protocols", McGraw-Hill, 1998 (3rd Edition).
[BOSS]	M. Bossert: Kanalcodierung. Teubner Verlag, Stuttgart, 1998
[CONC-TE]	K. Takashima, K. Nakamichi, and T. Soumiya, "Concept of IP Traffic Engineering", <draft-takashima-te-concept-00.txt>, work in progress, October 1999.
[CRLDP]	B.Jamoussi, "Constraint-Based LSP Setup using LDP", draft-ietf-mpls-cr-ldp-03.txt, Internet Draft, September 1999.
[DIFF-NEW]	D.Grossman, "New Terminology for DiffServ", draft-ietf-diffserv-new-terms-02.txt, Internet Draft, November 1999.
[DPS]	I.Stoica et al, *"Per Hop Behaviors Based on Dynamic Packet States"*, <draft-stoica-diffserv-dps-00>, work in progress, February 1999
[DPT-1]	Cisco Systems, "Cisco Optical Internetworking". http://www.cisco.com/warp/public/779/servpro/solutions/opt/oi_brochure.html
[DPT-2]	Cisco Systems, "Dynamic Packet Transport Technology and Applications Overview". http://www.cisco.com/warp/public/cc/cisco/mkt/servprod/opt/dpt/dpta_wp.htm

[DPT-3] Internet Draft (Work in progress), "The Cisco SRP MAC Layer Protocol <draft-tsiang-srp-01.txt>", D. Tsiang, G. Suwala, March 2000 (This draft expires on 1 September 2000) http://search.ietf.org/internet-drafts/draft-tsiang-srp-01.txt

[DPT-4] Executive Summary, "Metro Optical Networks: Metro DWDM and the New Public Network", Pioneer Consulting

[DPT-5] Cisco DPT Products http://www.cisco.com/warp/public/cc/cisco/mkt/servprod/opt/dpt/

[DPT-6] RingStar8000 Product Data Sheet http://www.penta-com.com/ringstar8000.htm

[DPT-7] Internet Draft (Work in progress), "Definitions of Managed Objects for Spatial Reuse Protocol (SRP) <draft-jedrysiak-srp-mib-00.txt>2", Stan Jedrysiak, March 2000 (This draft expires in September 2000) http://search.ietf.org/internet-drafts/draft-jedrysiak-srp-mib-00.txt

[EN 301164] EN 301 164, "SDH leased lines; Connection characteristics", 1999

[EN 301165] EN 301 165, "SDH leased lines; Network and terminal interface presentation", 1999

[ENCAPS] D. Farinacci et al., "MPLS Label Stack Encoding", draft-ietf-mpls-label-encaps-07, work in progress.

[ETS 300 417-4-1] ETS 300 417-4-1, "Generic requirements of transport functionality of equipment; Part 4-1: Synchronous Digital Hierarchy (SDH) path layer functions" ,1999

[ETS 300 462-2-1] ETS 300 462-2-1, "Generic requirements for synchronization networks; Part 2-1: Synchronization network architecture", 1999

[ETS 300 417-3-1] ETS 300 417-3-1, "Generic requirements of transport functionality of equipment; Part 3-1: Synchronous Transport Module-N (STM-N) regenerator and multiplex section layer functions", 1999

[FR] A. Conta, P. Dool an, and A. Malis, "Use of Label Switching on Frame Relay Networks Specification", draft-ietf-mpls-fr-03, work in progress.

[G.ason] ITU Rec. G.ason

[Ghani 2000] N. Ghani, "Lambda-Labeling: a Framework for IP-over-WDM using MPLS", Optical Networks (Spie), April 2000.

[G.852.16] ITU-T Draft Rec. G.852.16, "Enterprise viewpoint for pre-provisioned route discovery".

[Hessing, 1999] Steven Hessing, "Effectively Running IP over SDH and WDM", IP over WDM conference, Geneva, July 8th-9th, 1999

[IETF-1] Internet Draftraft-ashwood-generalized-mpls-signaling-00.txt June 2000

[IMEC1] Presentation of IMEC during WP2 meeting in Darmstadt on 13th of July, 2000. LOWS-address: ftp://lionp@ektor.telecom.ntua.gr/usr/WWW/htdocs/lion/private/management/WPG1/WP2/MS_WP2_IMEC_000613.ppt

[ITU-G114] ITU-T Recommendation G.114: Transmission systems and media - One way transmission time, February 1996.

[ITU-G691] ITU-T G.691, "Optical interfaces for single channel STM-64, STM-256 systems and other SDH systems with optical amplifiers", draft Recommendation, April 2000

[ITU-G703] ITU-T G.703, "Physical/electrical characteristics of hierarchical digital interfaces", 1998

[ITU-G707] ITU-T G.707, "Network Interface for the synchronous digital hierarchy (SDH)", 1996

[ITU-G709] ITU-T Recommendation G.709 Draft: NETWORK NODE INTERFACE FOR THE OPTICAL TRANSPORT NETWORK (OTN)

[ITU-G781] G.781, "Synchronization layer functions", 1999

[ITU-G782] G.782, "Types and general characteristics of synchronous digital hierarchy (SDH) equipment", 1994

[ITU-G783] G.783, "Characteristics of synchronous digital hierarchy (SDH) equipment functional blocks", 1997
[ITU-G803] ITU-T Recommendation G.803
[ITU-G825] ITU-T G.825, "The control of jitter and wander within digital networks which are based on the synchronous digital hierarchy (SDH)", 1993
[ITU-G826] ITU-T G.826, "Error performance parameters and objectives for international, constant bit rate digital paths at or above the primary rate", 1999
[ITU-G827] ITU-T G.827, "Availability parameters and objectives for path elements of international constant bit-rate digital paths at or above the primary rate", 1996
[ITU-G828] ITU-T G.828, "Error performance parameters and objectives for international constant bit rate synchronous digital paths", draft new Recommendation, October 1999
[ITU-G872] ITU-T Recommendation G.872
[ITU-G957] ITU-T G.957, "Optical interfaces for equipments and systems relating to the synchronous digital hierarchy", 1999
[ITU-H323] ITU-T Recommendation H.323: Packet-based multimedia communications systems, February 1998.
[ITU-I 432] I.432, "B-ISDN User-Network Interface – Physical layer specification", 1996
[ITU-M3010] M.3010, "Principles for a Telecommunications management network", 1996
[ITU-T G841] ITU-T Recommendation G841, "Types and characteristics of SDH network protection architectures", October 1998
[ITU-T Gason] ITU-T, Recommendation draft G.ason.
[ITU-T M3000] ITU-T, Recommendation M-3000.
[Jamousi] Jamousi et al. "Constraint-Based LSP Setup using LDP", draft-ietf-mpls-crldp-03.txt
[KURT] C. Kurtzke, "Kapazitätsgrenzen digitaler optischer Übertragungssysteme", TU Berlin, Dissertation D83, 1994
[LANE] "LANE v2.0 LUNI Interface", (af-lane-0084.000) July, 1997.
[LDP] L.Andersson, P.Doolan, N.Feldman, A.Fredette, B.Thomas., "LDP Specification", draft-ietf-mpls-ldp-06.txt, Internet Draft, October 1999.
[LION-WP1-T1] "Network requirements – Analysis and Services", LION/WPG1/WP1/T1, July 2000, working document, DT-LION-WP1Task1_01j.doc
[LOADCONTR] L.Westberg, Z.R.Turanyi, *"Load Control of Real-Time Traffic"*, <draft-westberg-loadcntr-03>, work in progress, April 2000
[LOOP1] Y. Ohba "Issues on loop prevention in MPLS networks", IEEE communications magazine, Januari 2000.
[LOOP2] Y. Ohba et al., "MPLS Loop Prevention Mechanism", draft-ohba-mpls-loop-prevention-02, work in progress.
[Manolessos] Yannis Manolessos, MSSP.xls
[MECN] K. K. Ramakrishnan et al., "A Proposal to Incorporate ECN in MPLS", draft-ietf-mpls-ecn-00, work in progress.
[Mephisto] Mephisto ACTS project D19 deliverable.
[MICMP] R.Bonica et al., "ICMP Extensions for MultiProtocol Label Switching", draft-ietf-mpls-icmp-01, work in progress.
[MPLS-DS] F. Le Faucheur et al., "MPLS Support of Differentiated Services", draft-ietf-mpls-diff-ext-03, work in progress.
[MPLS-FRAME] R.Callon, P.Doolan, N.Feldman, A.Fredette, G.Swallow, A.Viswanathan, "A Framework for Multiprotocol Label Switching", draft-ietf-mpls-framework-05.txt, Internet Draft, September 1999.
[MPLS-QoS] J. Ash et al., "QoS Resource Management in MPLS-based Networks", <draft-ash-qos-routing-00.txt>, work in progress, 1999
[MPLS-TE] D. Awduche, "MPLS and Traffic Engineering in IP Networks.", IEEE Communications Mag., December 1999.
[MPOA] "Multi-Protocol Over ATM v1.0", (af-mpoa-0087.000) July, 1997.
[MYTH] G. Armitage, "MPLS: The magic behind the myths", IEEE communications magazine, January 2000.

[NR-AoS] "Network Requirements – Analysis of Services", July 2000, LION/WPG1/WP1/T1 – working document: DT-LION-WP1Task1_01j.doc.
[NR-P&T] "Network Requirements – Products and Technologies", June 2000, LION/WPG1/WP1/T2 – working document: CM1T2-IMEC_v201.doc.
[Ocakoglu, 1999] Gzim Ocakoglu, "The Deployment of an IP over WDM Network: a Pan-European Perspective", IIR WDM congress, Cannes, June 28th – July 1st, 1999
[OTN-DELIV] Eurescom P918, D1, "IP over WDM, Transport and Routing", October 1999.
[OTN-PUBLI 1] L.Thylén et al., "Switching technologies for future guided wave optical networks: potentials and limitations of photonics and electronics", IEEE Comm. Mag., pp. 106-113, February 1996.
[OTN-PUBLI 2] A. Ware (Agilent) "New photonic-switching technology for all-optical networks" – Lightwave, March 2000
[OTN-PUBLI 3] Communication Magazine (March 2000).
[OTN-PUBLI 4] "Switching in IP Networks", Morgan Kaufmann Publishers, Inc., 1998.
[OTN-PUBLI 5] Report from Communications Industry Researchers, Inc. (CIR) on "Wave Division Multiplexing, Photonic Switching and the Coming of All Optical Networks 1999-2000, Volume1"
[OTN-STD 1] IETF draft "IP over Optical Networks - A Framework"
[OTN-STD 2] IETF draft "An architecture for MPLS control plane for Switched Optical Networks"
[OTN-STD 3] IETF draft "Multi-Protocol Lambda Switching: combining MPLS traffic engineering control with Optical Cross-connects"
[OTN-STD 4] IETF RFC 2702 "Requirements for Traffic Engineering Over MPLS"
[OTN-STD 5] ITU-T G.ason "Architecture for the Automatic Switched Optical Network"
[OTN-STD 6] ITU-T X.ipos "IP over SDH using LAPS"
[OTN-STD 7] ITU-T X.eos "Ethernet Frame over SDH/WDM"
[OTN-STD 8] ANSI T1X1 draft 99-269 (Nortel) "Generic Format for Carrying Ethernet MAC Frames over SONET"
[OTN-STD 9] "ODSI Functional Specifications"
[OTN-WEB 1] www.cisco.com
[OTN-WEB 10] www.lucent.com
[OTN-WEB 11] www.agilent.com
[OTN-WEB 12] www.tellabs.com
[OTN-WEB 13] www.telcordia.com
[OTN-WEB 2] www.marconicomms.com
[OTN-WEB 3] www.sycamorenet.com
[OTN-WEB 4] www.ciena.com
[OTN-WEB 5] www.alcatel.com
[OTN-WEB 6] www.xros.com
[OTN-WEB 7] www.siemens.com
[OTN-WEB 8] www.nortelnetworks.com
[OTN-WEB 9] www.juniper.com
[OTN-WHITE 1] C. Webb, T. Krause (Alcatel) "Drivers and Enabling Technologies for Wavelength Based Services in the All Optical Network".
[OTN-WHITE 10] Monterey White Paper: "Scaling Optical Data Network with Wavelength Routing"
[OTN-WHITE 11] N. De Vito (Tellium) "Delivering Optical Layer Services with Optical Switching".
[OTN-WHITE 12] C.A. Brackett (Tellium) "Cost Effective Optical Networks: the role of Optical Cross-Connects".
[OTN-WHITE 2] Ciena White Paper: "The new economics of Optical Core Networks"
[OTN-WHITE 3] "Cisco's Packet over SONET (POS) technology support: mission accomplished".
[OTN-WHITE 4] "Cisco Optical Internetworking."
[OTN-WHITE 5] Dynarc White Paper: "Distributed Switching and Routing"

[OTN-WHITE 6] R.A. Goudreault et al. (Lucent) "Impact of an Integrated Architecture for Bandwidth Management in a Broadband network Infrastructure"
[OTN-WHITE 7] J. Anderson et al. (Lucent) "Protocols and architectures for IP optical networking"
[OTN-WHITE 8] P. Bonenfant (Lucent) "WaveWrapper Technology, A major leap in optical transport networking".
[OTN-WHITE 9] B.T. Doshi (Lucent) "A Simple Data Link Protocol for high speed packet networks"
[PoS-1] James Manchester, et altri, "IP over SONET", IEEE Communications Magazine, May 1998.
[PoS-10] http:/www.lucent.com
[PoS-2] Kevin Thompson, Gregory J. Miller and Rick Wilder, "Wide-Area Internet Traffic Patterns and Characteristics", IEEE Network, November/December 1997.
[PoS-3] W. Simpson, "PPP in HDLC-like Framing", RFC 1662, July 1994, ftp://ftp.upc.es/pub/doc/rfc/16xx/1662
[PoS-4] W. Simpson, "PPP over SONET/SDH", RFC 2615, June 1999, ftp://ftp.upc.es/pub/doc/rfc/26xx/2615
[PoS-5] http://www.lucent.com/micro/tic/docs/sdl.pdf
[PoS-6] Integration of IP over Optical Networks: Networking and Management, Eurescom project P918-GI, Deliverable 1 (IP over WDM, Transport and Routing). http://www.eurescom.de
[PoS-7] W. Stallings, "ISDN and Broadband ISDN with Frame Relay and ATM". Ed. Prentice Hall, 1998 (4th. Edition).
[PoS-8] R. Ramaswami, K.Sivarajan "Optical Networks, A practical Perspective", Morgan Kaufmann Publishers, 1998.
[PoS-9] http:/www.cisco.com
[RFC-1349] P. Almquist., "*Type of Service in the Internet Protocol Suite*", RFC 1349, July 1992
[RFC-1633] R. Braden, D. Clark, S. Shenker, R. Braden, L. Zhang, S. Berson, S. Herzog, S. Jamin, "*Integrated Services in the Internet Architecture: an Overview*", RFC 1633, June 1994.
[RFC-2205] R. Braden et al., "*Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification*", RFC 2205, September 1997
[RFC-2206] F. Baker, J. Krawczyk, A. Sastry, "*RSVP Management Information Base using SMIv2*", RFC 2206, September 1997
[RFC-2210] J. Wroclawski, "*The Use of RSVP with IETF Integrated Services*", RFC 2210, September 1997
[RFC-2211] J. Wroclawski, "*Specification of the Controlled-Load Network Element Service*", RFC 2211, September 1997
[RFC-2212] S. Shenker, C. Partridge, R. Guerin, "*Specification of Guaranteed Quality of Service*", RFC 2212, September 1997
[RFC-2213] F. Baker, J. Krawczyk, A. Sastry, *"Integrated Services Management Information Base using SMIv2"*, RFC 2213, September 1997
[RFC-2214] F. Baker, J. Krawczyk, A. Sastry, *"Integrated Services Management Information Base, Guaranteed Service Extensions using SMIv2"*, RFC 2214, September 1997
[RFC-2215] S. Shenker, J. Wroclawski, "*General Characterization Parameters for Integrated Service Network Elements*", RFC 2215, September 1997
[RFC-2216] S. Shenker, J. Wroclawski, *"Network Element Service Specification Template"*, RFC 2216, September 1997
[RFC-2430] T.Li, Y.Rekhter, "A Provider Architecture for Differentiated Services and Traffic Engineering (PASTE)", Informational, October 1998
[RFC-2474] K. Nichols et al., "*Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers*", RFC 2474, December 1998
[RFC-2475] S. Blake et al., "*An Architecture for Differentiated Services*", RFC 2475, December 1998

[RFC-2597] J. Heinanen et al., "*Assured Forwarding PHB Group*", RFC 2597, June 1999

[RFC-2598] V. Jacobson, K. Nichols, K. Poduri, "*An Expedited Forwarding PHB*", RFC 2598, June 1999

[RFC-2638] K. Nichols, V. Jacobson, L. Zhang, "*A Two-bit Differentiated Services Architecture for the Internet*", RFC 2638, July 1999

[RFC-791] J. Postel., "*Internet Protocol*", RFC 791, September 1981

[RMMIB] A. Bierman, "*Remote Monitoring MIB Extensions for Differentiated Services Enabled Networks*", <draft-bierman-dsmon-mib-01>, work in progress, June 1999

[RSVPT] D. Awduche et al., "Extensions to RSVP for LSP Tunnels", draft-ietf-mpls-rsvp-lsp-tunnel-04, work in progress.

[STAL] William Stallings, "Data Computer Communications", Prentice Hall, 1999, (5th Edition).

[Struyve et al, 2000] Kris Struyve, Nico Wauters, Pedro Falcao, Peter Arijs, Didier Colle, Piet Demeester, Paul Lagasse, "Application, Design and Evolution of WDM in GTS's Pan-European Transport Network", IEEE Communications Magazine, page 114, Vol 38, No. 3, March 2000

[Struyve, 2000] Kris Struyve, "Survivability in Optical Transport Networks: Standardization and Requirements from a Pan-European Carriers' Carrier Perspective", DRCN2000 workshop, Munich, April 9th-12th.

[SWITCH] B. Davie, P. Doolan, and Y. Rekhter, "Switching in IP Networks", Morgan Kaufmann, 1998.

[TECH-SOFT] "The Lucent Technologies Softswitch – Realising the Promise of Convergence", Bell Lab Technical Journal, April-June 1999.

[TETZ] F.Tetzlaff, R.-P. Braun, N. Hanik, "QoS Monitoring of Optical Transmission Links using Amplitude Histograms: Field Implementation and Experimental Evaluation", NOC 2000

[URL EBONE] http://www.ebone.com

[URL GTS] http://www.grtgroup.com

[VCID] K. Nagami et al., " VCID Notification over ATM Link", draft-ietf-mpls-vcid-atm-04, work in progress.

[VCSW] B. Davie et al., "MPLS using LDP and ATM VC Switching", draft-ietf-mpls-atm-02, work in progress.

[Wauters 1999] N. Wauters, G. Ocakoglu, K. Struyve, P. Falcao Fonseca, "Survivability in a New Pan-European Carriers' Carrier Network based on WDM and SDH Technology: Current Implementation and Future Requirements", IEEE Communication Magazine, page 63-69, vol. 37, No 8, August 1999

[WP1-M1] Milestone WP1 M1: "Network Requirements"

[WP1-M2] Milestone WP1 M2: "Network Scenarios and Guidelines for the Testbed Configuration"

[WP2-D7] Deliverable WP2 D7: "Failure Scenarios of Resilience in multi-layer networks"

[WP2-M1] Milestone WP2 M1: "First indications of Failure Scenarios and Resilience Strategies"

[WPG1-DISC] Answers from Discussion Meeting, April 13th-14th 2000

1.1.2 Abbreviations

AAL5	ATM Adaptation Layer 5
AAU	ATM to ATM User
ADM	Add Drop Multiplexer
AF	Assured Forwarding
AIS	Alarm Indication Signal
APD	Avalanche PhotoDiode
APS	Automatic Protection Switching
ASON	Automatically Switched Optical Network
ATM	Asynchronous Transfer Mode
AUP	Acceptable Use Policy
AWG	Arrayed Wave Guide
BA	Behavior Aggregate
BDI	Backward Defect Indication
BEI	Backward Error Indication
BGP	Border Gateway Protocol
BIP	Bit Interleaved Parity
BSHR	Bidirectional Self Healing Ring
CDN	Cable Data Network
CNI	Client Network Interface
CNM	Customer Network Management
CoS	Class of Service
CRC	Cyclic Redundancy Check
CR-LDP	Constraint Routed – LDP
CU	Currently Unused bits
DBR	Distributed Bragg Reflector
DCC	Data Communication Channel
DCF	Dispersion Compensating Fibre
DFB	Distributed FeedBack
DiffServ	Differentiated Services
DLCI	Data Link ConnectionIdentifier
DPT	Dynamic Packet Transport
DS	DiffServ
DSCP	Differentiated Service Code Point
DTM	Dynamic Transfer Mode
DWDM	Dense Wavelength Division Multiplexing
DXC	Digital cross Connect
ECN	Early Congestion Notification
EDFA	Erbium Doped Fiber Amplifier
EDFFA	Erbium Doped Fluoride Fibre Amplifier
EF	Expedited Forwarding
E-LSP	EXP-inferred-PSC LSP
ER-LSP	Explicit Routed – LSP
FBG	Fibre Bragg Grating
FDM	Frequency Division Multiplexing; is used as a synonym for DWDM
FEC	Forward Equivalent Class (in context of MPLS)
FEC	Forward Error Correction
FR	Frame Relay
GAN	Global Area Network
GbE	Gigabit Ethernet
GSMP	General Switch Management Protocol
GSR	Giga-Switch Router

GTS	Global TeleSystems
HDLC	High-level Data Link Control
HEC	Header Error Correction
HO	Higher Order
IAB	Internet Architecture Board
IANA	Internet Assigned Number Authority
ICMP	Internet Control Message Protocol
IESG	Internet Engineering Steering Group
IETF	Internet Engineering Task Force
Intserv	Integrated Services
IP	Internet Protocol
IPS	Intelligent Protection Switching
IS	IntServ
ISOC	Internet Society
ISP	Internet Service Provider
ITU	International Telecommunication Union
IWF	Inter-working Functionality
LAN	Local Area Network
LAPS	Link Access Procedure on SDH
LDP	Label Distribution Protocol
LER	Label Edge Router
L-LSP	Label-only-inferred-PSC LSP
LSA	Link State Advertisement
LSP	Label Switched Path
LSR	Label Switch Router
MAC	Media Access Control
MAN	Metropolitan Area Network
MAPOS	Multiple Access Protocol Over SONET
MEMS	Micro-Electro-Mechanical System
MPLS	Multi-Protocol Label Switching
MPS	Multiplex Section Protection
MPLS	Multi-Protocol Lambda Switching
MS	Multiplex Section
MSP	Multiplex Section Protection
MTBF	MeanTime Between Failure
MTU	Maximum Transfer Unit
NE	Network Element
NHRP	Next Hop Resolution Protocol
NL	Network Level
NM	Network Management
NNI	Network to Node Interface
NSP	Network Service Provider
O/E/O	Opto-Electro-Optical
OA	Ordered Aggregate
OA&M	Operation Administration & Maintenance
OADM	Optical Add Drop Multiplexer
OC	Optical Carrier
OCh	Optical Channel
ODU	Optical Data Unit
OMS	Optical Multiplex Section
OPU	Optical Payload Unit
OSC	Optical Supervisory Channel
OSNR	Optical Signal to Noise Ratio

OSPF	Open Shortest Path First
OTN	Optical Transport Network
OTS	Optical Transmission Section
OTU	Optical Transport Unit
OXC	Optical cross Connect
PHB	Per-Hop-Behavior
PLC	Planar Lightwave Circuit
POL	Packet Over Lightwave
POS	Packet Over SONET/SDH
PPP	Point-to-Point Protocol
PSC	PHB Scheduling Class
PVC	Permanent Virtual Circuit
QoS	Quality-of-Service
RDI	Remote Defect indication
REI	Remote Error Indication
RIP	Routing Information Protocol
RS	Regenerator Section
Rspec	Request Specification
RSVP	Resource reSerVation Protocol
SAP	Service Access Point
SDH	Synchronous Digital Hierarchy
SDL	Simple Data Link
SLA	Service Level Agreement
SLA	Service Level Agreement
SNCP	Sub Network Connection Protection
SOA	Semiconductor Optical Amplifier
SONET	Synchronous Optical NETwork
SPE	Synchronous Payload Envelope
SRP	Spatial Reuse Protocol
SRP-fa	Spatial Reuse Protocol Fairness Algorithm
SSG	Super-Sampled Grating
STM	Synchronous Transfer Module
STS	Synchronous Transport Signal
SVC	Switched Virtual Circuit
TCB	Traffic Conditioning Block
TCM	Tandem Connection Monitoring
TCP	Transmission Control Protocol
TDM	Time Division Multiplexing
TE	Traffic Engineering
TMN	Telecommunication Management Network
ToS	Type of Service
Tspec	Traffic Specification
TTL	Time-To-Live
UDP	User Datagram Protocol
UPSR	Unidirectional Path Switched Ring
VC	Virtual Channel (ATM context)
VC	Virtual Container (SDH context)
VCI	Virtual Channel Identifier
VP	Virtual Path
VPI	Virtual Path Identifier
VPN	Virtual Private Networks
WAN	Wide Area Network
WDM	Wavelength Division Multiplexing

1.1.3 Definitions

Connectionless	Network property, specifying that the network treats each individual packet of a session independently
Best-effort	Service property, specifying that the network attempts to deliver the data to the right destination but without any guarantee
Connectivity	Network property, specifying if there exists a route from source to destination node
Flow	Sequence of packets belonging to the same session or connection
End-node	Network node residing at the edge of the network, where flows originate or terminate
Router	Network node at the network layer (L3). A router is based on the store-and-forward principle and routes each packet independently, based on its destination address
Congestion	Overloaded network status, resulting in packet losses
Packet Loss	Loss of packets due to buffer overflow. Packet losses can be specified on a node or network level
Packet Delay	Time required for a packet to travel across the network form source to destination node.
Streaming Application	delay-sensitive (real-time) traffic [EN 301164]
Multimedia Application	Application based on more than one medium (e.g., video-conferencing which is based on voice and video media
Multicast	One-to-many communication
Guarantee	In this context referring to QoS guarantees. Note that such a guarantee is not necessarily very strict (e.g., elastic traffic)
Controlled Load Service	A service defined in the IS framework and perceiving a quality as it would receive on an unloaded network
Guaranteed quality of service	Another service defined in the IS framework, receiving a QoS which is mathematically guaranteed for bandwidth and delay (but not delay variation) requirements
Soft-State Protocol	A protocol which requires to refresh the states in all positions over the network from time to time.
EF	A PHB defined in the DS framework and aiming to provide leased-line or premium services
AF	Another PHB defined in the DS framework and giving the traffic at least a certain probability to be forwarded
Intserv	An architecture providing QoS and reserving resources on a per-microflow basis
DiffServ	Another architecture aiming to provide QoS by providing aggregated traffic with a particular per-hop-behavior
RSVP	Soft-state signaling protocol to request and confirm resource reservations
PATH-message	An RSVP message sent from source to destination specifying the traffic requirements (Tspec) and being used in the intermediate nodes to find the previous upstream node
RESV-message	RSVP message send from destination to source on receipt of a PATH message, requesting to reserve the required resources in intermediate nodes
LDP	Protocol to distribute the label bindings across the network
LSR	An MPLS router, which switches LSPs based on its label (similar to ATM nodes)
LSP	A concatenation of label bindings representing a logical connection from ingress to egress LSR.
CR – LDP	Label distribution taking into account some constraints in the routing of LSPs

ER – LSP	An LSP which is not routed following the IP routing tables
FEC	An IP prefix or host address.
Flow Driven	Same as data driven
Topology Driven	LSP setup is triggered by and following the layer 3 routing table changes
Downstream Allocation	Upstream node receives label mappings from the next-hop or downstream node. (LDP supports only downstream allocation)
Upstream Allocation	Downstream node receives label mappings from upstream nodes (this mode is not supported by LDP)
Unsolicited Distribution	Label distribution, which is not triggered by label requests
On-demand Distribution	Label distribution, which is initiated by label requests of upstream nodes
Pushed Distribution	See unsolicited distribution
Pulled Distribution	See on-demand distribution
Independent Control	Label distribution, towards the previous upstream node, which is not triggered by a label distribution from the next downstream node
Ordered Control	Distribution, allowing label advertisement, towards the previous upstream node, which is triggered by the receipt of a downstream label mapping
Liberal Retention	Mode which keeps temporary unused LSPs.
Conservative Retention	Mode which destroys LSPs which are not used anymore due to routing table changes
HELLO LDP-message	Message used by LDP to discover LDP neighbors and remote peers
INITIATION LDP-message	Message used in LDP, for negotiation about LDP session parameters
Label Request Message	Message used in LDP, to request a label binding advertisement from the next downstream LSR
Label Mapping Message	Message used in LDP, to advertise label bindings to the previous upstream LSR
Label Withdraw	Message sent upstream in LDP, to request the tear down of an LSPs
Label Release	Message sent downstream in LDP, to tear down LSPs
Notification LDP-message	Message used in LDP, to report failures and exceptions
Traffic Trunk	An aggregation into a single LSP.of traffic flows belonging to the same class
Induced MPLS graph	Represents the logical topology induced by the end-to-end LSPs (links) between LSRs (nodes)
Traffic Trunk Attributes	Attributes describing the behavior of the traffic
Resource Attributes	Attributes constraining the placement of traffic trunks through the corresponding resource
BA	Aggregation of traffic with similar behavior characteristics. The BA is based on the DSCP.
OA	A set of BAs that share an ordering constraint
PSC	A set of PHBs for traffic preserving an ordering constraint
E-LSP	An LSP with a PHB to EXP-field mapping
L-LSP	An LSP with a PSC to label mapping and drop preference to EXP-field (or appropriate field in other headers than the shim header) mapping

1.2 Deliverable history

Version	Date	Authors	Comments
DT-LION-WP1Task1_01j.doc	30 June 2000	WP1 Task1	WP1T1 working document
CM1T2_IMEC_v201.doc	28 June 2000	WP1 Task2	WP1T2 working document
CM_WP1T3_UPCv04e.doc	04 August 2000	WP1 Task3	WP1T3 working document
WP1_T4_v202.doc	24 July 2000	WP1 Task4	WP1T4 working document
LION_WP1M1(v100).zip	11 September 2000	LION	LION Milestone WP1M1
WP1-M02_v1.doc	04 September 2000	LION	LION Milestone WP1M2
D06_WP1_v_01	25 September 2000	LION	LION Deliverable D6 - draft
	30 September 2000	LION	LION Deliverable D6
D06_WP1_v3.doc	18 October 2000	LION	LION Deliverable D6
D06_WP1_FINAL.doc	27 October 2000	LION	LION Deliverable D6 - Final

2 Drivers and Network Requirements

Before considering requirements of evolving networks, it should be briefly considered the tipology of business models that may be applicable. The following business models have been identified:

- ASP: companies offering business software applications and solutions to enterprise remotely across the network which they operate. The revenue per data-bit is greatest at the value chain.
- ISP: companies offering Internet access, as well as some value-added services including e-mail, web-site hosting, content provisioning and some IP telephony services. ISP can either own the Layer 1 infrastructure or lease it.In the latter case there is a client server relatioship between the IP layers and the infrastructure layers.
- NSP and carriers: companies offering retail services. They include ex-incumbent telecom operators and some next generation carriers. They players may build their own networks or lease bandwidth from lower in the valuer chain.
- Carriers'Carrier: these companies build out the infrastructure supporting all network services. The client network is likely to be a circuit network, potentially in the same layer (even if taking part in another operators network).

More stringent requirements should be applied for 3 and 4 as both these cases have trust and sicurity issues between server and its client business, and both can have multiple instances of client networks. This need indicates that there is a greater need for partitioning than by technology. In the case of 1 and 2 the only partitions are driven by technology and layer bounderies.

These potential business models show that the support for different type of protocol and carrier services is required. In this sense, multiple protocol and carrier services is acting as a strong drivers for designing network architectuire **carrying multiple client types**.

Another well-known driver is the data traffic growth. The volume of data traffic has grown significantly in the last years and it is expected to continue its growth in the future: within ten years the network capacity demand could be up to 100 times the current one. With the rapid growth in demand for network resources, not only **bandwidth** but also **traffic engineering** has become another key requirement. As a matter of fact, IP-based networks are expanding rapidly and encountering many scalability problems. As the volume of network traffic increases, congestions may occur more and more frequently. Hence traffic engineering is needed to redirect traffic away from shortest path to less congested routes. With the increasing of router throughput, ruoted cores have become competitively fast, but still remaining the traffic engineering rather poor.

The level of integration of the network services within the infrastructure impacts the network capability for advanced features such us **policy management**. By integrating the network services in the network itself, will result benefit from the reduced complexity and costs of the overall network environment.

Policy management seems to be the next success critical factor. Policy management allows for the definition of rule sets for operating and administrating the network. For example a NSP can create dinamically the definition of classes of services to be provided for different users, subnets, applications or a combinations of these variables. The Policy Management approach will allow service-based virtual networks to make the information available in a secure way to the targeted people at the right time.

A multitude of service providers are migrating towards the next generation networks from different backgrounds. The tendency is moving towards **operating a single multi-service network for voice, data and video**.

In summar, the following new network requirements are emerging:

- client independency
- scalability
- policy management
- efficient and cost-effective resilience
- automatic end-to-end provisioning
- fast and efficient routing (even at Layer 1)
- policy-based traffic engineering for QoS
- support of Optical Virtual Private Networks

3 Application and transport services

The section "Network requirements – analysis of services in transport networks" is based on the results of the defined in LION WP1 Task 1 (Network requirements – analysis of services). The working document of Task 1 has been included in Milestone WP1M1 Network Requirements.

Two categories of services have been defined and analysed: application-oriented and transport services. The former are services which are offered to the customers of the network. The latter are services which may be offered by the lower layers (transport layers) of the network, such as the OTN. The focus within this document is on the transport services.

3.1 Application oriented services

The subsection presents results of the studies on the most important application oriented services, which in opinion of the authors have to be taken into account in evaluation of the Optical Transport Network Requirements. The pre-selected set of application oriented services has been described in the WP1M1 Milestone in order to provide clear depiction of functionality of applications characterized. The following applications have been included: POTS (Plain Old Telephony Service), Voice over IP, Videotelephony, Videoconferencing, Internet Access, TV Broadcast, TV Delayed Broadcast, TV Listing, Movies on Demand, News on Demand, Teleworking (Telecommuting), Distance Learning, Tele-shopping and Virtual CD-ROM. Description of each of the applications has been concluded with Application Technical Requirements specification. Description of the specified applications and theirs technical requirements can be found in WP1M1. In this deliverable only the most important results of the work are included.

Application defined and characterized has been classified into two, very general categories. The classification has been based on the technical requirements determined for given applications. The first category of applications is based on the information sent to user in a streaming manner. The source continuously produces data stream during the end-user session. The source is bursty, mostly due to the adaptive nature of video coding schemes. The application classified into the ***Streaming Service*** is sensitive or highly sensitive to the delay introduced by the transport network.

The other group of applications are characterized as a much less delay sensitive. This group has been assigned to the ***Elastic Service***. Data bit rate variation of this group is much higher in comparison to the applications of Streaming Service. Elastic Service applications are in most cases based on information retrieval. Next parameter which differentiate Streaming and Elastic services is delay sensitivity.

The application classification depends in some cases on the content of the service is offering. For example News on Demand application categorized into Elastic Service group can be moved to the Streaming Service class in the case of on-line financial markets tracking. Tele-shopping service can be redefined with on-line auction systems taken into account. It is also possible to assign some Streaming Service applications to the Elastic Service group by the specific modification of the application definition.

Two parameters should be defined to identify an application services scenario: the bandwidth requirements and the service's degree of penetration in terms of percentage spread across customers for each type of service. The second parameter is the percentage of customers who subscribe the service, and generally it is obtained through market surveys.

Applications described and analyzed in Milestone WP1M1 are summarized in the Table 1. The assignment of given application to the Streaming or Elastic Transport Service group is also indicated in the following table. Applications are summarized first of all with indication of the data bit rate specific to the service and level of expected data rate variation. Application delay sensitivity has been indicated with the grade ranging form 0 up to 5. The highest grade means that the application is highly delay sensitive, the grade o 0 would mean that the delay in the transport network can be neglected. The same rating has been used for need for protection. Each of the applications has been also analyzed with regard to its expected importance in the short and mid-term perspective. It was assumed that short-term prognosis means prognosis in three years perspective and mid-term forecast relates to the application importance in 10 years. The importance of the application is give in the scale from –5 to +5. The rate of 0 means no changes in the applications importance. The negative ratings is adequate to forecasted degradation of the application, the positive value corresponds to increased importance of the service. The aggregate rating of importance prognosis is higher then zero due to the belief of continuing increase of networked services. The applications were analyzed with regard to the fiber optic based transport networks. Degradation of the TV broadcast services are resulted from belief of fast expansion of digital TV systems based on satellite links.

Most of the applications were find to be appropriate to the Streaming Service. The importance of the streaming services probably will increase in the future. ATM LAN Market Analysis: Quantitative Market Research report prepared for the ATM Forum by Sage Research presents the forecast of delay-sensitive traffic volume increase from 17% in 1997 up to 28% in 1999. The forecast has not been verified (nor data are accessible). Own analysis performed for this deliverable shown that the trend probably will observed and even intensified in the future. The basis of this prognosis is active support of services like teleworking and distance learning from many governments and organizations (including EU) and increased interest in these services from end-users.

Table 1: Application overview and its relationship to the Application Services

	Application:	Bit rate	Bit rate variation	Delay sensitivity	Need for protection	Importance	
						Short-term	Mid-term
Streaming Service	POTS	64 Kbps–32Kbps	Const.	5	5	-1	-3
	VoIP	8 Kbps–32 Kbps	Const.	5	5	3	5
	Video-telephony	256Kbps–1920Kbps	High	5	5	2	3
	Video-conference	256 Kbps at least	High	5	5	3	4
	Tele-working	64 Kbps–2 Mbps	Very high	5	4	3	5
	TV Broadcast	2 Mbps–8 Mbps	High	4	4	1	0
	Delayed TV Broadcast	2 Mbps–8 Mbps	High	4	4	1	0
	Distance Learning	64 Kbps–2 Mbps	Very high	5	5	2	4
	MoD	750 Kbps–4Mbps	High	4	3	1	3
Elastic Service	News on Demand	64 Kbps	Very high	2	2	2	4
	Internet Access	64 Kbps–2 Mbps	Very high	1	2	5	5
	TV Listing	64 Kbps	Very high	2	2	1	2
	Tele-shopping	64 Kbps–2 Mbps	Very high	2	2	5	5

Optical Transport Network has to be designed in a way enabling easy and reliable transport of this type of data. Data bit rate of streaming applications vary from 8 Kbps up to the 8 Mbps. The data bit rate required in the transport network by specific application vary with coding scheme applied to the incoming stream (video, audio stream). Notable progress in the video coding is still observable. New, faster and more effective codecs are available at the market. Application of new coding techniques probably will result in decreasing requirements for data rate. To forecast required throughputs complex techno-economic studies would be necessary. This is far away from the scope of the LION Project.

Streaming Service data is highly sensitive to the networking delay. Even in the case of TV like services short system (video server) response time in necessary, mostly due to the requirements of assuring VCR like functions (freeze, stop, forward, etc). The value of 250 ms is the most common in the subjective literature.

Need for protection for given applications varies with the type of end-user. Transport of digital video signal between movies server and end-user and between TV studios is characterized with different protection requirements. The most important conclusion is that Streaming Service need to be supported with protection mechanism of future transport network.

Forecasting of importance of the given applications are very difficult. Presented rating was based on available market research reports and discussed in the team working in LION project. The assigned importance of the applications is highly reliable to its definition. Presentation of application definitions were one of the goals of this deliverable. Importance rating will be discussed during project duration. Change of the application importance rating at the beginning and at the end of project was found to be the most reliable and interesting parameter of applications analysis.

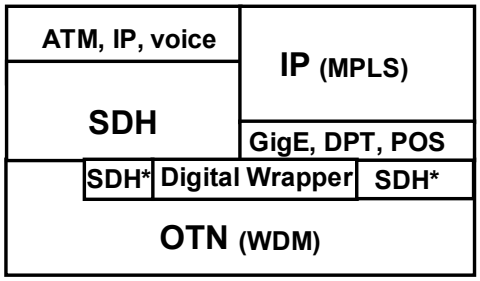
Some applications have been assigned to the Elastic Service. Data bit rate of this application group vary from 64 kbps up to 2 Mbps. Data bit rate variation is very high. This effect is observed because Elastic Service applications are in most cases based on information retrieval. The requested information is gathered by the end user (session is seen as an active one) and then some period of gathered data analysis can be noticed (at the transport network session seen as an inactive).

Elastic Service is not as highly as Streaming Service delay sensitive. Probably this parameter will have to be modified in the near future because of the possible change in the end users expectations. WWW service is likely to evolve toward fast, on-line services with no delay toleration.

The work done on the applications analysis should be continued. There is necessary to carry out the quantitative analysis of most likely applications in order to provide WP3 with clear view what share bandwidth will be consumed by Streaming Service. At this moment simple extrapolation of available data shows that streaming services will take about 46% of bandwidth of future transport networks.

3.2 Transport Services

The transport services described in the following are based on the network architecture depicted in Figure 1.



SDH* = SDH framing

Figure 1: Network architecture

Network transport services are those functions and utilities supporting connectivity, communications and control required by applications operating across the network applications. The definition of these network transport services will determine the flexibility of an enabling infrastructure to support current and unforeseen new network application. The tight link between network infrastructure and transport services will benefit Network and Service Providers from the reduced complexity of the overall network environment. This complexity reduction could lead to simpler management and cheaper overall network costs.

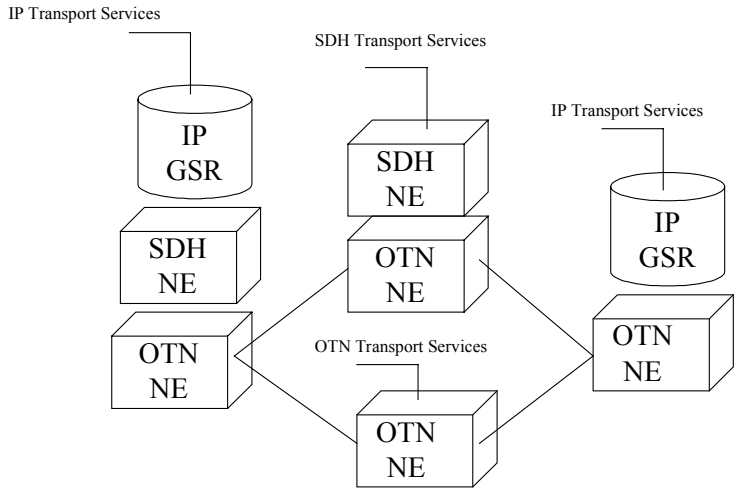


Figure 2: - Example of Transport Services

3.2.1 TDM-Transport Connection Service

3.2.1.1 SDH Leased Line Services

SDH Leased Line Services provide for transport of bi-directional lower order and higher order VC. Interconnection takes place at STM-N interfaces. A subset of possible solutions is defined in EN 301 164 [EN 301164] and EN 301 165 [EN 301165]

Timing Tolerance

The leased line connection shall carry user timing with a tolerance of ±4,6 ppm. It is recommended to use SDH synchronization as specified in ETS 300 462-2-1 [ETS 300 462-2-1]. The network termination point shall transmit a SDH section signal that could be used for generating the VC timing at the terminal equipment.

Transfer Delay

The one way end-to-end transfer delay of the connection mainly depends on geographical distance and the transmission media.

Jitter

In order to meet the jitter requirements the AU/TU pointers shall comply to the specification in EN 300 417-3-1 [ETS 300 417-3-1] and EN 300 417-4-1 [ETS 300 417-4-1] respectively. The leased line shall operate as specified when the jitter at the STM-N interface is within the limits given in Recommendation G.825 [ITU-G825].

Information Transfer Susceptance

The connection should be capable of transparently transferring a complete VC expect of the N1/N2 byte of the path overhead, which is used for Tandem Connection Monitoring application.

Error Performance

Measurement of the error performance of an SDH connection is based on the following parameters which are defined in Recommendation G.826 [ITU-G826].

- Errored Block (EB):

  An EB is a block in which one or more bits are in error.

- Errored Second Ratio (ESR)

  An errored second (ES) is a one second period with one or more errored blocks or at least one defect. The ESR is the ratio of ES to total seconds in available time during a fixed measurement interval.

- Severely Errored Second Ratio (SESR)

  A severely errored second (SES) is a one-second period which contains ≥ 30% errored blocks or at least one defect. SES is a subset of ES. The SESR is the ratio of SES to total seconds in available time during a fixed measurement interval.

- Background Block Error Ratio (BBER)

  A background block error (BBE) is an errored block not occurring as part of an SES. The BBER is the ratio of BBE to total blocks in available time during a fixed measurement interval. The count of total blocks excludes all blocks during SESs.

- Unavailable Seconds (UAS)

  A period of unavailable time begins at the onset of ten consecutive SES events. These ten seconds are considered to be part of unavailable time. A new period of available time begins at the onset of ten consecutive non-SES events. These ten seconds are considered to be part of available time. A bi-directional path is in the unavailable state if either one or both directions are in the unavailable state.

Performance objectives are given in Recommendation G.828 [ITU-G828].

Availability Performance

In Recommendation G.827 [ITU-G827] the following availability performance parameters are defined:

- Availability Ratio

  Availability ratio (AR) is defined as the proportion of time that a path element (PE) is in the available state during an observation period.

- Mean Time Between Digital Path Outages

  The mean time between digital path outages (Mo) for a digital path portion is the average duration of any continuous interval during which the portion is available. The reciprocal of the Mo is called Outage Intensity (OI).

Provisioning Time

The time to set-up a VC connection (with existing infrastructure) is mainly determined through provider's administrative processes. The duration of the provisioning action through the NMS is in the range of minutes depending on the size of the network.

3.2.1.2 SDH concatenated services

TDM multiplexing and TDM framing in new optical transport networks is focused on higher order SDH path layer (VC-4). In order to achieve payload bandwidths larger then the basic VC-4 container the principle of concatenation has been defined. The following description of the concatenated formats is based on ITU-T G.707 [ITU-G707].

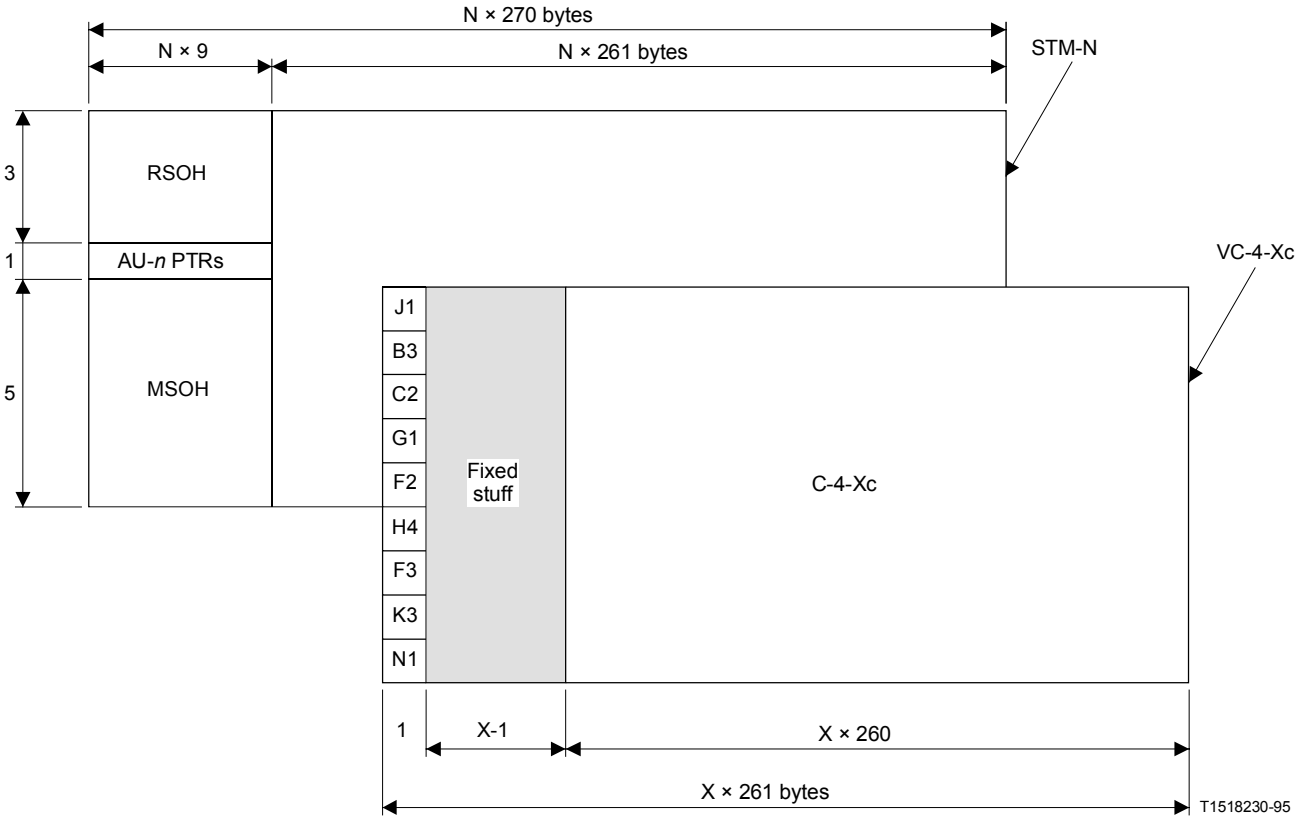
Concatenation definition: A procedure whereby a multiplicity of Virtual Containers is associated one with another with the result that their combined capacity can be used as a single container across which bit sequence integrity is maintained

Concatenation of AU-4

AU-4s can be concatenated together to form an AU-4-Xc (X concatenated AU-4s) which can transport payloads requiring greater capacity than one Container-4 capacity. In principle it is distinguished between contiguous concatenation and virtual concatenation

Concatenation of contiguous AU-4s

A concatenation indication, used to show that the multi Container-4 payload carried in a single VC-4-Xc should be kept together, is contained in the AU-4 pointer. The capacity available for the mapping, the multi Container-4, is X times the capacity of the Container-4 (e.g. 599 040 Mbit/s for X = 4 and 2 396 160 kbit/s for X = 16). Columns 2 to X of the VC-4-Xc are specified as fixed stuff. The first column of the VC-4-Xc is used for the POH. The POH is assigned to the VC-4-Xc (e.g. the BIP-8 covers 261 X columns of the VC-4-Xc). The VC-4-Xc is illustrated in Figure 3.



PTR Pointer

Figure 3: VC-4-Xc structure

Virtual concatenation of AU-4s

Up to now G.707 [ITU-G707] has not specified the virtual concatenation of VC-4. However it is specified the method for TU-2s, and the virtual concatenation of VC-4 will be based on the same basic principles.

Based on the specification in G.707 [ITU-G707] this method of concatenation allows for the transport of a single VC-2-mc in m × TU-2 without the use of Concatenation Indication in the pointer bytes. The method only requires the path termination equipment to provide concatenation functionality.

Virtual concatenation requires the concatenated Tributary Unit signals at the origin of the path to be launched with the same pointer value. The so formed Tributary Units at each interface shall be kept in a single higher order VC-4.

When the higher order VC-4 is terminated, the restrictions that apply in passing the concatenated Tributary Units from one interface to another is that all of the concatenated Tributary Units are connected to a single higher order VC-4 and that the time sequencing of the concatenated Tributary Units is not altered. Differences in delay of the individual concatenated VC-2 signals may occur due to pointer processing at intermediate equipment. The maximum difference in pointer value within a concatenated group at any interface is for further study. At the path termination the VC-2-mc can be reconstructed by using the pointer values for alignment.

Each concatenated VC-2 signal will carry its own POH. At the VC-2-mc path termination, the individual BIP-2s are aggregated to give a single BIP error monitor.

3.2.1.3 SDH VPN

SDH VPN is a service provided by the SDH TMN and does not require any equipment functions in addition to classical SDH equipment. A VPN comprises a set resources to which authorized customers can apply special customer network management functions according to different authorization levels.

VPN Resources

- Ports/Interfaces of Network Elements
- Connections between Network Elements
- Possible Sub-VPN levels

CNM Functions/CNM Interface

- Configuration Management Functions
- Fault Management Functions
- Performance Management Functions
- CNM Security Functions
- Accounting Functions

3.2.2 Basic Mapping of native Signals into TDM – Frame

Table 2: Mapping of Protocols

Protocol	Bitrate	SDH mapping	Utilisation
VC-12	2 Mbit/s	X	
VC-3	34 Mbit/s	X	
VC-4	140 Mbit/s	X	
STM-1	155 Mbit/s	X	100 %
STM-4-4c	622 Mbit/s	X	100 %
STM-16-16c	2.5 Gbit/s	X	100 %
STM-64-(64c)	10 Gbit/s	X	100 %
(STM-256-256c)	40 Gbit/s	X	100 %
Ethernet	10 Mbit/s	VC-3	ca. 30 %
Fast Ethernet	100 Mbit/s	STM-1	ca. 65 %
Gigabit Ethernet	1.25 Gbit/s	STM-16c	ca. 50 %
10 Gigabit Ethernet	10 Gbit/s	?	
FDDI	100 Mbit/s	STM-1	ca. 65 %
Serial HIPPI-800	800 Mbit/s	STM-16c	ca. 32 %
Serial HIPPI-1600	1.6 Gbit/s	STM-16c	ca. 64 %
Serial HIPPI-6400	6.4 Gbit/s	STM-64c	ca. 64 %
ESCON	200 Mbit/s	STM-4-4c	ca. 32 %
(3 x ESCON)	600 Mbit/s	STM-4c	ca. 96 %
Fiber Channel/FICON	1.0625 Gbit/s	STM-16c	ca. 42 %
Coupling Facilities	1.0625 Gbit/s	STM-16c	ca. 42 %
Digital Video	270 Mbit/s	STM-4c	ca. 43 %
Serial Digital HDTV	1.485 Gbit/s	STM-16c	ca. 60 %

Example of application of concatenated VC-4 – mapping of ATM-cells

The mapping of ATM cells is performed by aligning the byte structure of every cell with the byte structure of the Virtual Container used including the concatenated structure (VC-x or VC-x-mc, $x \geq 1$). Since the relevant Container-x or Container-x-mc capacity may not be an integer multiple of the ATM cell length (53 bytes), a cell is allowed to cross the Container-x frame boundary.

The ATM cell information field (48 bytes) shall be scrambled before mapping into the VC-x or VC-x-mc. In the reverse operation, following termination of the VC-x or VC-x-mc signal, the ATM cell information field will be descrambled before being passed to the ATM layer. A self-synchronizing scrambler with generator polynomial $x^{43} + 1$ shall be used. The scrambler operates for the duration of the cell information field. During the 5-byte header the scrambler operation is suspended and the scrambler state retained. The first cell transmitted on start-up will be corrupted because the descrambler at the receiving end will not be synchronized to the transmitter scrambler. Cell information field scrambling is required to provide security against false cell delineation and cell information field replicating the STM-N frame alignment word.

When the VC-x or VC-x-mc is terminated, the cell must be recovered. The ATM cell header contains a Header Error Control (HEC) field which may be used in a similar way to a frame alignment word to achieve cell delineation. This HEC method uses the correlation between the header bits to be protected by the HEC (32 bits) and the control bit of the HEC (8 bits) introduced in the header after computation with a shortened cyclic code with generating polynomial $g(x) = x^8 + x^2 + x + 1$.

The remainder from this polynomial is then added to the fixed pattern "01010101" in order to improve the cell delineation performance. This method is similar to conventional frame alignment recovery where the alignment word is not fixed but varies from cell to cell. More information on HEC cell delineation is given in Recommendation I.432 [ITU-I 432].

The ATM cell stream is mapped into a Container-4-Xc with its byte boundaries aligned with the Container-4-Xc byte boundaries. The Container-4-Xc is then mapped into VC-4-Xc together with the VC-4-Xc POH and (X–1) columns of fixed stuff. The ATM cell boundaries are thus aligned with the VC-4-Xc byte boundaries. Since the Container-4-Xc capacity (X × 2340 bytes) is not an integer multiple of the cell length (53 bytes), a cell may cross a Container-4-Xc frame boundary.

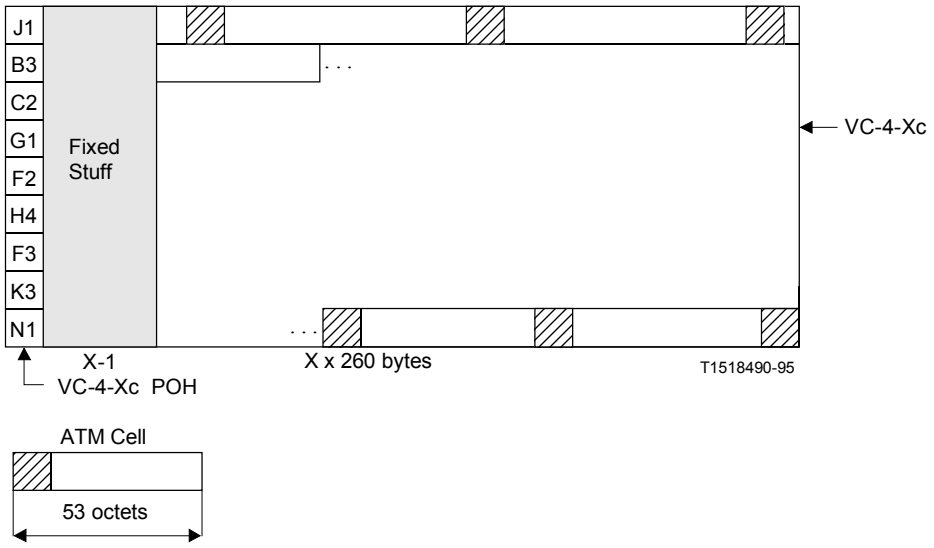


Figure 4: Mapping of ATM cells into VC-4-Xc

3.2.3 Optical Channel Transport Service

An OTN is a transport network bounded by optical channel access points.

An OTN Transport Service could be defined as the capability of an OTN to transport client information between two optical channel access points with a given degree of transparency.

Three main kinds of Optical Transport Services can be identified:

- Leased OCh service
- Leased Wavelegth
- Leased Dark Fiber

An OTN should be capable of supporting the following types of services for Leased OCh:

- Provisioning of leased OCh service

Two classes of leased OCh services, based on the digital container, are identified.

- Leased OPU

This service offers a defined digital container restricted to the Optical Payload Unit (OPU) area as specified in G.709. Customer can use the OPU data area freely to convey any stream type digital client signals. Customer needs to fill the OPU with fixed stuffing if the data rate of his concern is lower than that of the offered OPU. The OCh itself is fully managed and supported by the management system and equipment under operator's responsibility.

- Leased ODU/OTU

This service offers a defined digital container bounded by the Optical Transport Unit (OTU) area as specified in G.709. Customer can manage OCh overheads with the capability of connection monitoring specified in G.872. The end-to-end ODU and the series of OTU may or may not be supported by the customer's equipment.

- Bandwidth management of leased OCh

Further special service option is considered as to the bandwidth management of leased OCh. This special service might be offered when the inverse multiplexing of OTN, e.g., OCh virtual concatenation is available.

- Different ways to set-up of OCh service:
  - a permanent optical channel set up from the network management system by means of network management protocols;
  - a soft permanent optical channel set up from the management system, which uses network generated signalling and routing protocols to establish connections;
  - a switched optical channel which can be set up by the customer on demand using signalling and routing protocols.

Leased Wavelegth is offered to clients equipped with coloured line terminals.

In case of Leased Dark Fiber who is providing this service doesn't control the degree of use of the fiber itself.

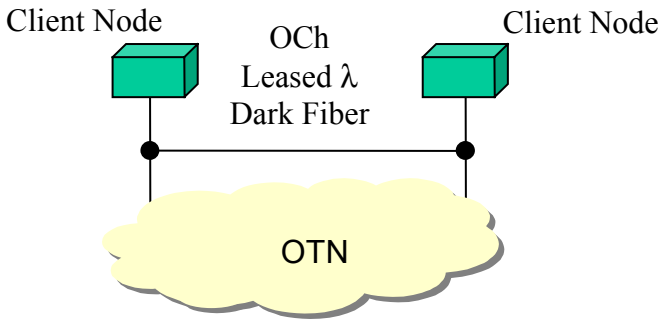


Figure 5 - Example of Optical transport service

A set of parameter could be used to qualify the Optical Channel Connection:

- access mode
- rate of service availability
- multiple service quality classes
- multiple service protection classes
- routing configuration (VPN)
- bandwidth management
- security
- provisioning time

3.2.3.1 Digital Wrappering

Draft Recommendation G.709 (April/2000) contains the basic information of an ONNI. Further details will be specified within the year 2000. The text hereafter describes the draft ONNI specification and addresses, in addition, the ongoing discussion regarding a detail specification of the ONNI frame structure.

The Optical Transport Network as specified in ITU-T Recommendation G.872 defines two ONNI interface classes:

- Inter-Domain Interface (IrDI)
- Intra-Domain Interface (IaDI)

Figure 6 gives a possible network configuration to illustrate the location of OTM IrDI and OTM IaDI.

The Optical Transport Network will consist of a set of network operator specific networks, which are interconnected via OTM IrDIs (Figure 7). Each network operator specific network may consist of one or more manufacturer specific subnetworks. Those manufacturer specific subnetworks are interconnected via OTM IrDIs. OTN equipment within a manufacturer specific subnetwork may be interconnected via OTM IaDIs and OTM IrDIs.

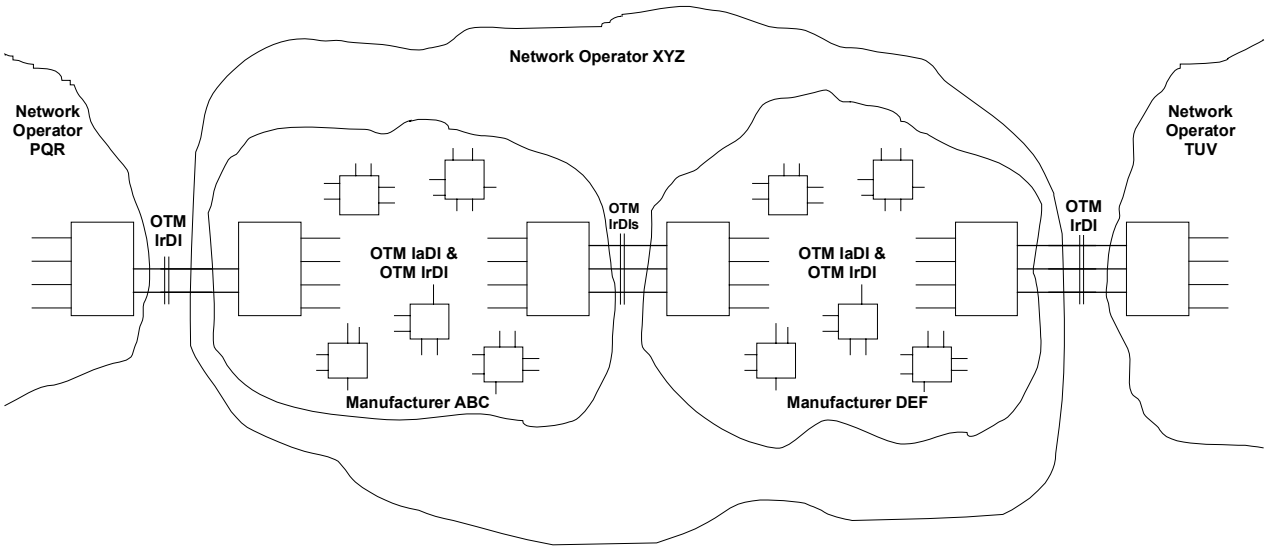


Figure 6 : Locations of OTM IrDI and IaDI

The IrDI is subject to standardization (see Figure 7). The IaDI is not subject to standardization, except for the properties necessary for network management. However it includes the ODUk as a standard information structure which is used through out the network in order to allow network wide interworking of optical channels.

NOTE – Any non-standard interface is defined as an IaDI. Any standardized interface is defined as an IrDI.

The Optical Transport Module-n (OTM-n) is the information structure used to support ONNI interfaces. Two OTM-n structures are specified per ONNI interface class:

- OTM-n (n≥1)
- OTM0

This results in four OTM-n interface structures: OTM-n IrDI, OTM0 IrDI, OTM-m IaDI, OTM0* IaDI.

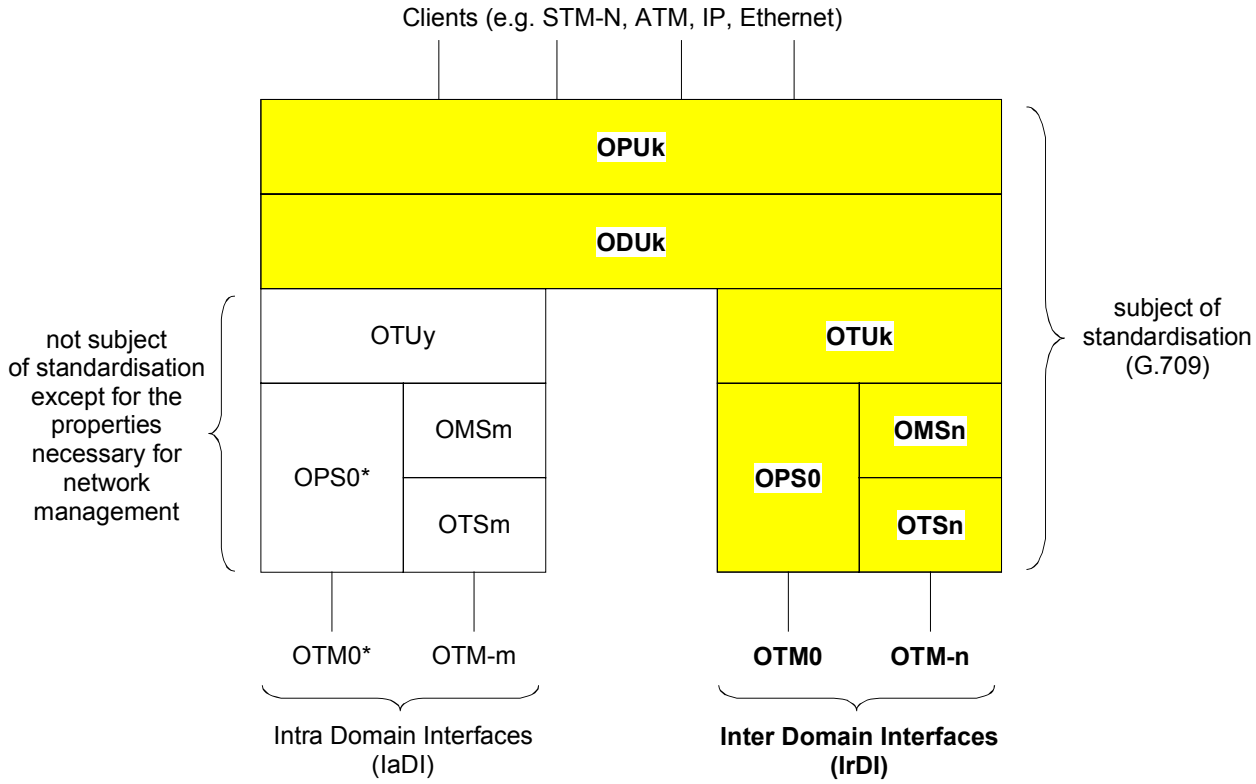


Figure 7 : Overview of the basic structure of standardized and propriety optical interfaces

The frame structure of the ODU/OTU bases on the Digital Wrapper (DW). The basic assignment of overhead functions to the frame is illustrated in Figure 8 and Figure 9. It is agreed to specify a 4 times 16 byte overhead plus a 4 times 256 bytes FEC field that may be unused or filled with RS239/255 FEC code. The size of the ODUk payload is not yet fixed. It depends on mapping scheme for STM-N (byte synchronous or byte asynchronous).

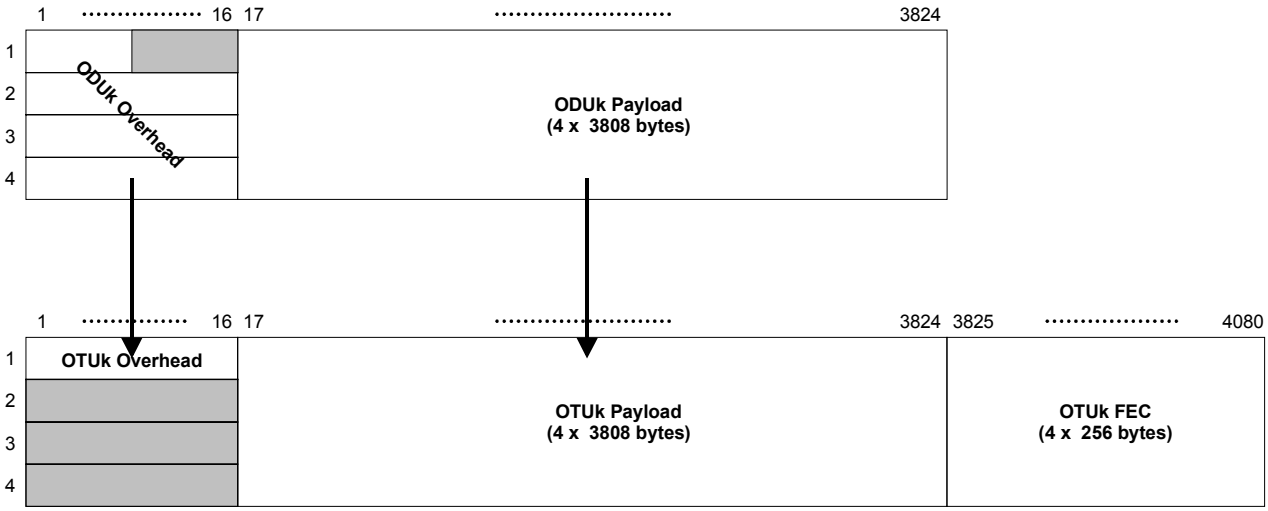


Figure 8 : Frame synchronous mapping of ODUk into OTUk signal

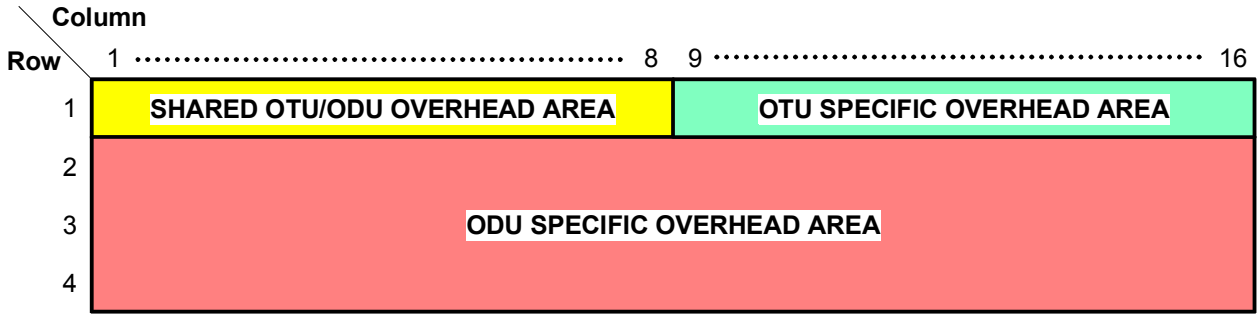


Figure 9 : OTUk and ODUk Overhead

Network functions provided by the DW

The DW provides functions for:

a) Continuity supervision (loss of frame, AIS generation and detection for OTU and ODU).

b) Connectivity supervision (trail trace for OTU and ODU).

c) Quality supervision (BIP-8 and 4 bit REI for OTU and ODU).

d) Transport for information for network operators.

The supervision functions listed under item a) to c) are provided for 8 nested levels and are called connection monitoring. The coding for each level is identical. One level is assigned for the end-to-end path and one for the OTU. Figure 10 shows 3 levels of connection monitoring of an ODU (OTU monitoring is not illustrated). One level is used by the user for end-to-end monitoring, one for the leased line (along domains A, B and C) and one for each domain. In addition, network operator B uses a non-intrusive monitor for reading the end-to-end quality and the leased line quality.

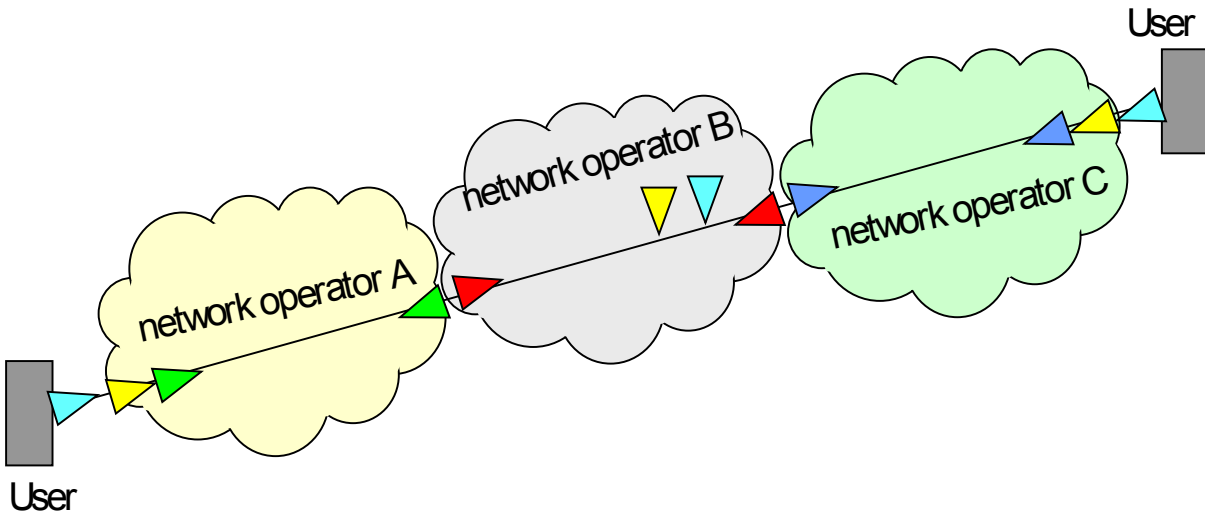


Figure 10 : Example of nested connection monitoring

Relation between transparent domains and DW

As shown in Figure 7 the ONNI consists of a common part (ODU) and parts that may be standardized or propriety (OTU, OMS and OTS). An ODU can be transported over administrative domains while an OTU is a signal within or between administrative domains or transparent domains respectively. This is illustrated in Figure 11.

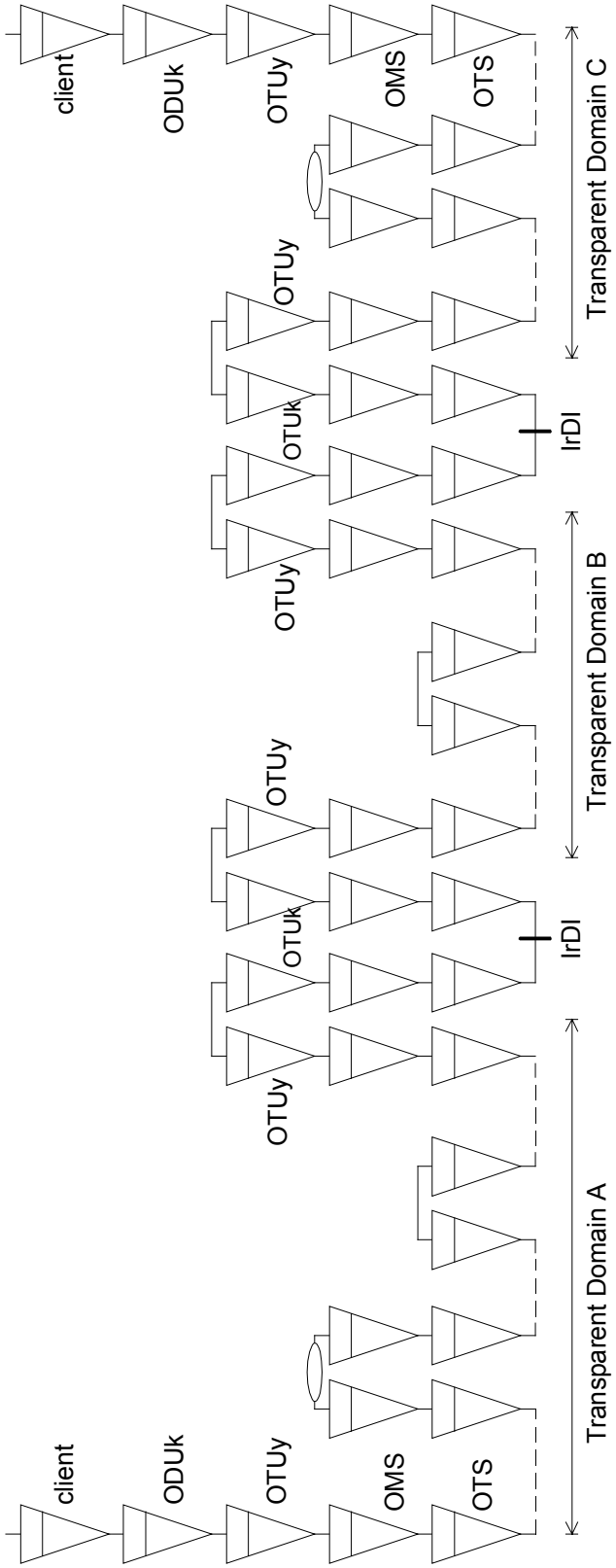


Figure 11 : Relation between transparent domains, OTU and ODU

The current specification of the ONNI include only a structure of a single ODU per OTU, i.e. multiplexing of ODUs in order to increase the bit rate per wavelength is foreseen for the 2nd version of the ONNI. However, the automatic switched optical network (ASON) and Time

Division Multiplexing (TDM) of DW signals influence the design of the OTN and, therefore, structure and bit rate of the DW.

Effect of ASON on OTN

The discussion on an ASON has just started. The relation between ASON and OTN is not yet defined nor which signal to be switched. It may be that ASON and OTN each uses a fibre exclusively or that they share a fibre. The latter case implies that that ASON switches uses transport capacities of the OTN. ASON will provide switched ODUs when it is intended to offer switch connections between any locations of the world. Switched OTUs can be only provided within a transparent domain and therefore cannot be used for a world-wide switching network.

It is assumed that ASON will offer switched ODUs. It is the purpose of switched network to provide quickly bandwidth with an appropriate granularity. On the other hand, it is the purpose of the OTN to design OTN spans and nodes for a maximum capacity or throughput respectively. Example: Switched ODU1s of 2,5 Gbit/s provide the smallest granularity for ASON while an OTN may be optimised for OTU2s (10 Gbit/s) or OTU3s (40 Gbit/s). Hence, the forthcoming ASON gives reason for the standardisation of ODU multiplexing in future.

TDM of ODUs

It is the intention of network operators to run the spans between nodes with the optimum capacity, which is the product of bit rate per channel and number of channels. A best bit rate or a best number of channels exist only for a certain period. Both, best bit rate and number of channels increases over time. Unlike digital cross connect, transparent OCX have fixed cost per cross connection that independent on the bit rate per channel. These are two further reasons for TDM in the OTN, even when there is currently no need for it.

Three alternatives are discussed for the TDM scheme:

a) Layered ODU scheme like PDH (see Figure 12)

b) Lower and higher order ODUs (see Figure 13)

c) Flat TDM hierarchy (see Figure 14)

The layered ODU scheme has two disadvantages. It requires stuffing overhead (about 1% of an ODU) for each multiplex level. Cross connections of ODU1s requires the termination of all higher order ODUs.

The scheme of lower and higher order ODUs offers both, traffic aggregation and transport of aggregated traffic over administrative domains. This scheme has the disadvantage that the higher order ODU has to be terminated in case of lower order cross-connections.

The flat TDM hierarchy provides traffic aggregation, but no transport of aggregated traffic over administrative domains. It has the advantage of direct access to each ODU level.

Two orders of ODUs and the flat TDM hierarchy consume probably the same amount of stuffing overhead, i.e. both schemes lead to the same OTU bit rate. For these reasons, the lower and higher order ODU scheme may be favoured.

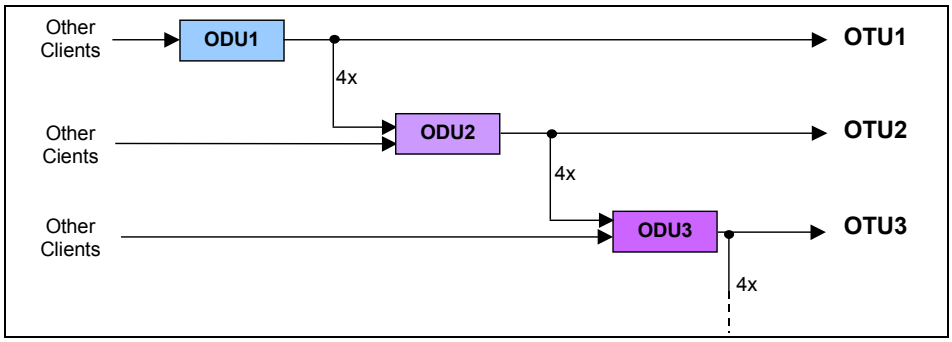


Figure 12 : Layered ODU hierarchy

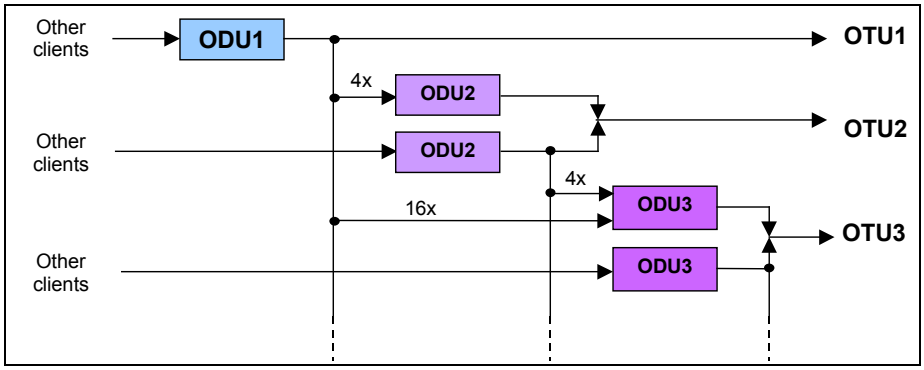


Figure 13 : Lower and higher order ODUs

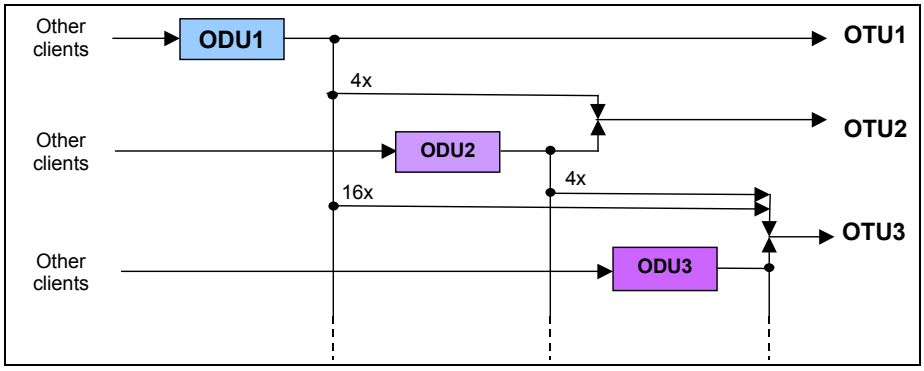


Figure 14 : Flat TDM hierarchy

SDH-Wrapped

Leased VC service provides leased lines with conventional granularities in the OTN. To achieve this service, three options of overhead-relay functionality is considered. The first one passes through all user overheads with the capability of B1 and J0 monitoring at the gateway equipment. The second one terminates RS at the gateway to obtain OCh overhead capacity, while passing through MS overheads. The last one terminates both MS and RS to make more room for OCh overheads. In this option user's VC network is completely separable from the OTN in terms of management capability if all the user overhead is terminated. As such, some portion of user overheads might be relayed to guarantee users' OA&M capability.

3.2.3.2 Service Parameters

Provisioning Time

Class of Restoration / Protection

Time to Restore

Type of Resilience

- Unprotected

  Two classes are recognized: normal unprotected OCh and extra traffic type OCh. Neither of them offers restorability in case of failure. The normal unprotected OCh is managed as a working OCh. It has no protection bandwidth, and as such shall be disrupted during the period of its failure. The extra traffic type OCh is offered by using the protection bandwidth (wavelengths). Its service shall be disrupted and replaced by protection OChs for other protected OChs in case of failure.

- Restored

  OChs categorized in this resilience class have no dedicated or shared protection OCh capacity. The only way to be recovered from a failure is the OCh restoration. This service has slower rapidity to restore traffic than the protected OCh service.

- Protected

  This resilience class is for the highest priority traffic. Normally, the dedicated OCh protection such as 1+1 and 1:1 is supported. The shared protection (1:N) architecture can be applied if a protection OCh shall not be accessed by multiple of working OChs at the same time. The example of 1:N protection is an OCh shared protection ring. For the multiple failure cases, since not all the protected traffic can be restored, the remaining portion of traffic will be restored by the restoration scheme.

WDM Inter-Channel Jitter/Delay

Performance Parameter

- Performance Parameter of DW bit interleaved parity level eight (BIP-8)
- Check Sum (Example SDH)

Bit Interleaved Parity (BIP)

The BIP calculation is used to measure the error performance of an SDH connection according to ITU-T G. 826/828 . The received BIP-Bytes are compared with the calculated BIP-Bytes on the receiver side. In a case of bit errors you get a difference between these compared BIP-Bytes

One byte is allocated for regenerator section error monitoring. This function shall be a Bit Interleaved Parity 8 (BIP-8) code using even parity. The BIP-8 is computed over all bits of the previous STM-N frame after scrambling and is placed in byte B1 of the current frame before scrambling.

The B2 bytes are allocated for a multiplex section error monitoring function. This function shall be a Bit Interleaved Parity N $\times$ 24 code (BIP-N $\times$ 24) using even parity. The BIP-N $\times$ 24 is computed over all bits of the previous STM-N frame except for the first three rows of SOH and is placed in bytes B2 of the current frame before scrambling.

One byte is allocated in each VC-4-Xc/VC-4/VC-3 for a path error monitoring function. This function shall be a BIP-8 code using even parity. The path BIP-8 is calculated over all bits of

the previous VC-4-Xc/VC-4/VC-3 before scrambling. The computed BIP-8 is placed in the B3 byte of the current VC-4-Xc/VC-4/VC-3 before scrambling.

Calculation of BIP-X:

The monitored signal is divided in bit sequences each with a size of X bits. The 1st bit of the BIP-X is the result of an even parity calculation over all 1st bits from all bit sequences. The 2nd bit of the BIP-X is the result of an even parity calculation over all 2nd bits from all bit sequences, etc.

The even parity calculation is done by set the corresponding bit in the BIP-X in such a manner that you have an even number of one's.

Supervision Functions / requirements

Continuity: support of loss of signal detection and loss of frame detection

Connectivity: support of 64-byte TTI aligned with 64-multiframe

TCM: support of eight level of nested tandem connection monitoring as specified in G.709 and G.872

Defect Indications: support of BEI, BDI, and ODU-AIS

FEC based Performance parameter

Forward Error Correction (FEC)

Error control to guarantee integrity of the transmitted data may be exercised using Forward Error Correction (FEC). In principle, the FEC encoder in the transmitter accepts data bits and adds redundancy according to a prescribed rule, thereby producing encoded data at a higher bit rate. The corresponding decoder on the receiver side exploits this redundancy to decide which message bits have been actually transmitted. The goal of this encoding-decoding mechanism is to reduce the degrading effects of transmission on the Bit Error Ratio to a tolerable level. The use of FEC, however, adds complexity to the system. Thus, an appropriate trade-off has to be found between acceptable error performance and increased bandwidth demand and system complexity

There is a high number of different error correcting codes which can be applied. Historically, these codes are classified into block codes and convolutional codes, depending on the implementation (or absence) of memory in the encoders. In optical transmission systems, block codes are used throughout, due to their efficient hardware implementation.

To generate a (n, k) block code, the channel encoder collects data in successive k-bit blocks. For each block, it adds n-k redundant bits generated by some mathematical rules, thus producing an overall encoded block of n bits (with n > k). Hence, the channel data rate coming out of the coder is always higher than the source data rate. Depending on the amount of added redundancy (i. e. the number of added n-k bits per block), and the type of the code, a specific amount of errors within every encoded block can be identified.

An important subclass of block codes are the so-called cyclic codes. A binary code is classified as cyclic, if the sum of any two code words is also a code word, any cyclic shift of a code word is also a code word.

For cyclic codes, an extended mathematical theory has been developed, using binary polynomials do describe the coding and decoding procedures. Linear feedback shift registers are applied to implement coding and decoding algorithms derived from these polynomials.

Examples of block codes are:

- Repetition codes, where a data bit is simply transmitted n times;

- Hamming codes, Maximum Length Codes, Cyclic Redundancy Check (CRC-) Codes, Bose-Chaudhury-Hocquenghem (BCH-) Codes are block codes with specific parameters;
- Reed Solomon Codes (RS-codes) are an important subclass of nonbinary BCH Codes, as the encoder operates on multiple bits (i. e. m-bit symbols) rather than individual bits. Hence, the encoding algorithm expands a block of k symbols to n symbols by adding n-k redundant symbols. These RS-codes make highly efficient use of redundancy, and block lengths and symbol sizes can be adjusted in a wide range of data sizes.

3.2.4 IP-based Transport Services

Before introducing the concept of IP Transport Service some preliminary definitions should be fixed. As known the OSI model provides a conceptual framework for communication between systems, but the model itself is not a method of communication. Actual communication is made possible using communication protocols. In this context the protocol is a formal set of rules that governs how nodes exchange information over a network medium. A protocol implement the functions of one or more OSI layers.

A given OSI layer generally communicates with three other OSI layers: the layer above, the layer below and its peer layer in the other networked systems. The Service User is the OSI layer that requests services from an adjacent OSI layer. The Service Provider is the OSI layer that provides a service to a Service User. OSI layers can provide services to multiple service users. The Service Access Point (SAP) is the conceptual location where one OSI layer can request a service to another OSI layer. The service provided by adjacent layers help an OSI layer to communicate its peer layer in another system.

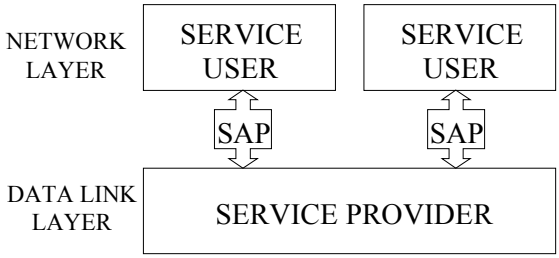


Figure 15 : Communication between OSI layers

The Internet protocols consist of a suite of communication protocols of which the two best known are the Transmission Control Protocol (TCP) and Internet Protocol (IP). The IP is a network-layer (Layer 3) protocol that contains addressing information and some control information that enables packets to be routed. IP has two primary responsibilities: providing connectionless, best-effort delivery of datagrams; providing fragmentation and reassembly of datagrams to support data links with different maximum-transmission unit size.

The Internet protocols consist of a suite of communication protocols of which the two best known are the Transmission Control Protocol (TCP) [RFC793 and RFC1122] and Internet Protocol (IP). TCP is a transport-layer (Layer4) protocol that controls the reliability of the end-to-end data transfer, while IP is a network-layer (Layer 3) protocol that contains addressing information and some control information that enables packets to be routed. Both software and hardware operating on TCP/IP networks typically consist of a wide range of functions to support data communications. In general, an IP Transport Service could be defined as the capability of an IP-based network to deliver datagram payloads from a Service Access Point to anyone of the interfaces with the IP-address for that SAP. In this sense, an IP Transport

Service is intended as an added value service to the simple data transmission. As such an IP Transport Service needs to be characterised.

If the end-user always acts via an application protocol, the Service Access Point of a Network Service Provider (NSP) could be either at the network-layer, or at the transport-layer, via intermediate protocols (e.g. TCP) or at the application layer via application protocols. This is shown in Figure 16.

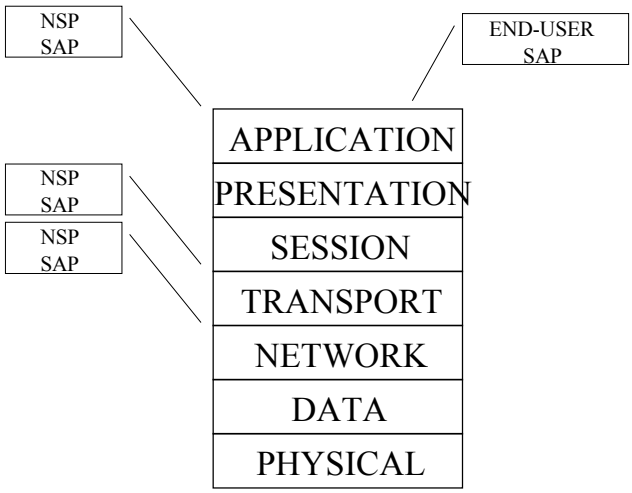


Figure 16: Definition of IP Transport Service on the OSI framework

Figure 16 depicts the IP Transport service on the conventional OSI framework. Nevertheless, the IP Transport service has to be placed on the Internet layer architecture shown in [BLAC], which is the OSI framework adapted to the TCP/IP networks.

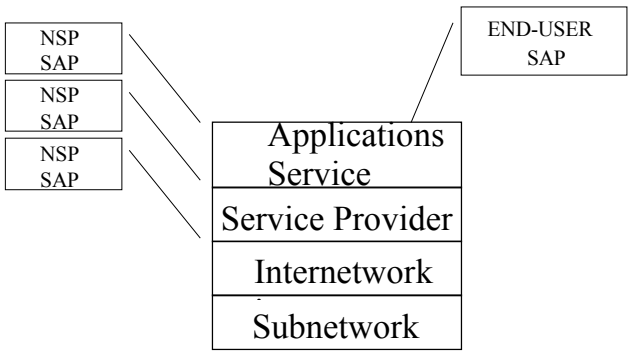


Figure 17: Definition of IP Transport Service on the Internet Layer Architecture

The bottom layer of the Internet model contains subnetworks and is thus called the *subnetwork* layer; it includes the subnetwork interfaces. Examples of subnetworks are Ethernet, in local area environment, and Frame Relay, ATM and OTN in wide area environments. The Internet subnetwork layer includes the link and physical layers of the OSI model.

The next layer is the *internetwork* layer. The internetwork layer provides the functions for connecting networks and gateways into one coherent system. This layer includes the IP and ICMP (Internet Control Message Protocol [RFC 792]) protocols.

The third layer is known as *service provider protocol* layer. This layer is responsible for end-to-end communications and includes the TCP and UDP protocols.

Finally, the upper layer is called the Application service layer. This layer supports the direct interface to an end-user application (FTP, HTTP, SNMP, SMTP, etc.)

According to this model, a more detailed definition of an IP Transport Service could be formulated including the following capabilities:

- Data transfer
- Reliability
- Flow and congestion control
- Operation and maintenance (O&M)
- Management

Data transfer is provided by IP. IP has two primary responsibilities: providing connectionless, best-effort delivery of datagrams; providing fragmentation and reassembly of datagrams to support data links with different maximum-transmission unit size.

Reliability: An IP transport based service can provide both reliable and unreliable data transfer between host computers attached to the Internet. TCP provides a connection-oriented reliable data transfer, while UDP provides a connectionless unreliable data transfer.

Flow and congestion control: TCP also provides end-to-end flow control. Flow control is performed in order to ensure the proper rate adaptation between sender and receiver, and secondly to prevent and survive possible congestion within the network.

Some sort of **Operation and Maintenance** capabilities can be set up using ICMP. ICMP includes two key mechanisms, namely *ping* and *traceroute*. The former mechanism provides an estimation of the mean delay to reach certain destination. The latter tells the user the current route provided by the network for reaching certain destination. These two mechanisms can be used, either manually or automatically, to monitor the IP transport service. For an automatic monitoring the ping and traceroute mechanisms can be enabled from an internal process, which would be periodically executed.

Management: The protocol used for the management of TCP/IP networks is the Simple Network Management Protocol (SNMP) [RFC1157 and RFC1441 to RFC1452]. SNMP includes three key capabilities, namely *get*, *set* and *notify*. Get enables the management station to retrieve the value of objects at the agent. Set enables the management station to set the value of objects at the agent. Finally, notify enables the agent to notify the management station of significant events. SNMP is at the application layer and goes on top of UDP.

3.2.4.1 IP Transport Service Characterization

According to the above definition, an IP Transport Service could be qualified by a set of parameters (to be adopted in the SLA), such as:

- Access mode
- Rate of service availability
- Multiple service quality classes vs performance
- Routing configuration
- Multicast

- Throughput management and traffic shaping
- Security
- Provisioning time
- Charging

The **access mode** defines how the user can assess the service: for example, switched access via ISDN or dedicated access via leased line such as 2 Mbit/s. The **rate of service availability** should indicate the service unavailability in hours per year.

Multiple service quality classes tied to the network latency or throughput characteristics provide an opportunity to differentiate between carrier services as well as to satisfy a full range of customer transport requirements via a single common interface. For example the DiffServ proposed by IETF, which makes use of the Type of Services (TOS) bit in the header of every IP datagram, permits the network to recognise how each packet should be handled. Premium (or Virtual Leased Line) service could include application services such as real time video-conferencing; on the other end the Assured service might include most of non-time sensitive application services such as file-transfers.

As the IP layer could run the role of integration layer for multiple services, particular attention should be paid in relating the IP performance parameters (such as packet loss ratio, end-to-end transfer delay, delay variance) with the performance requirements of the services. As an example, in Table 3 performance requirements for both standard services and IP-based services are reported.

Table 3 – Application Service Performance Requirements

APPLICATION SERVICE	STANDARDS	BER	JITTER	DELAY
Voice	G.711- G.723.1 G.727 – G.729	$< 10^{-4}$	< 3 ms	< 200 ms
Data Transmission	RFC 959	n/a	n/a	< 60 s
On – Line	RFC 1945	n/a	n/a	< 400 ms
Broadcast-Multicast	MPEG1 – MPEG2	$< 10^{-6}$	< 3 ms	n/a
Video Communication	H.320 – H.323	$< 10^{-6}$	< 3 ms	< 250 ms

Routing configuration should indicate, for example, the belonging to one or more IP-based VPN.

As the offering of **multicast** services for streaming audio and video-conferences are also foreseen, the IP transport services should support also multicast protocols. There a number of multicast protocols being developed and adopted such as MOSPF (Multicast OSPF), DVMRP (Distance Vector Multicast Routing Protocol) and PIM (Protocol Independent Multicast).

Throughput management and traffic shaping relates to the capability to be used to decrease the burstiness of UDP and TCP traffic an as such to decrease the load on traffic buffers as well as the latency jitter caused by long queues. Traffic shaping does this by identifying traffic flows and then managing the maximum transmission rates. In this sense the rate control works directly with TCP or UDP to manipulate transmission rate, instead of relying only on the existing TCP channel-capacity feedback mechanism.

Security relates to the capacity of end-users to select both secure and non-secure exchange. Support for this feature in mandatory in IP transport service based on IPv6 and optional for IPv4. In both cases the security features are implemented as extension headers that follow the main IP header. Two extension header are defined, one is designated for authentication (the Authentication Header, AH) and the other for privacy (the Encapsulating Security Payload, ESP) [RFC1825 to RFC1829].

The **provisioning time** is the time required to set-up the service.

Charging is the way the NO get paid for the provided service. Charging implies a tariff specification and an accounting policy. For an IP Transport service, the accounting policy is a still open critical issue. Yet it has to be defined its parameters and implementation.

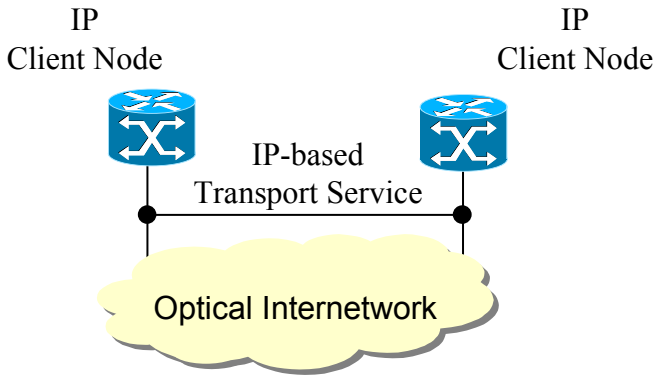


Figure 18 - Example of IP-based transport service

4 Multi–layers network models

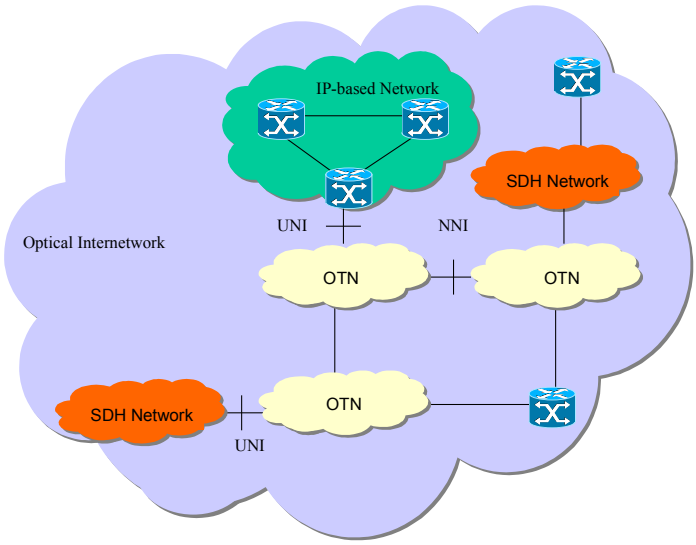


Figure 19 - Reference scenario of an optical internetwork

This section initially provides an overview of different mappings of IP over currently deployed transport networks. Then, the attention is focussed on analysing architectures of IP over

Optical Transport Networks as recognised on of the most promising solutions for Next Generation Networks.

4.1 IP–based network transmission

Classical mapping: IP/ ATM/ SDH/ fibre

Simplified mappings:

- Simplified layer 2: IP/ PPP-HDLC/ SDH/ fibre (*Packet over SONET*)
  IP/ GbE / SDH / fibre (*Ethernet over SONET*)
  IP/ SDL/ SDH/ fibre (*SDL*)
  IP/ LAPS / SDH / fibre (*ref. X.ipos*)
  IP/ MAPOS / SDH / fibre (*Packet over Lightwave*)
- Simplified layer 1: IP/ ATM (*cell-based*)/ fibre

"Intelligent" mappings: IP/ MPLS *(with ATM switches)*/ SDH/ fibre
IP/ GbE *(with GbE switches)*/ SDH/ fibre
IP/ DPT/ *(SDH frame)*/ fibre
IP/ DTM/ *(SDH frame)*/ fibre

4.1.1 The "Classical Telecom" mapping

The classical telecom mapping **IP/ ATM/ SDH/ Fibre** is:

- a viable option for the Internet world-wide extension:
  - lowing the extension of IP protocol beyond the native LAN/Ethernet environment,
  - exploiting the existing and effective networks – SDH at layer 1 and ATM at layer 2 - each one optimised to provide the functions of its layer,
- an effective solution in the present scenario, where:
  - IP is just one of the clients of the transport network (voice-centric scenario),
  - P router interfaces and links between routers (< 622 Mb/s) are not yet "very big pipes".

In this scenario, IP packets are segmented into ATM cells, assigned different Virtual Connections (switched through ATM core switches) packed into an SDH frame, according to the SDH multiplexing hierarchy, and transported through the SDH transport network. ATM provides variable granularity by the Permanent Virtual Channels using the ATM management system or by Switched Virtual Channels (SVC) dynamically set-up, all within Virtual Paths (VP). It can also use statistical multiplexing to allow certain users to access extra bandwidth for short bursts. This can help to guarantee a fixed and arbitrary bandwidth from less than 1Mb/s to several hundred Mb/s to many different customers. In addition the fine granularity can enable IP routers to be connected into a logical mesh easily, thus minimising delays from intermediate routers.

Several critical issues can be however identified if this solution is applied to the emerging network requirements:

- Entire networks are necessary at each layer:
  - Difficult end-to-end provisioning.
  - High global cost.
  - Possible multiple bottlenecks.
  - Partial overlapping of functions (e.g., protection, management functions).
  - Heavy overall overhead.

    → (1) How to simplify the "downstairs" path from IP to WDM?
- It is hard to face the exponential increase of the overall traffic:
  - Possible IP router congestion (especially in the backbone).

    → (2) How to reduce the IP Router congestion / latency delay?
- It is difficult to face the increase of the traffic granularity (bit-rate of router I/F)
  - IP Routers perform statistical "packet" multiplexing into 'high capacity pipes', through high bit-rate I/F.
  - ATM bandwidth flexibility is no longer useful on these large granularities.
  - SDH equipment is partially redundant (e.g., the use of SDH multiplexing hierarchy is no longer necessary).
  - A new "transport" service for IP is necessary.

    → (3) How to optimise this transport service?

4.1.2 Mappings simplifying Layer 2

The mappings simplifying layer 2 are based on the following steps:

- enhancement of IP Router performances:
  - higher internal throughput (from Giga to Tera Routers),
  - increased number of interfaces (to support enhanced logical mesh),
  - increased bit-rate per interface,
- addition of new functions to IP Routers,
  - statistical packet multiplexing,
  - traffic aggregation into large pipes,
  - layer 2 point-to-point data link functions (e.g., packet delineation, error correction),
- consequent simplification of layer 2, avoiding the necessity of a whole ATM network,
- necessity of an efficient underlying transport network, providing fixed bit-rate circuits (close to the peak rate, to guarantee QoS) to realise "rich mesh" connectivity between routers. Actually, SDH can offer these capabilities. The same rationale – just scaled towards higher traffic, higher bit-rates, higher throughputs – is pushing requirements for the OTN (see section 4.3).

The advantage of this solutions, with respect to classical mapping, is mainly the simplification made possible by the elimination of the ATM network. On the other hand, the fact that the SDH network is kept at layer 1 still ensures all the advantages of a transport network, amongst which protection, rich overhead for OA&M and digital performance monitoring, complete network management capabilities.

The disadvantage of this solution is mainly the difficult scalability with respect to the emerging scenarios (increase of overall traffic, traffic granularity, etc.). Disadvantages are exacerbated by the typical multi-ring topology of SDH networks: it is difficult (and expensive) to guarantee a logical mesh on a physical ring. Moreover, bandwidth efficiency is reduced by the necessity of dimensioning the interfaces and the related fixed SDH circuit to a bit-rate close to the maximum peak and not to the average traffic. This may lead to inefficiencies, in spite of the fact that, on the other hand, the weight of the overhead is fairly reduced.

Some of these solutions are listed here below:

4.1.2.1 IP / PPP-HDLC / SDH / Fibre (POS)

The basic IP over SONET uses PPP encapsulation and HDLC framing. This is also known as POS (Packet over SONET). POS (one of Cisco's solutions) places the IP layer directly above the SONET layer, and while offering QoS guarantees, eliminates the overhead needed to run IP over ATM over SONET. It is possible to transport the SDH-framed IP over an SDH transport network along with other traffic, which may then use WDM links. SDH can be used to protect IP traffic links against cable breaks by automatic protection switching (APS) in different ways.

PPP is a standardized way to encapsulate IP and other types of packets for transmission over many media from analogue phone lines to SDH, and also includes functionality to set up and close links. The HDLC framing contains delimiting flag sequences at the start and end of the frame, and also has a CRC field for error control.

The line-card in the IP router can be enhanced to perform PPP/HDLC framing. The optical signal is then suitable for transmission over optical fibre either into an SDH network element, an adjacent IP router, or a WDM transponder. There are different types of IP over SDH interfaces:

- VC4 or Concatenated VC4 "fat pipes" which provide aggregate bandwidth without any partitions between different IP services which may exist within the packet stream.
- Channelized interfaces, where an STM16 optical output may contain 16 individual VC4s, with a possible service separation for each VC4.

This solution is already implemented by some of the major IP carriers:

- Sprint is deploying packet over SONET to boost its Internet backbone speed to 622 Mbps.
- Qwest's OC-192 SONET network is using packet over SONET to connect its regional terabit points of presence (TeraPOPs), which are essentially IP-based central offices (COs).
- UUNET's recently announced OC-12 network will use packet over SONET technology.
- GTE Internetw. is deploying Cisco 12000 routers to operate over its nationwide mesh of OC-3 SONET circuits.
- Many European operators have introduced SDH elements with add/drop bit rate up to 140 Mb/s or STM1. In order to support POS, these SDH elements should be upgraded to have the STM-4c and STM-16c (concatenated) interface.

4.1.2.2 IP / GbE / SDH / Fibre (Ethernet over SONET)

A large part of native IP traffic comes from Ethernet/Gigabit Ethernet LAN (Ethernet accounts for over 85% of LANs world-wide). The new Gigabit Ethernet standard can be used to extend high-capacity LANs to MANs and maybe even WANs, using Gigabit line-cards on IP routers, which can cost 5 times less than SDH line-cards with similar capacity. For this reason, Gigabit Ethernet could be a very attractive means to transport IP over "metropolitan" WDM rings, or even over longer WDM links. Furthermore, 10Gbit/s Ethernet ports are likely to be standardized in the near future.

Ethernet is both a layer 1 physical interface and the most diffused layer 2 protocol. It can be used:

- on the LAN/access side, for layer 1 and layer 2 functions (e.g., access to shared resources)
- on the WAN side, for layer 2 data link functions, while it is interfaced to the SDH/SONET physical layer for transport functions (transmission over medium/long distances, protection/ restoration, fault and performance management, etc.). In this case, the Ethernet layering is valid at level 2 for MAC (the 8B/10B encoding may be avoided). On the contrary, the physical Ethernet 802.3 layer is substituted by SDH/SONET. It should be recalled that when Gigabit Ethernet (1000Base-X) is used in full-duplex mode, it becomes just an encapsulation and framing method for IP packets, and the CSMA-CD functionality is not used.

This solution may assume a particular interest if Ethernet switching is enabled (see section 4.1.4). Some options for the mappings Data (IP) over Ethernet and Ethernet over SONET/SDH have been proposed by Nortel (e.g., in ANSI T1X1.5).

4.1.2.3 IP / SDL / SDH / Fibre (SDL)

Simple Data Link (SDL) is a framing method proposed by Lucent Technologies, which can replace HDLC framing for PPP-encapsulated packets. It is simpler than HDLC and advantageous at high bit rates. Compared with the HDLC frame the SDL frame has no delimiting flag sequences. Instead the SDL frame is started with a packet length field. This is advantageous at high bit rates where synchronisation with the flag sequence is difficult. The SDL format can be inserted as a payload in an SDH frame or directly onto a WDM optical channel. SDL associated functionality is the minimum required. It does not support protection functionality.

4.1.2.4 IP / LAPS / SDH / Fibre (X.ipos)

LAPS (Link Access Procedure SDH) is a type of HDLC including data link service and protocol specification to adapt directly IP to SDH. It is a substitute for PPP-HDLC. SDH is seen as an octet-oriented synchronous point-to-point links. The SDH frames are an octet-oriented synchronous multiplex mapping structure which specifies a series of standard rates, formats and mapping method. The LAPS shall be encapsulated into a 32-bit word-oriented frame as for the need of providing frame delineation, transparent transferring, and error detection.

4.1.2.5 IP / MAPOS / SDH / Fibre (POL = Packet Over Lightwave)

Multiple Access Protocol over SONET (MAPOS): it is a link layer protocol supporting IP over SDH. It is a connectionless packet switching protocol based on a simple extension of POS framing.

4.1.3 Mappings simplifying Layer 1

The mappings simplifying layer 1 are based on the following rationale:

- exploitation of layer 2 switching,
  - traffic by-pass at layer 2,
  - packets are segmented / re-assembled in cells, and handled as virtual circuits,
- use of ATM as transport infrastructure,
- consequent simplification of layer 1 - avoiding a further underlying transport network, possibly using layer 1 standardized physical interfaces.

Although ATM can be an option for this approach, there are severe limitations related to the max. capacity that can be served and the max. granularity which can be offered. This solution is good when the bit-rates of the interconnection pipes are relatively low (155 Mb/s or less) but not suitable to the emerging scenario for the core backbone. It is mentioned here just for completeness.

4.1.3.1 IP / ATM (Cell Based) / Fibre

It is possible to have a scenario where ATM cells are transported directly on an optical channel. ATM cells are not encapsulated into SDH frames; instead, they are sent directly on the physical medium by using an ATM cell-based physical layer. Cell-based physical mechanisms have been developed specifically to carry the ATM protocol; this technique can not support any other protocol except if these protocols are emulated over ATM. Some benefits of using a cell-based interface instead of SDH are:

- Simple transmission technique: ATM cells are directly sent over the physical medium after scrambling.
- Lower physical layer overhead (around 16 times lower than SDH).
- As ATM is asynchronous, there is no stringent timing mechanism to be put on the network.

The drawbacks are that the overhead (i.e., the cell tax) is the same as for transport on SDH, the technology has not been endorsed by the industry yet and this transmission technique can only carry ATM cells.

4.1.4 "Intelligent" Mappings

In order to answer the second issue raised in section 4.1.1 - (2) how to reduce the IP Router congestion/latency delay? – the following approaches can be adopted:

- Traffic by-pass at lower layers: typically, exploitation of layer 2 switching.
- Enhancement of the logical connectivity seen by IP Routers, through improvement of the physical connectivity at lower layers.

- Introduction of "interworking" at control plane level.

Several interesting new technologies are emerging, but, for the moment, as single wavelength solutions limited to metro-net scenarios and not yet demonstrated for very high capacity backbone. Moreover, these technologies are not yet ready to face the remarkable increase of traffic granularity currently required.

Some solutions are here just mentioned (since they are dealt with in other contributions to the same WP 1 – Task 2):

4.1.4.1 IP / MPLS (with ATM switches) / SDH / Fibre

This is a successful solution, emerged from the "IP standardization world". It is la label switching, driven by the IP control plane, and actuated by ATM switches. It allows to by-pass the express traffic at layer 2, giving relief to the layer 3 routers. A whole part of the WP1 Task 2 document is devoted to this technology.

4.1.4.2 IP / GbE (with GbE switches) / SDH / Fibre

The idea is that of exploiting Ethernet switching at each access node to a ring. In this way, the flow of transit packets is distinguished from the flow of packets destined to (or coming from) the local IP Router. A layer 2 by-pass is performed, and the sharing of a common resource (ring big-pipe) is allowed (up to the 1.25 Gb/s allowed by Gigabit Ethernet) achieving all the benefits of a statistical multiplexing applied by means of packet switching.

To obtain these advantages, it is necessary to change (and enhance) the Ethernet MAC through proper SRP protocols and fairness algorithms. Different physical options can be implemented (according to whether Ethernet switch cards are integrated in the layer 3 router or in the layer 1 node, or in a stand-alone further node).

An example of product proposal is the interWAN Packet Transport (iPT), by Nortel, which is claimed to deliver lower cost IP-optimized networking capabilities to today's wide area networks over existing SONET/SDH and DWDM optical networks. IPT essentially multiplexes and switches packet traffic in its native format over SONET/SDH platforms. The benefits are optimized use of transport network bandwidth, reduced port costs, and the ability to support a mix of packet and TDM traffic over the same network. With Ethernet add/drop and switching capabilities at each node, interWAN Packet Transport optimizes the optical network usage through packet multiplexing and through a unique spatial reuse algorithm. Instead of dedicating a number of links for each connection, iPT provides logical connectivity for optimum bandwidth sharing.

4.1.4.3 IP / DPT / (SDH frame) / fibre

Dynamic Packet Transport – DPT (Cisco proprietary). It is associated to a transport ring topology. It provides the Spatial Reuse Protocol (SRP) to optimize bandwidth optimization and a protection mechanism called Intelligent Protection Switching (IPS). It is based on a further enhancement of IP Router functions: traffic by-pass, queue handling at layer 2, statistical "packet" multiplexing on shared pipes, protection features. This approach not only avoids the use of ATM, but also of SDH (only SDH interfaces are kept).

4.1.4.4 IP / DTM (with DTM switches) / (SDH frame) / fibre

DTM is an effort to combine the advantages of asynchronous and synchronous data transfer. It is a TDM scheme, which guarantees each host a certain bandwidth and uses a large fraction of the available bandwidth for effective data transfer. The DTM scheme has in

common with ATM, support for dynamic reallocation of bandwidth between hosts. This means that the network can adapt to variations in the traffic and divide its bandwidth between hosts according to their demand.

Hosts connected to a DTM network communicate with each other on channels (circuits). A DTM channel is a dynamic resource that can be set up with a bandwidth ranging from 512 kb/s in quantum steps of 512 kb/s up to the maximum bandwidth. The total capacity is divided into frames of 125 microseconds, which are further divided into 64-bit time slots. This framing structure makes it interoperable with SDH/SONET. Different types of slot reservation can be assigned according to the QoS a client wants: constant delay, minimum bandwidth, best effort.

To interconnect different DTM links, DTM switches should be used. The switching in DTM is synchronous, which means that the switching delay is constant for a channel.

DTM channels are multicast in nature: any channel at a given time can occupy one sender and any number of receivers. DTM is suited as a backbone technology because of its high bit rate throughput. DTM is seen as an alternative for ATM over SDH because it operates at layer 1 to 3 and includes switching and a signaling protocol.

Flows of IP can be mapped on DTM channels. However, DTM is slightly inefficient for IP since it uses channels with a minimum capacity of 512 kb/s.

DTM has enough switch capacity to handle WDM. Hereby the basic assumption is that one WDM colour will go on one DTM channel, which is only possible when the transmission method is supported by DTM.

A drawback of DTM is that currently only three vendors are supporting DTM, Dynarc, Net Insight and Ericsson. Moreover, their solutions are proprietary and not compatible with each other.

4.2 IP–based networks over WDM

All the mapping solutions mentioned in chapter 4.1 may be ultimately carried by WDM, rather than being transmitted through a dark fibre. This is advantageous, since very high-capacity WDM systems have already been deployed (especially in the backbone transport). Moreover, this is becoming strictly necessary, because WDM is the only technology whose capacity can support the traffic growth related to the new data-centric network.

More specifically:

- All the mapping solutions including SDH as layer 1 may be supported by the WDM transmission systems already installed for the long-haul SDH network. What is necessary is simply a transponder between the SDH STM-16/STM-64 termination and the "coloured" interface towards the WDM line system.
  - IP/ ATM/ SDH/ WDM
  - IP/ PPP-HDLC/ SDH/ WDM
  - IP/ GbE / SDH / WDM
  - IP/ SDL/ SDH/ WDM
  - IP/ LAPS / SDH / WDM
  - IP/ MAPOS/ SDH/ WDM
  - IP/ MPLS (*with ATM switches*)/ SDH/ WDM
- Even the other mappings – which do not include SDH – can be transported by WDM, through proper transponders between the router/switch and the ends of the WDM line

system. Once more, this is done just to increase the capacity of the links, without changing the working principle of the adopted technique.

- IP/ ATM (cell-based)/ WDM
- IP/ DPT/ (SDH frame)/ WDM
- IP/ DTM/ (SDH frame)/ WDM

Therefore, "indirect" IP over WDM is already a reality, especially in backbone network segments. The huge capacity increase allowed by WDM, although not involving advanced network aspects, is by itself an unavoidable pre-requisite for all the consequent progresses, and should not be under-evaluated.

However, this is not yet "IP directly over WDM": layer 2 and layer 1 functions (and often entire layer 2 and layer 1 networks) are still necessary in between. In other words, WDM plays here the role of a "layer 0".

In general, a layer 2 network might be useful to allow:

- fine granularity bandwidth management,
- differentiation of QoS on different flows,
- traffic by-pass at layer 2,
- point-to-point layer 2 functions (packet delineation, error correction, etc.).

In general a layer 1 network might be useful to provide:

- high-bit-rate interconnections, through high capacity point-to-point links,
- advanced protection/restoration functions,
- performance, fault, configuration management on connections,
- traffic by-pass at layer 1.

For clarity's purpose, the following table is proposed:

Table 4: Overview on different interpretations of the layering concepts

OSI layer	**#**	**Function**	Handled Entities	**Network**	**Ref. Technology**
network	3	Routing	Packets	Internet	IP
data link	2	Switching	cells, virtual circuits	Switching Network	ATM
physical	1	cross-connection	"TDM" circuits	Transport Network	SDH
-	0	high-capacity transmission	wavelengths	-	WDM

4.2.1 Enabling technologies

WDM technologies for line systems are mature and effective. Nonetheless, new technologies are emerging to allow further performance improvements. A quick overview of the main optical components of a WDM line system and of the related enabling technologies is here recalled.

Table 5: Overview on different interpretations of the layering concepts

Functional Block	Consolidated Technology	Advanced Technology
Source	Distributed FeedBack (DFB) Lasers Distributed Bragg Reflector (DBR) Lasers	SSG (Super-Sampled Grating) DBR for wide tunability Vertical Cavity Surface Emitting Lasers (VCSEL) for low-cost arrays
Intens. Modulator	Mach Zehnder interferometer in $LiNbO_3$	Integrated semiconductor Electro-Absorption modul.
Receiver Photo-Detector	PIN photodetectors (up to 10 Gb/s ch.) Avalanche-Photo-Diode - APD (up to 2.5 Gb/s ch.)	Avalanche Photo-Diodes – APD (also for 10 Gb/s ch.)
WDM Mux/Demux	micro-optics diffraction gratings cascade of interferential filters cascade of Fibre Bragg Gratings (FBG)	Arrayed WaveGuides (AWG), built as Planar Lightwave Circuit (PLC)
Optical Amplifier	Erbium-Doped Fibre Amplifiers (EDFA) (optimized for multi-channel WDM, also through external control gain systems)	Different dopings for extended bandwidth (e.g. Erbium-Doped Fibre Fluoride Amplifiers –EDFFA): the objective is to have two usable bandwidths, C and L, over a whole range between 1530 and 1620 nm
Dispersion Compensation Module	Dispersion Compensating Fibre - DCF (on multi-channel WDM) Chirped Fibre Grating (single channel)	Chirped Fibre Grating (also on multi-channel WDM)

4.3 IP–based networks over OTN

4.3.1 The role of a Transport Network

The advent of packet-switched data-centric networking does not eliminate the necessity of a transport network. In the rush to get to service convergence, it is easy to neglect the fundamental value of transport networking. This is a mistake, because transport networking continues to be essential for real-world networks:

- to provide/manage high bit-rate interconnection pipes between Giga/Tera-Routers in the core backbone; transport networking, increasingly based on the optical layer technology, continues to be essential, because broadband data networks will need more transport functionality than just point-to-point "big dumb pipe" interconnects between routers and switches,
- to establish a unifying infrastructure for multiple-service layers providing large-granularity bandwidth management,
- to support the world-wide extension of the Internet, where the network usage patterns are completely changing. The old rule was that network traffic was 80 % local, and 20 % wide area. Today, the ratio is closer to 50/50, and the growing number of applications that require all communication from clients to traverse the backbone to reach central servers means more stress on network backbones. It is expected that the 80/20 rule will eventually return, but with wide-area applications accounting for 80 % of traffic,
- to achieve, on the whole, a lower-cost network.

A transport network ensures several important functionalities, among which:

- capability of traffic by-pass at layer 1:

this implies the availability of more high-capacity circuits for improved connectivity and less "hops" for end-to-end paths, the reduction of congestion and latency time. On the contrary, without a transport network, the mesh between routers (through point-to-point links) may be relatively poor, so that a large fraction of the traffic traversing each router is "through" traffic,

- scalability for the service layer:

  without an efficient transport network, it may be necessary to expand the entire network to face sudden spikes in demand for a sub-set of client nodes, even the intermediate service-layer nodes that provide only transport-like bandwidth management. The problem is that the service layer logical topology is tied to the network physical link topology. Transport networking provides a solution, by freeing the service-layer logical topology from the physical link topology, and by implementing networking and bandwidth management in a most efficient way:

  - upgrade only the routers where spikes occur,
  - provide a new high-capacity end-to-end circuit
  - keep the rest of the network unchanged

- fast survivability and support of service-layer restoration with shared-protection architectures:

  IP can achieve survivability by its re-routing algorithms, but when each fibre carries a huge traffic, the convergence time of IP re-routing may be very slow. Network survivability/reliability improves if large aggregates of traffic are recovered by fast layer 1 protection/restoration (at least in case of fibre breaks).

Today's transport network is SDH/SONET (already mentioned as "layer 1 reference technology") which exploits WDM as a transmission support. SDH is an excellent solution under many aspects:

- Layered network functional modeling.
- Managed Network (TMN).
- Rich "overhead" for Operation & Maintenance purposes.
- Advanced protection schemes supported (e.g., Optical Multiplex Section Shared Protection Ring - OMSPRing).

However, it is optimised for "voice / PSTN primary clients", not as a "data-centric" network.

The data-centric network operates in a fundamentally different way from the voice-centric network. Aggregation of physical circuits tends to happen at the edge of the carrier network. The speed of the transmission facilities between switches directly affects the performance of these networks. This drives the evolution towards routers and switches characterised by higher and higher capacity interfaces. The newest IP routers and ATM switches for the core are scaling to Tb/s throughputs, with 2.5 and 10 Gb/s ports to match. The rapid emergence of high bit-rate ports on routers and switches is probably the single most disruptive force in today's transport networks. The data-service layers are suddenly scaling faster than SDH/SONET.

When OC-192c/STM-64c interfaces are used, the current transport network architectures would call for OC-768/STM-256 ADMs; alternatively, for DXCs with OC-192c/STM-64c interfaces and an accompanying switch fabric. These upgrades will not only be expensive, but they also will not be readily available in the near future.

Another possible answer by SDH is the "overlay rings" architecture. However, from an operational standpoint, the overlay model is complex because each overlay ring must be

managed individually. This introduces the issue of selecting the correct ring to provision services, which becomes a challenging exercise in large networks. Provisioning across rings adds even more complexity because ring selection must be performed across multiple rings. To provision capacity in today's network, there typically are a number of steps and different organisations involved.

On the other hand, the availability of OC-48c interfaces on a router/switch (and the subtended aggregation function) may eliminate the need for an entire layer of TDM multiplexing. There is little benefit to connecting a switch OC-48c output to an OC-48 SONET multiplexer, because the port speeds have already reached the interface speed of the optical layer.

The obvious consequent solution is a "direct" connection of routers/switches to the already deployed and/or emerging WDM systems (described in chapter 4.2), taking advantage also of the fair matching between the mentioned IP Gigarouters interfaces and the typical bit-rate modulation of wavelength channels in WDM. The number of OC-48 channels made possible by DWDM seems to be ideally suited to broadband IP routers. In this case, however, traffic engineering (for provisioning and protection) and more generally all the function of a transport network should be performed by the WDM layer on the wavelength granularity.

To summarize, the future Core Transport Network will be required to offer a substantially increased overall capacity and a flexible provision of long-reach high bit-rate pipes:

- to meet the huge overall traffic growth,
- to answer the new traffic distribution paradigm (80% long reach),
- to serve the future main client: IP Giga/Tera Router with several high-capacity ports,
- to allow switching at a higher level of granularity.

The Optical Transport Network is actually the only candidate solution to fulfil these requirements.

4.3.2 The "IP over OTN" scenario

As mentioned due to DWDM technology, an OTN can provide a large amount of raw bandwidth supporting the delivery of big volumes of IP traffic. Particularly:

- IP Core Routers will continue to do what they can do well:
  - routing, forwarding, packet switching and aggregation are performed by IP Routers,
  - "point to point" layer 2 (data link) functions are added to IP Routers,
  - high bit-rate interfaces SDH- or WDM- compliant are included in the IP Routers.
- WDM supports IP by offering it the functions of a more advanced transport network:
  - cross-connection on the new wavelength granularity,
  - transport network (layer 1) functions: protection/restoration, monitoring, circuit provisioning, management and supervision – again, on the wavelength granularity.

The high bit rate data pipe – internally organised as a flow of aggregated and forwarded IP packets, with a given priority – is mapped into one wavelength, which is the new "managed entity" in the new OTN transport network. Actually, IP nodes extends their functions to layer 2 (at least point-to-point data link functions) whereas WDM extends its functions from layer 0 (transmission support) to layer 1 (transport network), thus becoming OTN.

The former "transmission support" and "transport network" are substituted by the new "OTN", realising the evolution from WDM to OTN.

With respect to the emerging data-centric backbone network, the situation appear as follows:

(3) how to optimise the transport service for the IP core backbone?

- OTN is the only solution.

(2) how to reduce congestion/latency delays in core IP Routers?

- It is possible to develop solutions for the traffic by-pass at layers lower than layer 3, driven by the layer 3 control plane (e.g., MPLS) – see section 4.1.4.1.
- OTN can be a solution, in that it allows traffic by-pass at level 1, on wavelength granularity.

(1) how to simplify the downstairs path from IP to WDM?

- several different "simplified mappings" have,
- OTN can be a solution, in that it may substitute SDH as transport layer.

This means that the evolution is from a conflicting to a converging telecom and datacom view:

- telecom view: IP is just one more client (the risk of this view is that of under-evaluating the emerging data-centric network requirements),
- datacom view: the IP router is the multi-service aggregation node, fulcrum of the new data-centric network (the risk of this view is that of under-evaluating the necessity and the role of the underlying transport network),
- converging view: the best network architecture for a cost-effective, reliable, scalable evolution employs both transport networking and enhanced service layers, working together in a complementary and interoperable way.

A summary of properties and advantages of the IP over OTN solution are summarised here below:

- Optimisation of the data transport plane, so that:
  - It is ready to serve the rapidly increasing long reach data traffic.
  - It allows the implementation of the "unlimited bandwidth" scenario imagined in forecasts.
  - It respects the evolving IP network (no major change required in the IP world).
  - It is a platform for the wide area extension of the network, able to accept different physical interfaces, e.g.,
    - SDH/Sonet OC-48/STM-16, OC-192/STM-64
    - SDH/Sonet compliant (DPT, EoS)
    - ATM 622 Mb/s
    - GbE – 10 GbE
  - It is consistent with the development of Gigabit Routers as the main backbone nodes

    For example: the Cisco 12016 GSR Terabit System offers modular scalability to 5 Tb/s by integrating up to sixteen 12016 GSR nodes using the GSR Scalability Module and the GSR Terabit Fabric interconnect. The system will support 256 slots and the entire range of existing and future GSR line cards from DS-3 to OC-192c/STM-64c. In particular, Cisco claims it has delivered thousands of OC-48/STM-16 interfaces for

the Cisco 12000 series gigabit switch router (GSR) to large IP network operators throughout the world. The vast majority of these interfaces connect directly to DWDM systems, making this a widely accepted approach to building high-capacity IP backbones.

Another example is the "Juniper M20 Backbone Router", whose architecture delivers a 4xOC-48/STM-16 router that runs in the most traffic-intensive parts of the Internet.

- Improvement of the logical connectivity for the Gigarouters:
  - enriched meshed connectivity at layer 1,
  - traffic by-pass at layer 1,
  - possibility of avoiding a layer 2 network,
- Improvement of the scalability for the service layers.
- Simplification of the network layering:
  - the necessity of an entire ATM (more generally, a layer 2) network is avoided,
  - the necessity of an entire SDH network with the related equipment is avoided,
  - the evolution is from TDM Voice Infrastructure (hierarchical model, cross-connection through 64-kbps crosspoints, capacities limited by TDM, restoration via full duplication) to a Data-Optimised Infrastructure (flattened network model, statistical multiplexing, dynamic usage characteristics, intelligent protection switching, WDM capacities),
  - keeping the already standardized SDH interfaces, when opportune – evolving from OC-48 (channelized) optimized for TDM Switching to OC-48c (concatenated) enabling Optical Internetworking,
  - the new mapping is: IP / layer 2 point-to-point functions / layer 1 interfaces / OTN.

Thus potentially achieving lower cost, complexity, overhead, and improved scalability.

- Simplification of the network structure, i.e., substantial reduction of the number of separate physical devices.
- Capability of switching at the right granularity.

  Historically, the increase of higher level switching granularity has been in proportion to the increase of the overall link capacity (between one and two orders of magnitude below). It is clear that a switching granularity of the order of magnitude of 2.5 Gb/s is becoming necessary in the new scenario. This is perfectly consistent with the wavelength-granularity switching – which is on the other hand the only practical solution to achieve this objective
- Potential to become a unifying transport layer.

  Once deployed for the new IP main client, the OTN may substitute the SDH network and accept the existing "SDH voice-centric network" just as a client. The evolution will be smooth:
  - IP will continue to be carried by ATM up to e.g., OC-3 capacity.
  - ATM and voice will continue to be carried by SDH up to e.g., OC-12.
  - SDH and IP will be direct client of OTN at OC-48 and above.

Rather than disappearing, SDH/SONET will transition through this evolution. SDH/SONET will continue to be competitive for lower-capacity transport, especially toward the network edge. And it will enjoy wide acceptance as a stable, interoperable format for multi-gigabit data-service ports.

5 Interworking Functionality

As data is the fastest growing segment of network traffic, transport network models are likely to evolve to data-centric solutions. Furthermore as the OTN can provide the large amount of raw bandwidth supporting the increasing data traffic, in the short term a client-independent OTN is likely to be the missing link between legacy (e.g. SDH) and data centric networks.

As a consequence Network and Service Providers (NSP) are strategically moving toward a single integrated voice and data infrastructure where IP is gaining the role of integration layer for multiple services. In this context, it is a foregone conclusion that the interworking between routers and optical network elements (e.g. OADM and OXC) will play a pre-eminent role in converged future networks.

Nevertheless NSP that build a multi-service IP network are going to need connectivity to its pre-existing legacy networks. The client-independence of the OTN will guarantee a smooth evolution from legacy over OTN to a data-centric OTN, both coexisting. As such, if the first step is the introduction of optical network elements (ONE) based on WDM, supporting also the transport of legacy clients, the immediately following one seems to be the interworking of ONE with IP routers (data-centric OTN).

From the controlling viewpoint, in the short-middle term the network model is likely to be client-server where the OTN will maintain its own control plane. In the middle-long term even the evolution towards peer-to-peer model could be considered.

Historically Network Operators have used several layers to build their transport networks: adopting for example IP routers over ATM switches over PDH or SDH network elements. As a consequence the set of networks functionality across the several layers needs to be optimised reducing potential duplications. For example, survivability in multi-layer networks is achieved by providing recovery mechanisms in multiple layers. In order to provide a proper working and to avoid an enormous waste of spare resources, it is required to co-ordinate these recovery mechanisms.

In this section, a list of network functionality required by an integrated multi-layer transport network over OTN, supporting envisaged transport services [WP1-M1]. Finally, some guidelines are provided to define the interworking functionality to be demonstrated in the LION Test-bed.

5.1 General transport functionality

A generic IP-based network over OTN should provide the following main functionality:

- Frame forwarding: the functionality that processes inbound traffic and forward this traffic to the appropriate outbound destination link. Frame forwarding can operate at different layers of the protocol stack:
- Layer 1 switching.

- Layer 2 switches (mainly for LAN) provide frame forwarding based on link layer information such as MAC address.
- Layer 3 switches and routers forward frames based on layer 3 address (e.g. IP address).
- Also Layer-4 switching could be required: this functionality permits the network to differentiate the way it treats network traffic by type of application. For example traffic for critical applications can be assigned different forwarding rules than HTTP-based Web traffic even if transmitted across the same set or router interfaces.
- Route-Path calculation: involve the identification of optimal routes (at layer 3, 2 and 1) through the network. For example, path and route calculation, operating in a router at layer 3 (RIP and OSPF), allow efficient decisions to be made for the transmission of traffic between nodes of the network.
- Multiple service quality classes to deliver to customers.
- Multicasting: as NSP in the future are expected to offer multicast services for streaming audio and video, multicast protocols and capabilities should be supported. Depending on the network topology and status, multicast functionality may be implemented at different layer levels. This also depends on the granularity of the flow to be distributed.
- Traffic Shaping: this is, for example, to decrease the burstiness of UDP and TCP traffic, thereby decreasing the load on the router buffers as well as the latency jitter caused by long queues. Traffic shaping does this by identifying traffic flows and then managing the maximum transmission rate.
- Bandwidth reservation: there are some applications such as interactive video and voice that may require limited network delays or bounded delay variation. For this type of traffic, bandwidth reservation may be required.
- Congestion control
- Restored IP connectivity
- Set-up and tear down of Optical Channel Connections
- Restored Optical Channel Connectivity
- Optical Performance Monitoring
- Regeneration of Optical Channels
- Added value functionality: it allows the dynamic definition of a rule set for operating and administrating the network. A centrally managed configuration of a wide range of network policies and service classes across switches, routers and optical network elements.

5.2 Network Functionality for Layers Inter-Working

In a multi-layer network the generic network functions may exist in each network layer. This duplication may produce different kind of conflicts and inefficiencies. Therefore, a key issue is to define how to distribute the functionality across layers, or how to implement functionality processes for interworking between the different layers. In this section those network functionality for which layer inter-working is required are preliminary selected.

In particular the attention will be focussed on the set-up and tear down of OCh Connections from a client network (e.g. IP) as one of the innovative features (potentially to be implemented in the LION test-bed) provided by a mdata-centric OTN accordingly to the ASON model.

The idea of a new converged network architecture, such as IP over OTN, which utilises routers or switches connected to an optical networking infrastructure opens the issues of how

to distribute this functionality across technologies and how to optimise the layer inter-working.

We have identified four main functionality categories requiring multiple-layer inter-working:

- Dynamic Configuration of Connections
- Bandwidth Provisioning
- Performance Monitoring
- Multi-Layer Survivability

5.2.1 Dynamic Configuration of Connections

Any client signal from incoming set-up service request should be managed by the server layer in order to find its route from end to end automatically, using the corresponding OA&M messaging and activating the necessary network resources; similarly, releasing them when finishing. These switching functions are to be automatic considering the great number of different flows in real time crossing the network and the desired fast service, and also dynamic, adapting to the variations of these flows. Both central and distributed connection set up should be considered.

To obtain a desirable level of flexibility, configuration celerity and efficient resource utilisation, dynamic traffic engineering algorithms for traffic grooming, flexible re-configuration, etc. must be applied.

The switch-router communicates with the optical network element via a logical CNI (ie UNI) interface. This interface defines a set of primitives to configure the optical network element and to convey information from the optical network element to the switch-router.

Given that both IP and OTN layers require control planes, two options arise concerning their inter-working [Jamousi]:

One option is to have two separate independent control planes. However, both planes have relatively similar requirements.

An alternative approach is to have an integrated control plane.

Therefore, two models can be adopted: 1) a single control plane that can be used by the IP control plane to establish OCh connections in the OTN layer, 2) two different control planes exchanging information through the CNI.

The approach adopted by the LION Project is to consider and experiment features of an OTN that has its own independent control plane (model 2). Nevertheless, model 1 is seen as a middle-long term scenarios for which theoretical investigations are seen within the Project.

The network scenario considered in the following part of this section is based on two independent IP and OTN control planes exchanging information across the CNI interface using a standard protocol.

As an example of interworking functionality, the "Dynamic Configuration of Connections" is considered. This functionality can be fractionated in a number of inter-related specific functions at the different layers that will be triggered successively to adapt to the instantaneous connection and bandwidth demands. These specific functions are listed below allocated in the different network layers, for an IP over MPLS over OTN scenario:

IP layer

N_Connect

N_Data

N_Reset

N_Disconnect

MPLS

LSP establishment

OTN layer

OCh Connection Set-up

OCh_route discovery

Create OCh_trail termination point

Set_up point_to_point OCh_trail

MPλS functionality

Optical Channel trail establishment

Alarm indications

OCh_Data

OCh_Disconnect

OCh_Bridge

OCh_Switch

The fuctions at the IP and MPLS level are already well known and deployed. The attention will be focussed on the specific functions at the OTN level. Particularly, the "OCh Connection Set-Up" function is taken as a preliminary example.

5.2.1.1 OCh Connection Set-up

In the dynamic OCh Connection Set-up the higher layer dynamically requests the establishment of OTN network connections, that also may lead to the dynamic need and reservation of optical channel path for subsequent data communication [ITU-G872], [Mephisto].

This ASON functionality is triggered by the upper layers and it leads to the interaction between layers, mainly with signalling.

Considering the directionality of the connections, the OTN can provide service for

Point-to-point unidirectional OCh connection.

Point-to-point bidirectional OCh connection.

Point-to-multipoint unidirectional OCh connection.

Path establishment service will also have to meet physical (e.g. OSNR) and logical limitations (e.g. MTBF).

Therefore, the service establishment in the OTN layer has to comply with the following points:

a) Identify a path (or a path sub-set) that meets service requirements like physical and logical requirements.

b) Limit as much as possible resource allocation (e.g. not use wavelength converters when useless).

c) Preserve as much as possible further service establishment capabilities.

The introduction of an underlying WDM network as service provider adds new flexibility to interconnect client nodes (IP GSR for the LION test-bed). To fully benefit from these, the client network operations system should be aware of all possible OCh connections between client nodes (GSR's). Consequently, the OTN operations system should provide this information.

When the overlay model of the control plane is considered, three specific processes can be defined in this functionality:

OCh route discovery

This function is aimed at identifying routes for trails and/or sub-network connections. It is also susceptible to be used for protection purposes. Selection of route/s will be done following certain criteria, some of which is particular to the case of the optical routes. It may apply traffic-engineering algorithms like ER-MPλS to wavelength-label-driven routing and use routing tables of OXC and OADM to find the best OCh route considering specific criteria.

The necessary parameters identifying a requested connection are:

- Router/port/client destination address
- Router/port/client source address
- Bit-rate
- Priority
- Availability level required
- Protection level required
- Security level required ?

This information needs to be adequately pre-defined to have a common compliance over the whole network. Correspondingly:

- Address scheme for OCh end points
- Naming schemes for OCh clients
- Scheme for specifying protection needs in connection requests
- Security parameters
- Provisioning time

Prior to selecting routes for trails or sub-network connections in the OCh layer, some parameters have to be measured/calculated. These parameters are the overall Chromatic Dispersion, the overall Polarisation Mode Dispersion (PMD) and the OSNR in every transport entity. The number of crossed Wavelength Converters is also limited since they add jitter to the optical signal. It is necessary to verify that the overall values are acceptable to the candidate routes for each wavelength (i.e. to verify that the system limit is not exceeded).

These calculations can be converted to Available Bandwidth before being used by the upper control planes when searching for adequate routes.

Create OCh_trail termination point

This function has been taken from ITU-T Rec. G.852.6. It creates a trail termination point. There is also the possibility of associating the trail termination point with an access group or a sub-network.

Set_up point_to_point OCh_trail

This function has been taken from ITU-T Rec. G.852.6. It associates two identified trail termination points or access groups in the OCh layer. Performs resources reservations along the selected path by the previous function, and correspondingly sets the OTN NE switches.

When considering the peer-to-peer or augmented model of the control plane, other processes can be defined.

MPλS functions

When MPLS functionality is applied at the OTN, a new protocol is derived. This customisation of MPLS is being done by IETF and it is named Multi-protocol Lambda Switching [Ghani 2000]. The main objective of MPλS is to have an efficient routing/switching technique with a fast protection/restoration at layer one.

Considering an IP network which is MPLS capable (MPLS domain), the configuration of the IP routers means to establish a Label Switched Path (LSP), what is considered that the IP routers are performing layer two switching. As stated before, adopting an MPLS technique means to collapse layer two and three functions.

Considering MPλS further investigations are required to reach an understanding on the control plane architectures and interworking.

Optical Channel trail establishment

Enhanced Interior Gateway Protocol (IGP) versions will have to distribute the State Information of the OTN including topology state information [Awduche]. This information will be used by the "route discovery" system to find corresponding paths for point-to-point OCh trails based on some constraints. The distributed information will include:

- OTN Topology
- Available bandwidth.
- Available channels per fibre.
- OTN topology State Information.

The control plane needs an MPLS signalling protocol to create a point-to-point OCh trail between identified trail termination points in the OTN. The control plane can use an out-of-band IP channel, or a dedicated supervisory optical channel (i.e., dedicated wavelength).

Hence a critical issue is the development of an OTN domain specific model to describe and abstract its relevant characteristics. This model could be divided into Network Element Level (NEL) and Network Level (NL). The NL model could be used to find an appropriate OCh trail from the client control plane, and the NEL model is primarily used to update the information in the NL model (other applications are out of the scope of this deliverable).

The OTN model is also susceptible to be used for the overlay model of the control plane to perform Traffic Engineering processes in the optical layer.

Another important question is scalability for which control plane partitioning should be possible. Then, each network partition might maintain a Link State Database containing information about fibre links, OCh trails and logical paths.

Other generic processes can also be defined for either control plane model.

Alarm indications

Message sent from the optical network element to the router to indicate that a failure has been detected.

OCh_Data

To begin a specific data stream transmission over the established OCh.

OCh_Disconnect

Release all the resources reservations of the path to tear down an Optical Channel Connection. This is a command sent from the router to the optical network element to disconnect an input port to an output port.

OCh_Bridge

This is a command sent from the router to the optical network element to bridge a connected input wavelength to another output port, triggered by a request at the upper layer.

OCh_Switch

This is a command sent from the router to the optical network element to tear down an OCh connection and immediately set up another one.

5.2.2 Bandwidth Provisioning

A Network Service Provider must be able to set up a specific trail at the request by the client, with a pre-negotiated QoS characteristics in terms of bandwidth, availability, quality of signal, delay, encryption, etc, according to the particular needs of the end-user or service.

The network operator can also see this functionality as a leased/private line service at customer request.

An innovative aspect here may consist in the direct and instant provisioning via a Web Interface by customer, performing reservations on-line, being billed only for the reserved bandwidth at each time.

This optical channel setting up by the customer on demand performs the processes of resources reservation making use of specific signalling and routing protocols affecting various domains and layers. These protocols are different from those used for the previous functionality, in the notion that priority is now given to QoS on demand, perhaps overlooking the time delay to initiate the connection.

Specifically, the Expedited Forwarding Class (EF) IP traffic would need to make intensive use of this functionality, to allow the customers to obtain the assured delay and high capacity bandwidth provisioning, probably using a SLA protocol. An interesting function that can be offered is a direct OCh provisioning from the upper IP level, perhaps with multicast option.

Obviously, it is necessary a complete knowledge of the network structure and network elements in the different domains and layers, and this knowledge should be accessible from any network edge node in a distributed servicing network.

The atomic primitives regarding this network function to be dealt with by the CNI and NNI interfaces are equivalent to the previous procedures to dial-up requests and to process them. The main differences are in the control and management planes.

We can differentiate two types of bandwidth provisioning, depending on the customer needs: a specific data bandwidth or a full wavelength.

TDM bandwidth provisioning

Service for a specific data bandwidth (e.g. 100 Mbit/s) with the corresponding traffic grooming strategies into a wavelength.

OCh provisioning

Service for a full wavelength with specific parameters. Its related processes are equivalent to the ones specified at the section 3.1.3.

5.2.3 Multi-Layer Survivability

Multi-layered network survivability means recovery, restoration and automatic protection switching across layers, to be implemented for a fastest and most efficient traffic protection.

The survivability facilities imply rerouting mechanisms (protection switching / restoration) than can be triggered by both the optical layers network or the client networks. Furthermore, the diverted traffic can be re-routed through different layers, depending on the cause, the scenario and the designed recovery mechanism. Also, the recovery can be triggered from different importance levels, e.g. drastic fault or performance degradation.

These mechanisms are then in close connection to the performance monitoring capabilities and the automatic switching methods, in order to obtain the aimed high QoS and availability of the integrated network.

The related processes include defect location signalling, route discovery, recovery decisions, protection switching orders via signalling, etc. The recovery decisions are taken upon the indication and knowledge of the necessary information regarding the network topology and network element status, critical traffic priority, etc.

The aim of this section is to explain why it is necessary to co-ordinate multiple layer recovery mechanisms. A first sub-section will explain how recovery at multiple layers may increase the reliability of the multi-layer network. A second sub-section will explain why these recovery mechanisms need to be co-ordinated. Detailed information on single layer recovery strategies can be found in [WP2-M1], and is thus not in the scope of this document.

Why survivability at multiple layers?

One may be wondering why it is not sufficient to deploy a recovery in only one single layer of the multi-layer network?

Consider a traditional SDH transport network. Then there exist well-known recovery strategies at several levels [ITU-T G841]: e.g., MSP 1+1 protection, SNCP at VC-4 or VC-12 level.

- MSP 1+1 protection seems to be very suitable for link failures (e.g., fiber cuts). But since digital cross-connects (DXC) terminate MS's, MSP cannot recover from a DXC failure (actually it is possible that the MS-level will not notice the existence of the DXC failure).
- Another solution is to deploy SNCP at the VC-4 level. In this case, it would be possible to recover VC-4 traffic transiting a failing HO-DXC (DXC-4/4). However, when a LO-DXC (DXC-4/1) is attached to this failing HO-DXC, then this LO-DXC will be isolated from the

rest of the network. SNCP at the VC-4 level is not able to prevent this LO-DXC isolation and thus is also not able to recover VC-12 traffic transiting this isolated LO-DXC. As shown in Figure 20, our presentation discussed a similar circumstance in the case of a OTN node failure when deploying a MPLS-over-OTN multi-layer network[1].

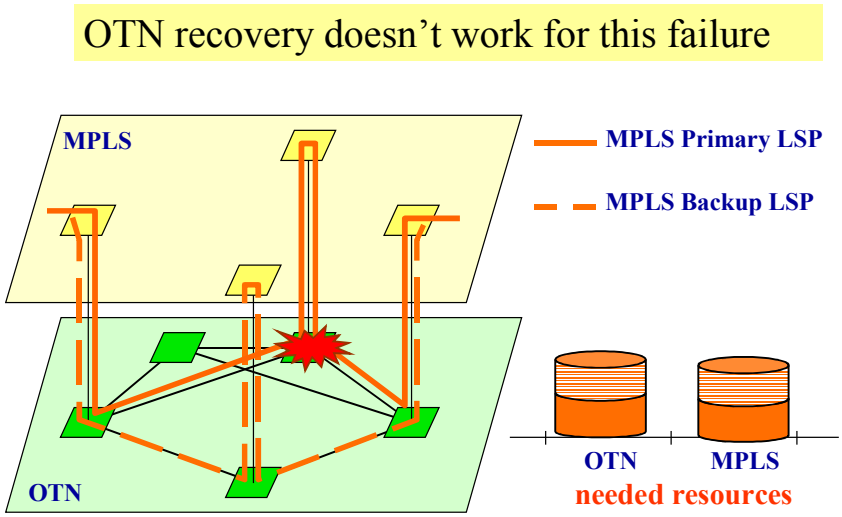


Figure 20: MPLS-LSR isolated due to a OTN-node failure

- The solution would be to perform SNCP at the VC-12 level, instead of at the VC-4 level. However, the drawback of this solution would be that no recovery is foreseen to protect native VC-4 traffic (e.g., a customer may want to lease a complete VC-4 instead of one or more separate VC-12s).

The conclusion is thus that it is not straightforward to decide at which layer to provide recovery. Even more: a reasonable trade-off would combine recovery mechanisms in more than one layer, in order to profit from the advantages of each layer.

The above three points only illustrate the complexity of this decision from a reliability point of view. They can be summarised in general respectively as follows:

- **Recovery at higher layers desired**: because lower layers will not notice failures of higher layer equipment.
- **Recovery at higher layers desired**: because higher layer equipment may be isolated due to lower layer failure scenarios (e.g., lower layer node failure). Only higher layer recovery can restore the traffic transiting this isolated equipment.
- **Recovery at lower layers desired**: since native traffic injected in lower layers is not inside the scope of higher layer recovery strategies.

Of course, considerations have also to be made from other viewpoints than just reliability.

Why co-ordination of recovery mechanisms in multiple layers?

When considering recovery mechanisms in multiple layers, the simplest way is to run these mechanisms in **parallel and independently** form each other. However, this solution may

1 The MPLS-over-OTN scenario fits better in the scope of the LION project, but we preferred to start with an SDH example to explain multi-layer survivability, since SDH is probably a better known technology, which makes the understanding of the general concepts more easy.

have some major drawbacks. In some cases recovery at a higher layer may be necessary (as discussed in the previous section). But, in severe failure conditions, it may be necessary that the lower layer recovers first, in order to restore simply the connectivity of this higher layer. The presentation of [IMEC1] gives such an example, which is shown in Figure 21. This illustrates clearly that both mechanisms have to be coordinated in one way or another. If the higher layer decides that recovery is not possible in its own layer, before the lower layer gets the chance to restore the connectivity, then it is useless to restore the connectivity with lower layer recovery strategies.

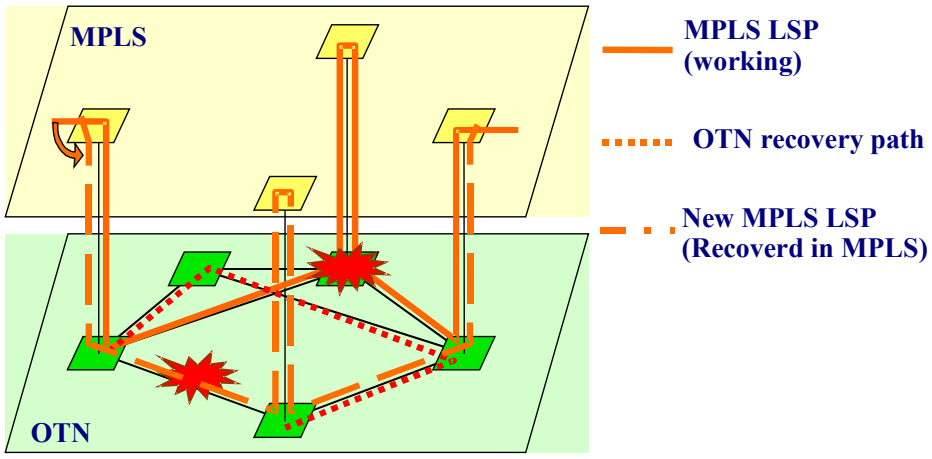


Figure 21: Double failure, requiring recovery in a disconnected client layer

The above discussion is again only from a reliability viewpoint. Of course, considerations from other viewpoints are also important.

An interesting way of co-ordinating these recovery mechanisms is a **sequential approach**: one layer tries to restore the traffic and only the following layer takes over if the current layer does not succeed to restore the traffic. Often a **bottom-up approach** is chosen: recovery is first done in the lowest layer and eventually handed over to a higher layer. Another sequential approach could be a **top-down strategy**. Note that this approach is not very intuitive and/or popular. However, when deploying MPLS as client layer, this approach may gain more interest, since recovery of high priority traffic first may become very easy in MPLS. MPLS recovery tends also to be rather capacity efficient, compared to recovery in its server OTN transport network.

The following section will discuss which layer interworking functionality may be required to co-ordinate recovery mechanisms in multiple layers and multilayer-survivability in general.

5.2.3.1 Layer Co-ordination for Multi-layer Network Recovery

In the previous section we have shown that the deployment of recovery at multiple layers significantly improves the network reliability. However, that section also has raised the need to co-ordinate recovery actions in these layers.

The goal of this section is to provide an overview of inter-working functionality that is needed to co-ordinate properly the several recovery mechanisms.

Failure propagation

In order to make recovery at a higher layer more attractive; a fast detection of failures is needed at this layer. However, when a failure occurs in a lower layer, the fault has to propagate towards the higher layer. When no failure indication signal is sent from server to client layer, the client layer has to rely on its own failure detection mechanism. In the IP/MPLS-over-OTN scenario, the client layer is the MPLS layer and uses the rather slow technique of liveness messages for failure detection. Therefore, such a signal from OTN to MPLS could be very desirable.

One may argue that triggering the higher layer recovery mechanism very fast may introduce the real need for the co-ordination of both layers. This is because the long failure detection time of the higher layer (when no failure indication signal exists) can be seen as a natural hold-off timer. Nevertheless, implementing a hold-off timer, in the case that such a signal is generated may speed-up the process significantly. For instance, when the OTN recovery has done its job within 50 ms, then a hold-off timer set at 100 ms for example, will be much faster then detecting the loss of liveness messages in the MPLS layer, which are sent each second for example.

Thus, we may conclude that an interlayer failure indication signal can be very important.

IWF requirement 1: a failure indication signal from lower layer to higher layer.

Independent Recovery Strategy

Failure detection and propagation is the first phase in the process of network recovery. The next step is to initiate and to perform the network recovery. Section 3.4.1.1 discusses why recovery at multiple layers can be important. On the other hand, section 3.4.1.2 shows that multiple recovery mechanisms should be co-ordinated in one or another way, in order to be effective and to work properly together.

The simplest way is to neglect this need for co-ordination and to hope that everything works fine. This way of working has the big advantage that **no layer interworking** is required. However, hoping that everything goes fine is not always enough. [Wauters 1999] illustrate with a lab test that even a simple switch in the optical layer may trigger the SNCP protection in the SDH layer, which implies that hoping that everything goes fine is an utopia.

Sequential Recovery Strategy

As mentioned in above a straightforward solution could be to try recovery at one layer and only to hand over control to another layer when the current layer fails to restore the traffic. This is the so-called **sequential approach**. We also showed that this approach covers two possibilities: a **bottom-up** approach, where the recovery starts at the lowest layer, and a **top-down** approach, where the recovery starts at the highest possible layer.

Hold-off Timer

Some recovery schemes may require a typical amount of time to perform their job. For example, protection may be terminated within 50-100 ms. Thus, after a certain period of time, elapsed since the failure occurred, one could be almost 100% certain that the recovery was successful or not. In such a case, a so-called **hold-off timer** could be used to delay the triggering of the next layer recovery scheme. The hold-off timer is initiated at the moment that the failure was detected. Only when the hold-off timer goes off and the current recovery mechanism did not yet restore the traffic the next recovery mechanism is triggered.

The advantage of a hold-off timer strategy is clearly that it can be implemented rather easily. However, a major drawback is the delay introduced by the hold-off timer.

As we described in [WP2-M1], the current proposals in the IETF already cover the possibility to implement a hold-off timer in MPLS-recovery schemes. This is a typical bottom-up approach. A similar approach could be deployed in a top-down strategy. However, this would require that the server layer is able to detect that its client layer was able to restore the traffic. This is clearly a violation of the network-layering concept. However, in the case of MPLS (or any other packet based network)-over-OTN, a client specific interface card in the OTN may notice that no packets are forwarded anymore via that interface, which may indicate a successful restoration of the client layer traffic.

IWF requirement 2a: a hold-off timer, delaying the triggering of the next recovery scheme

Recovery Token

In the case that a hold-off timer would introduce too long delays, one may opt to deploy a recovery token strategy. In this case, a **recovery token** is sent towards the next recovery mechanism, as soon as it is known that the current recovery scheme is unsuitable to restore the traffic. In this way, the next recovery scheme has to wait for the receipt of a recovery token instead of a fixed hold-off timer going off.

It is clear that this way of working is beneficial from a speed point-of-view. However, implementing such a recovery token signal from one layer to another is less straightforward than implementing a hold-off timer. Another advantage of a recovery token, is that it fits better in a top-down (and of course also in a bottom-up) approach than a hold-off timer, which really fits only in a bottom-up approach without violating network-layering concepts.

IWF requirement 2b: sending a recovery token signal from one layer to the other, as soon as it is known that the current layer is unable to restore the traffic.

Integrated Recovery Strategy

Another solution to deploy network survivability at multiple layers is to avoid the deployment of multiple recovery mechanisms, by integrating them into a single multilayer recovery mechanism. This integrated single multilayer recovery mechanism requires a full view of network; thus a view covering all network layers.

The implementation and deployment of a mechanism covering all network layers is at least not straightforward or trivial[2]. However, the recent proposals on MPλS see opportunities to integrate the control plane of the electrical and optical nodes. Such an integrated control plane may be a big step forward in the study of single multilayer recovery mechanisms.

IWF requirement 3: deployment of a single integrated recovery scheme, covering multiple layers

5.2.3.2 MPLS recovery plus OTN re-optimization

We have discussed an opportunity for lambda switching in the context of network recovery. If one decides to do the recovery at the MPLS layer, then it can be interesting to establish a short-cut wavelength along the alternative path used by the MPLS layer. This may be done in order to avoid the overload of LSRs along the alternative path.

2 As far as we understand the current status of the LION test-bed, we don't believe that such an integrated recovery mechanism is possible in the LION test-bed. The software controlling the optical nodes runs on a separate PC, independent of the control software inside the GSR routers.

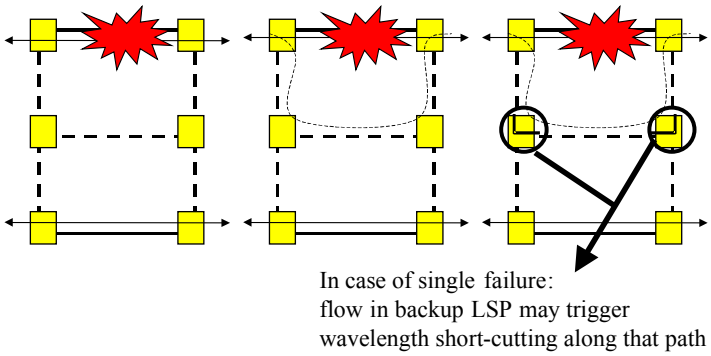


Figure 22: OTN re-optimization after MPLS recovery

Figure 22shows an example, where in the case of a failure, MPLS is sending traffic along a backup LSP. This flow going through the backup LSP may trigger the set-up of a new wavelength path along the backup LSP, in order to avoid the overload of the LSR in at the middle row of the network.

Note that this has nothing to do with recovery at multiple layers! The fast reconfiguration of the optical layer may make higher layer recovery a little bit more attractive.

IWF requirement 2a: fast reconfiguration of server layer, after recovery in a client layer

5.2.3.3 Summary

Table 6gives an overview of the discussed interworking functionality, which may make multilayer recovery possible or more attractive. Note that we also tried to indicate in this table which functionality may be realistic to implement in the LION test-bed (at least as far as we have a correct view of the possibilities of the LION test-bed).

Table 6: Summary of the required interworking functionality

IWF req #	Description	Realization in test-bed
1	Interlayer failure indication signal	Easy
2a	Hold-off timer	Easy
2b	Recovery token	Harder
3	Single integrated recovery scheme	Not possible
4	Fast reconfiguration of server layer, after higher layer recovery	Easy

6 Preliminary LION Roadmap

SDH Networks

The introduction in the '90s of SDH allowed Network Operators to deploy the wide-area electronic networking. The need of having greater and greater capacities on the transmission networks has brought to increase the bit rate of such TDM systems up to 10 Gbit/s or 40 Gbit/s (STM-64 and STM-256). This last bit rate could be the upper practical limit, even if not the theoretical one, for the capacity of the TDM systems for two main reasons:

- technological difficulties to make regenerators operating at higher bit rates;

- effects of chromatic and polarisation dispersion and optical fibres non linearity.

Historically Network Operators have used several layers to build their transport networks: adopting for example IP routers over ATM switches over PDH or SDH network elements. Figure 23 reports a typical layered architecture based on legacy systems.

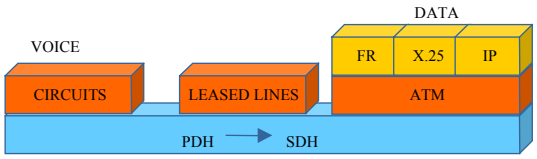


Figure 23 - Example of legacy current scenario

WDM and Optical Transport Network (OTN)

In order to overcome this capacity limit (10-40 Gbit/s), since '95 WDM point-to-point systems are under deployment in long distance networks. They use the transparency of the optical fibres on a very large bandwidth in order to transmit several wavelengths on the same fibres. Each wavelength of a WDM system carries a TDM digital stream and it is subject to all the transmission limitations that are typical of the TDM technique.

Furthermore today WDM is almost mature to provide networking functionality by means of Optical Network Elements such us OADM. ITU-T SG13 recommended the functional architecture of the Optical Transport Network in the Recommendation G.872. As such the OTN could represent another potential layer added to a multi-layer transport network.

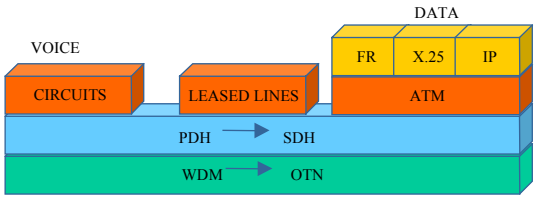


Figure 24 - Introduction of the OTN

Considering that today the estimated growth of data traffic is from 35% to 60% per year and that within ten years the network capacity demand could be up to 100 times the current one, the OTN can provide such a large amount of raw bandwidth supporting (even directly) data traffic.

Automatic Switched Optical Networks (ASON)

Very recently (March, 2000), ITU-T SG13 started to define the Recommendation G. ason. This Rec. describes the control and management architecture for an Automatic Switched Optical transport Network (ASON). The recommended architecture recognises that the optical transport network is capable of supporting multiple clients and enforces separacy of control from client networks.

The ASON is still an Optical Tranport Network but with a further fuctionality: the capability of switching Optical Channels automatically. Three kind OCh connections are envisaged:

- permanent: the set up is done from the management system with network management protocols (NMI)

- soft permanent: the set up is done from the management system which uses network generated signalling and routing protocols to establish connections (NMI and NNI)
- switched: the set up is done directly by the customer on demand by means of signalling and routing protocols (UNI and NNI)

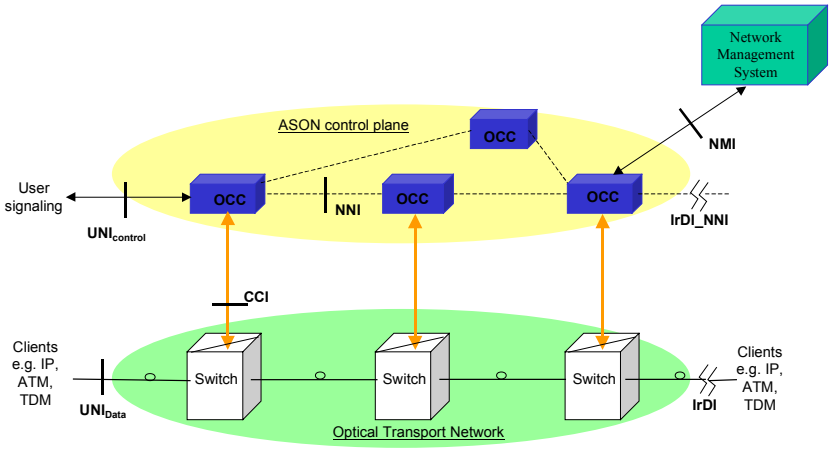


Figure 25 - Automatic Switch Optical Network (Source ITU-T Rec. G.ason).

Generalised Multiprotocol Label Switching (G-MPLS)

Generalized MPLS differs from traditional MPLS in that it supports multiple types of switching, i.e., the addition of support for TDM, lambda, and fiber (port) switching.

The original architecture has recently been extended to include LSRs whose forwarding plane recognizes neither packet, nor cell boundaries, and therefore, can't forward data based on the information carried in either packet or cell headers. Specifically, such LSRs include devices where the forwarding decision is based on time slots, wavelengths, or physical ports.

Given the above, LSRs, or more precisely interfaces on LSRs, can be subdivided into the following classes:

1. Interfaces that recognize packet/cell boundaries and can forward data based on the content of the packet/cell header. Examples include interfaces on routers that forward data based on the content of the "shim" header, interfaces on ATM-LSRs that forward data based on the ATM VPI/VCI. Such interfaces are referred to as Packet-Switch Capable (PSC).

2. Interfaces that forward data based on the data's time slot in a repeating cycle. An example of such an interface is an interface on a SONET Cross-Connect. Such interfaces are referred to as Time-Division Multiplex Capable (TDM).

3. Interfaces that forward data based on the wavelength on which the data is received. An example of such an interface is an interface on an Optical Cross-Connect that can operate at the level of an individual wavelength. Such interfaces are referred to as Lambda Switch Capable (LSC).

4. Interfaces that forward data based on a position of the data in the real world physical spaces. An example of such an interface is an interface on an Optical Cross-Connect that can operate at the level of a single (or multiple) fibers. Such interfaces are referred to as Fiber-Switch Capable (FSC).

Using the concept of nested LSPs (by using label stack) allows the system to scale by building a forwarding hierarchy. At the top of this hierarchy are FSC interfaces, followed by LSC interfaces, followed by TDM interfaces, followed by PSC interfaces. This way, an LSP that starts and ends on a PSC interface can be nested (together with other LSPs) into an LSP that starts and ends on a TDM interface. This LSP, in turn, can be nested (together with other LSPs) into an LSP that starts and ends on a LSC interface, which in turn can be nested (together with other LSPs) into an LSP that starts and ends on a FSC interface.

Summary

G-MPLS is a first step to propose (by IETF) the use Internet protocols to accomplish the functions needed to automate the OTN. These functions are protocols to discover relevant topology in the OCh layer, protocols to disseminate some or all of this topology information to the switch elements comprising the OTN, and protocols to carry call setup and restoration control among the switches.

The current ITU position express some doubt whether that proposal can actually be made to work at all: this is due to the fundamentally different nature of an Optical Channel Circuit and a label Switched Path.

The most obvious disadvantage is that using the routing infrastructure of the client to control the OTN effectively makes the OTN a single client network. A second obvious disadvantage is that the inner detail of a server network is the intellectual property of that network owner, and should not be shared with a client.

For these reasons ASON is designed as an independent network, and particular attention must be paid to its connection service definition so that future services can be supported.

As data is the fastest growing segment of network traffic, **transport network models are likely to evolve to data-centric solutions, primarly ASON and then potentially G-MPLS based**. As the OTN can provide the large amount of raw bandwidth supporting the increasing data traffic, **in the short term a client-independent OTN is likely to be the missing link between legacy (e.g. SDH) and data centric networks.**

7 Conclusions

This Deliverable provides a preliminary description of services and requirements of next generation networks. Particularly it starts from a description of the main business drivers moving the network evolution. An overview of the envisaged application and transport services for the next generation networks is considered the starting point to identify innovative functionality (e.g. interworking) for multi-layer networks.

The state of art of enabling technologies and a general survey of currently deployed transport networks allows to set the starting point of the project roadmap.

The main preliminary conclusion is: as data is the fastest growing segment of network traffic, **transport network models are likely to evolve to data-centric solutions, primarly ASON and then potentially G-MPLS based**. As the OTN can provide the large amount of raw bandwidth supporting the increasing data traffic, **in the short term a client-independent OTN is likely to be the missing link between legacy (e.g. SDH) and data centric networks.**

APPENDIX 1
Mapping Solutions taken into consideration in LION

A.1.1 Multi-Protocol Label Switching

Because of its features, MPLS is widely considered as a viable technology for traffic engineering. This section first presents MPLS notions, concepts and the architecture and then focuses on MPLS-based traffic engineering aspects and approaches.

A.1.1.1 What is MPLS?

MPLS can be a lot of things depending on the point of view. The following sections will describe what MPLS is, seen from different points of view. These sections also try to point out what MPLS is not.

MPLS is a technique for IP over ATM inter-working.

MPLS is a convergence of a number of "IP switching" schemes. IP switching is a technique that uses ATM hardware to speed up the forwarding of IP packets. It is important to know that the ATM hardware is controlled by IP routing and not by ATM signalling. There are a number of different IP switching implementations: Cisco Systems Tag Switching, IBM's Aggregated Route based IP Switching (ARIS), Toshiba's Cell Switch Router (CSR) and NEC's Ipsofacto (now renamed to Lcatm) [SWITCH]. In order to standardise all these IP switching techniques a new IETF working group came to life in 1997. The MPLS working group has since then been working on forming a common technology for IP switching.

MPLS is NOT an overlay technique.

There are a number of techniques for IP over ATM which are overlay techniques, MPLS is not an overlay technique. Examples of overlay techniques are LANE [LANE], MPOA [MPOA] and the work done in the IETF ION working group. In the case of overlay techniques there are two different networks: a network at layer 2 (ATM) and a network at layer 3 (IP). This leads to a number of disadvantages: the two networks both have to be managed (more management overhead), both L2 (switches) and L3 (router) equipment is necessary (more equipment) and the scalability is limited (due to full meshed peering) [MYTH].

An MPLS network is one single network in contrast with the overlay techniques. In MPLS there is however a separation between control and forwarding. The control is based on standard IP control mechanisms like IP routing and an extra control protocol called Label Distribution Protocol (LDP) [LDP]. LDP will be introduced in the next section and will be described further in detail in the following sections. The advantage of separating forwarding and control is that changing control will have no (or limited) consequences for the forwarding. This makes MPLS a very flexible approach because it makes it easier to introduce new features in the control layer.

MPLS is an advanced forwarding scheme.

In regular (that is non-MPLS) IP networks packets are forwarded in a router after consulting the routing table. A routing table contains information about the next hop and the outgoing interface for a certain destination address (found in the IP header). The destination addresses in this table are aggregated in order to reduce the number of entries in this table. The entries are aggregated, by indicating the length of the destination addresses (from 0 to 32 bits). If n is the length of address a then only the first n (most significant) bits of a are considered. The resulting address is called a prefix. This aggregation of addresses has the drawback that searching through the table becomes more complex. Instead of looking for an

exact match in the table, the result of the search must be the entry with the longest address that matches the address. This process is called a longest prefix match.

MPLS takes another approach to route the IP packets through the network: labels. Labels are fixed length entities that have only a local meaning. In the case of ATM based MPLS a label is a VPI, VPI/VCI or a VCI identifier[3] [VCSW], [VCID]. In traditional ATM networks these identifiers are installed with UNI[4] or PNNI signalling. Since MPLS does not use ATM signaling another protocol is needed. The MPLS working group developed the Label Distribution Protocol (LDP). LDP is responsible for distributing labels across the different MPLS routers (called LSRs: Label Switching Routers) so that a packet can travel across the network by switching labels in the core. The concatenation of these installed labels in the different LSRs is called a Label Switched Path (LSP). How these labels are installed and what event (trigger) leads to the installation of the labels will be described later on.

MPLS is NOT a QoS framework, however facilitates QoS delivery.

MPLS is sometimes seen as a technique that stands on the same level as Intserv and DiffServ. This point of view is not correct: MPLS is an advanced forwarding scheme that can have a better forwarding performance then traditional routers (although gigabit routers can be made as fast as MPLS routers) but this performance gain does not make it a QoS framework.

There are however a number of relations between QoS and MPLS. With the use of CR-LDP (extensions to LDP for traffic engineering) [CRLDP] it becomes possible to signal the requested QoS level for a LSP. Another approach is to use a modified version of RSVP for traffic engineering [RSVPT]. Even with the use of CR-LDP or RSVP topics like admission control algorithms, scheduling, constrained based routing etc. are not a part of MPLS.

MPLS is NOT constrained to only ATM as a link layer.

Although the IP switching started out as a technique for IP over ATM this is no longer true, MPLS is now available over a number of link layers. These link layers can be divided in two categories. To the first category belong the link layers that already switch on "labels". Examples are of course ATM but also Frame Relay (the DLCI values are used as label) [FR]. The second category contains link layers that do not have a field in the header that can be used to transport the labels. In this case an extra header, the shim header is used [ENCAPS]. This shim header is not link layer specific. The header is inserted after the layer 2 header and before the IP header, this is necessary to support IP fragmentation. Ethernet and PPP are examples of link layers that use a shim header.

A.1.1.2 MPLS architecture

The MPLS architecture consists of a control layer and a forwarding layer [ARCH], [MPLS-FRAME]. The functionality of the forwarding layer is simple: to switch labels. The control layer is responsible for the IP control functionality of a regular IP router and the distribution of labels.

A closer look at labels.

It is apparent that "labels" constitute the center of the MPLS architecture. However these labels do not look the same for every link layer. There are two types of link layers in this context: link layers that already support a kind of "label" to switch and link layers that need to use a "shim header" which contains a label. The former category includes ATM and Frame Relay, the latter includes PPP and Ethernet.

---

[3] From now on we will use VPI/VCI for the three possibilities.

[4] Or CNI in LION terminology

ATM and Frame Relay already have an identifier that can be used as a label. In ATM this is the VPI/VCI field in the cell header and in Frame Relay this is the DLCI field. These two link layers have support for switching on these fields in hardware.

On the other hand PPP and Ethernet need a shim header. This shim header (see Figure 26 and Figure 27) contains everything that is needed to forward packets through a MPLS domain. First of all the shim header contains a label but the header also contains three experimental bits (used in DiffServ [MPLS-DS] and ECN [MECN]), a bottom of stack indicator (needed for label stacking, see next section) and a TTL field (will be explained later).

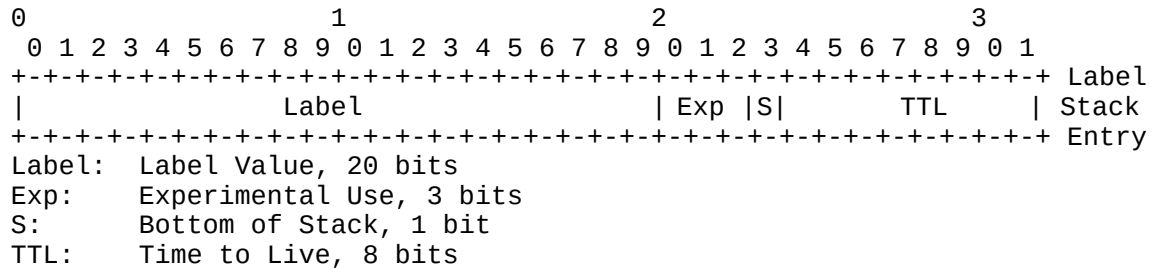


Figure 26: The shim header

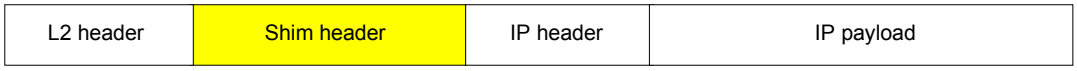


Figure 27: the position of the shim header

Label stacks

The labels used in MPLS (ATM VPI/VCI, Frame Relay DLCI and the label field in the shim header) are just identifiers; they are used to create label switched paths. In traditional ATM the VPI field can be used to aggregate multiple PVCs or SVCs. MPLS has a similar but more generic concept to aggregate LSPs. Multiple shim headers can be stacked. We will first look at link layers that only use shim headers (like PPP and Ethernet) and afterwards look at ATM and Frame Relay.

At every node only the top label is considered when making the forwarding actions. On inspection of the top label the LSR can do the following operations on the label stack: replace the top label with a new label (swap), pop the label stack or replace the top label and push a number of new labels on the label stack.

ATM and Frame Relay switch on VPI/VCI and DLCI identifiers and not on labels in the shim headers. However (limited) support is provided for label stacks in an ATM or FR segment of a MPLS network. When the packet enters the ATM/FR segment the top entry of the label stack is copied into the VPI/VCI or DLCI field. Then the value of the label at the top entry of the label stack is set to zero. In the ATM/FR segment of the MPLS network it is not possible to do operations on the label stack, the only operations possible are the operations on the copy of the top entry of the label stack. In the ingress of the ATM/FR segment it is possible to push the label, in the core to swap the label and in the egress to pop the label. After the packet has passed the ATM/FR segment the VPI/VCI or DLCI label is copied back into the top entry of the label stack.

TTL

The TTL field in the IP header has to be decreased at every hop in an IP network. This mechanism is used to prevent packets from living forever. The TTL field is also used in utilities like ping and traceroute. The section describes the support for TTL decrement in MPLS networks.

Again there's a difference between ATM/FR and PPP/Ethernet. Since the shim header contains a TTL field the LSRs are able to decrement the TTL just like in the normal IP forwarding case.

ATM/FR LSRs can not decrement a TTL field since they can only act on their native identifiers (VPI/VCI, DLCI). The solution for these link layers is to compute the whole TTL decrement for the ATM/FR segment a LSP traverses [MICMP]. This TTL decrement is computed before the LSP is used. When a packet arrives at the first ATM/FR LSR the calculated TTL is subtracted from the current TTL (found in the shim header). Appropriate actions must be taken on this result (e.g. sending ICMP messages). If the result is positive then this result is written in the top entry of the label stack. So a packet travelling through a LSP which goes through an ATM/FR segment will have the correct TTL before the ATM/FR segment and after the ATM/FR segment. The TTL will have a constant value in the ATM/FR segment (the same value as just after the ATM/FR segment).

Label merging

Some link layer technologies are capable of merging labels. Label merging means that if there's a merge point (a multi-point to point subtree) in the network then only one label is needed as outgoing label.

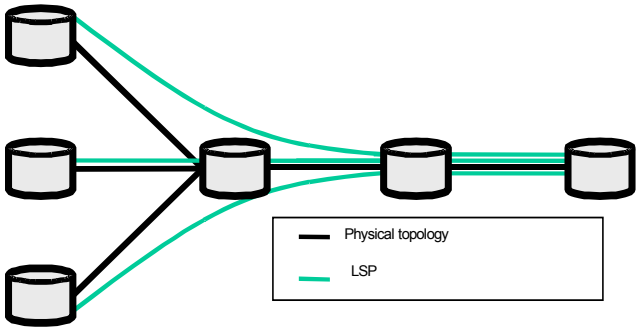


Figure 28: multiple LSPs with non-merge capable LSRs

When dealing with non-merge capable media one must use a different outgoing label for every incoming branch of the merge point. This leads to a situation where multiple LSPs needs to be set up between two adjacent LSRs, as shown in Figure 28. Most ATM switches and Frame Relay equipment are non-merge capable.

The distribution of labels

There are two issues when looking at the distribution of the labels in a MPLS network. The first one is what effect triggers the distribution of the labels (data flowing through the network, routing changes or reservations) and the second issue is which protocol is used to distribute the labels. Also some of the protocols have a number of modes in which to operate.

When looking at the protocols to distribute the labels there are again two possibilities: devise a new protocol (LDP – Label Distribution Protocol- and Constraint Routed-LDP) or piggy back it on an existing protocol (tunnel extensions for RSVP or BGP [BGPLAB]).

In the following sections we will look at hop-by-hop routed LSPs. This means a LSP that is set up according to the routing tables of the LSRs in the MPLS network. The alternative is to

use an explicit routed LSP (ER-LSP): a LSP that is set up following the path that is included in the request message.

The control protocol that is proposed by the MPLS working group as the control protocol for hop-by-hop routed LSPs is LDP. Two approaches are discussed for explicit LSP set up in the MPLS working group: CR-LDP and extensions to the RSVP protocol. The following section describes the LDP.

A.1.1.3 Label Distribution Protocol (LDP)

As mentioned before, LDP distributes labels in the MPLS network. Labels are distributed for "Forward Equivalent Classes" (FECs). A FEC can be either an IP prefix or an IP host address. Addresses that belong to the same FEC can be forwarded in the MPLS domain in the same way. As a result it suffices to construct one label switched path for each FEC[5]. When a label is mapped to a certain FEC one speaks of a FEC-Label binding. A LSP can be regarded as a concatenation of FEC-Label bindings over a number of LSRs. An important consequence of this definition is that a LSP is an unidirectional path.

LDP gives the user a great deal of freedom how to set up the LSPs. This is reflected in a number of different modes [ARCH]. These modes will be described in the following subsections. The reason for the multitude of modes is that support is needed for multiple link layers (some link layers mandate some modes), the desire to "emulate" the proprietary IP switching techniques and the desire for flexibility.

Topology driven versus data driven

The LSPs for the different FECs in a network can be set up according to the routing table before traffic begins to flow through to the network (topology driven). Alternatively the LSPs can be set up after a certain amount of traffic has passed through two points of the network (data driven or flow driven). Topology driven MPLS has the advantage that the number of LDP messages is small (LDP activity takes place initially when MPLS is enabled and after routing changes). However topology driven MPLS has the disadvantage that generally a full mesh of LSPs is constructed between all the edge LSRs (also called Label Edge Routers LERs). Depending on whether or not the LSRs are merge capable the number of LSPs is O(n) or O(n2) respectively[6].

Data driven MPLS has the advantage that it potentially requires less LSPs. The disadvantages of flow driven MPLS are the greater LDP overhead, extra functionality needed (flow detection), the set up delay and a more unpredictable network behaviour (different behaviour when traffic is switched).

Downstream versus upstream allocation

Since labels only have a local meaning, these labels can be allocated decentralised by the switch controllers. A core LSR has always two neighbors: one upstream and one downstream. It also has an incoming label and an outgoing label. With that in mind there are two approaches possible: the LSR supplies his upstream neighbor with a label and receives a label from his downstream neighbor or the other way around. When the LSR receives the label from his downstream neighbor this is called "downstream allocation". LDP only supports downstream allocation.

---

[5] one LSP per ingress and FEC in the case of non-merge capable switches

[6] Note that the peering is always O(n) in contrast with overlay models which require $O(n^2)$ peering which leads to $O(n^3)$ IGP rerouting.

Unsolicited distribution versus distribution on demand

The previous subsection described that labels are chosen (allocated) by the downstream LSRs. When these labels are distributed spontaneously this is called "unsolicited label distribution". When the upstream LSR always sends a request to his downstream neighbor this is called "distribution on demand". Unsolicited label distribution is also called "Pushed" distribution while distribution on demand is called "Pulled" distribution.

Independent versus ordered control

In "independent control" an LSR, recognizing a particular FEC, makes an independent decision to bind a label to that FEC and to distribute that binding. In "ordered control" an LSR only binds a label to a particular FEC if it is the egress LSR for that FEC, or if it has already received a label binding for that FEC from its next hop for that FEC.

Liberal retention versus conservative retention

Consider the situation where an upstream LSR has received and retained a label mapping from his downstream peer. When the routing changes and the original downstream peer is no longer the next hop for the FEC then there are two possible actions the LSR can take. The LSR can release the label, this is called "conservative retention" or it can keep the label for later use: "liberal retention".

Conservative retention ("release on change") has the advantage that it uses fewer labels, liberal retention ("no release on change") has the advantage that it allows for faster reaction to routing changes.

Label use method

Labels can be used as soon as a LSR receives them ("use immediate") or the LSR can use the label unless a loop has been detected ("use loop free") [LOOP1], [LOOP2].

Supported LDP modes

Media that do not support label merging (most ATM and Frame Relay switches) must use distribution on demand.

Media where the labels are a scarce resource[7] (again ATM and FR) should use conservative retention and distribution on demand.

Important LDP messages

This section briefly describes the important LDP messages. The "Hello" message is used to discover LDP neighbors and remote peers (with the use of "Targeted Hello"). The "Initialisation" message is used to negotiate about the LDP session parameters. LSPs are set up with the "Label Request" and "Label Mapping" messages. "Label Withdraw" (upstream) and "Label Release" (downstream) messages are used to tear down LSPs. Exceptions and failures are reported with "Notification" messages.

A.1.1.4 DiffServ in MPLS-based networks

The recent work of the MPLS workgroup has been described in [MPLS-DS]. Special about the MPLS architecture is its connection-oriented nature, which implies that packets of a particular LSP will not be delivered disordered. Therefore the Ordered Aggregate (OA) concept has been defined [DIFF-NEW]. An OA is a set of BAs that share an ordering constraint. The corresponding set of PHBs is called a PHB Scheduling Class (PSC). More precisely, a PHB can be seen as the PSC plus the drop precedence.

To support DiffServ on an MPLS-based network, there exist mainly two possibilities

[7] And wavelengths in MPλS.

1. **EXP-Inferred-PSC LSPs (E-LSPs)**: in this case a single LSP can support at most 8 BAs, by mapping the EXP-field in the shim header (see Figure 26) to the appropriate PHB. This mapping can be pre-configured or explicitly signaled at label setup.
2. **Label-Only-Inferred-PSC LSPs (L-LSP)**: this case requires that at LSP setup the label-to-PSC mapping is explicitly signaled. The drop precedence is carried in the MPLS header. In the case of the shim header, the EXP field is used for this purpose. In the other case, the right field in the link layer header has to be used.

Note that these binding schemes don't specify anything about resource allocation. Resource (e.g. bandwidth) requirements may be signaled explicitly at LSP setup. Those signaled requirements can then be used to perform admission control or to update the current amount of allocated resources. In the case of L-LSP binding resources can be explicitly allocated for the PSC corresponding to the LSP being established. In the other case, for E-LSP bindings, resources can only be allocated for the whole set of PSCs corresponding to that particular LSP.

A.1.1.5 Traffic Engineering (TE) support in MPLS

Although MPLS is not a framework developed for traffic engineering, it provides some significant tools for this purpose due to its path-oriented nature [CONC-TE], [MPLS-TE]. Concepts like Forwarding Equivalent Class (FEC), traffic trunk and explicit routed LSPs (ER-LSPs) are important features for traffic engineering. A traffic trunk aggregates traffic flows belonging to the same class into a single LSP [RFC-2430]. It can be seen as a routable object: this means that its path can be changed.

These concepts make it possible for MPLS to allow sophisticated routing control capabilities and QoS resource management techniques [MPLS-QoS], needed for traffic engineering.

Important to support TE in MPLS, is the induced MPLS graph [MPLS-TE]: the nodes of this graph represent the LSRs of the network and the links represent the logical end-to-end links between these LSRs, provided by the LSPs. There are two main problems to solve:

Which traffic flows to aggregate in the same traffic trunk or LSP? And what resources (e.g. capacity) is required for each traffic trunk?

How to map this logical induced MPLS graph on the physical topology, taking into account the limited resources on this physical topology?

To be able to solve this problem traffic trunks and resources are assigned a set of attributes.

Traffic trunk attributes: describe the behavior of the corresponding traffic trunk (e.g., traffic parameter, generic path selection and maintenance attribute, priority, pre-emption, resilience and policing attributes).

Resource attributes: constraint the placement of traffic trunks through the corresponding resource.

A constraint-based routing framework should come up with a route for each LSP, without violating the constraints implied by the traffic trunk and resource attributes. Note that the framework may calculate these routes off-line.

A.1.2 Dynamic Packet Transport

A.1.2.1 Introduction

Telecommunications networks around the world are facing a significant change in their network architectures: the shift from a voice based (circuit) to a data based (packet) network. The explosion of bandwidth fueled by the growth of data traffic, particularly IP traffic, is

disrupting the nature of networks and is driving next generation architectures. Data traffic, which has been accommodated as well as possible on the voice-centric network, can no longer be handled by optimizing the old circuit-switched infrastructure.

The existing circuit-oriented public infrastructure, based primarily on time-division multiplexing (TDM), must evolve and incorporate data packet technology that combines high-performance switching and routing with new and existing optical technologies and standards: it requires optical internetworking.

Optical internetworking combines high-performance data and optical networking technologies to create new optical networking solutions that can efficiently support the exponential growth of data traffic.

Cisco offers a five-phase plan to face the deployment of optical internetworking while optimizing the existing technology and investment (Figure 29). The latter phases of the plan enable the optical internetworking concept providing data optimized transport over optical transmission technologies.

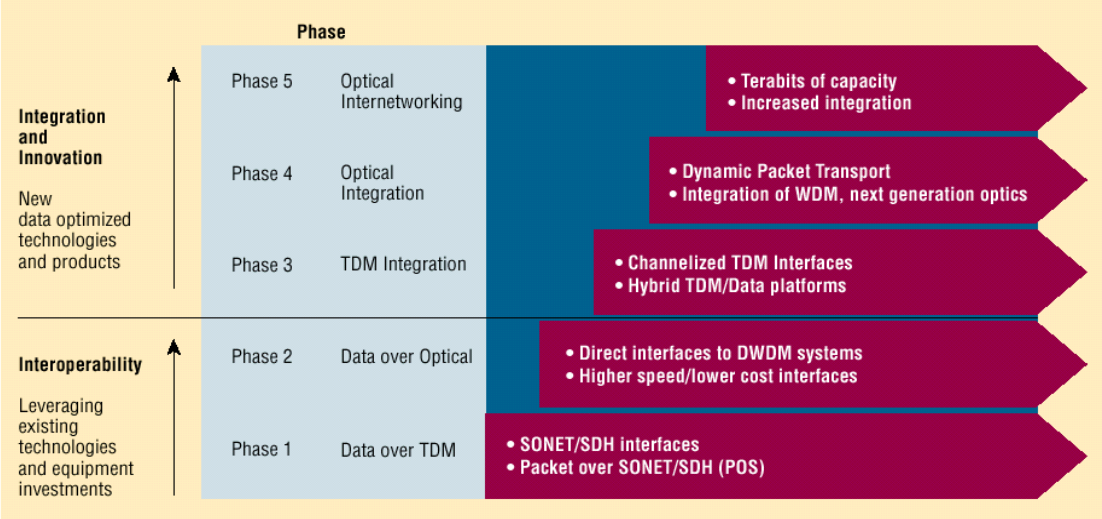


Figure 29: Cisco's Optical Internetworking Strategy.

The Dynamic Packet Transport (DPT) is placed in the fourth phase of Cisco's optical internetworking strategy as metropolitan and wide area network solution providing an innovative and optimized network architecture for ring-based delivery of IP services [DPT-1]. Even if the advantage and applicability of DPT as wide area network solution is not clear in the moment, it seems to be a interesting solution for local - and metropolitan area applications.

Some key benefits of DPT technology are listed below:

- Like other optical internetworking solutions, DPT reduces costs and complexity by eliminating intermediate layers between the IP layer and the optical layer. In this case, it would be eliminated the SONET/SDH equipment.
- DPT increments bandwidth efficiency because of the spatial reuse of bandwidth and the statistical multiplexing of packets.
- DPT enables services such as voice and video over IP and Virtual Private Networks (VPN) to be transparently and robustly extended to the metropolitan and wide area networks.

The DPT functionality is enabled by a new Media Access Control (MAC) layer protocol called Spatial Reuse Protocol (SRP) operating over a dual-ring network topology.

A.1.2.2 DPT Foundations

This section presents basic terminology used with the DPT system and the SRP protocol.

As shown in Figure 30, a DPT ring is composed of nodes that are interconnected by a dual ring consisting of two counter-rotating fibers. In order to differentiate between the two rings, the rings are referred to as "outer" ring and "inner" ring. Both rings transport concurrently data and control packets and while data information is sent in one direction, called 'downstream', the corresponding control messages are sent in the opposite direction, called 'upstream' [DPT-2]. Therefore, control signal propagation that is necessary for access control and self-healing purposes can be accelerated following the shortest path.

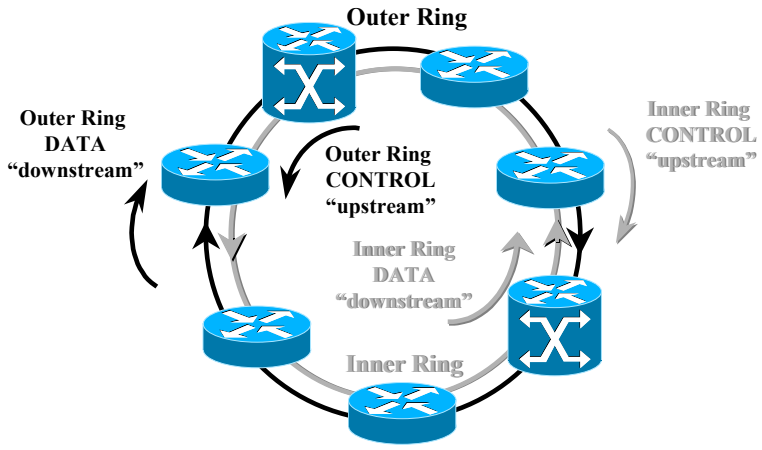


Figure 30: DPT Ring.

DPT provides the following features inherited from the SRP protocol:

- Efficient use of bandwidth.
- Support for priority traffic and multicasting.
- Scalability across a large number of nodes.
- Topology discovery.
- Fairness among nodes accessing to the ring.
- Redundancy and protection in the event of a failed node or fiber cut.

These features are explained in the next section when defining and describing the SRP protocol.

A.1.2.3 The Spatial Reuse Protocol

The Spatial Reuse Protocol is a new MAC layer protocol for ring configurations and takes its name from the *spatial reuse* concept. The entire protocol is exhaustively treated in an Internet Draft [DPT-3].

In principal the SRP protocol is layer 1 (media) independent and can be used over a variety of underlying technologies such as SONET/SDH, WDM, dark fiber, or mixed environments. The initial SRP implementation makes use of SONET/SDH framing. This mapping is the same as the Packet over SONET (POS) mapping, using the point-to-point protocol (PPP) and the High-level Data Link Control (HDLC) to encapsulate SRP packets within a SONET/SDH frame. This initial implementation also supports SDH OAM-flows.

The use of DPT on top of a SDH or WDM infrastructure will enhance the possible link span because the maximum link span depends on these underlying transport technologies only.

Among other important features such as bandwidth efficiency or priority and multicast supporting, two protocols or algorithms lay the foundations of the SRP protocol: the SRP fairness algorithm (SRP-fa) and the *Intelligent Protection Switching protocol* (IPS). The former controls the access to the shared media ensuring fairness, bounding latency and avoiding privileged nodes or conditions while undertaking to prevent congestion, and, the latter consists of a protection scheme of the dual-ring. These two algorithms will be explained later in this section.

A.1.2.3.1 SRP Features

As mentioned before, DPT provides a set of features inherited from the SRP protocol. Next, these and other features are explained:

Bandwidth Efficiency

The *spatial reuse* concept, which is used in rings to increase the bandwidth available of the ring, refers to the fact that packets (unicast) only circulate along spans between the source and the destination node rather than the whole ring as in other protocols such as FDDI and Token Ring.

This operation, in which destination nodes remove the packets from the ring, is known *as destination stripping*. Unlike protocols such as Token Ring and FDDI, there are no shared tokens to access to the ring (source stripping) and it is used the SRP-fa to control the access to the media. Due to the spatial reuse (or destination stripping), the bandwidth is used in a more efficient manner since bandwidth is only consumed on traverse segments allowing nodes to transmit concurrently. In addition, the use of the bandwidth is dynamic; there are no bandwidth reservation or provisioned connections.

SRP Packet Formats

As depicted in Figure 31, there are two types of SRP packets: SRP data packets and SRP control packets. The figure shows the packet formats of the SRP version 2 where the SRP header has passed from 4 bytes to 2 bytes, thus, reducing overhead.

The maximum transfer unit (MTU) for data packets is 9216 octets and the minimum transfer unit is 55 octets. The minimum limit corresponds to ATM cells transported over SRP since the version 2 of the protocol has added a SRP cell mode to carry ATM as payload (2 bytes SRP header + 53 bytes ATM). Nevertheless, IP datagrams are the main payload of SRP packets.

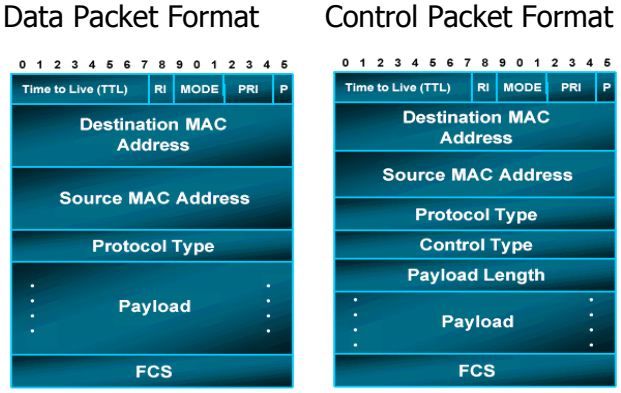


Figure 31: SRP Data and Control Packet Formats.

There are three different control messages used in SRP rings*: Usage, Topology Discovery and Intelligent Protection Switching control packets*.

Packet Prioritization and Processing

In order to provide support for real time, mission critical applications and control traffic, the SRP protocol has been designed to support packet prioritization.

The node utilizes a mapping between the IP precedence bits in the Type of Service (ToS) field into the SRP MAC header priority field (mapping 16 levels of priority to 8 levels).

It is worth noting that, although the SRP priority field size are 3 bits (8 possible levels), there are only two levels of priority working within SRP rings. Thus, a priority threshold is used to determine if the packets should be placed in the high or low priority queues (mapping 8 levels of priority to 2 levels). However, the full 8 levels of priority in the SRP header can be used prior to transmission onto the ring as well as after reception from the ring.

The motivation of these two levels of priority is the use of a transit buffer in the node that consists of two separate fifo queues (high and low priority) in order to forward packets to other nodes. To send their own packets, the nodes make use of a transmit buffer also divided into two different queues (high and low). The following figure illustrates the SRP packet processing (reception and transmission) in which both the transit and the transmit buffer are shown.

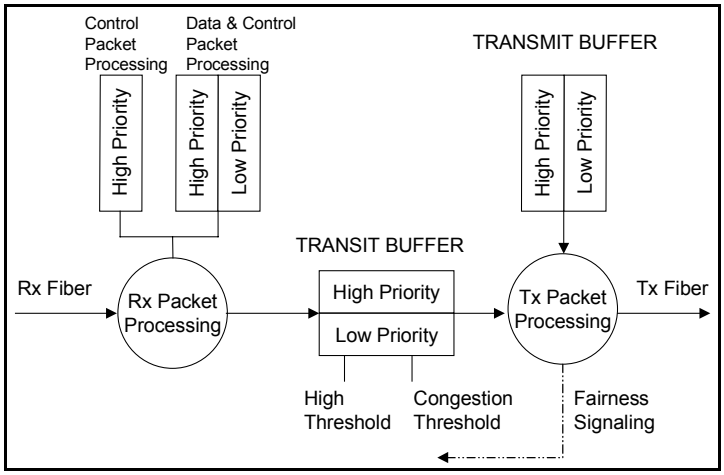


Figure 32: SRP Packet Processing Flow.

The figure shows only the operations made on a single ring (the outer or the inner). The other part of the dual-ring is symmetric. When a packet arrives to the node, it is checked its source and destination address (address lookup) and its mode (type of packet) since packets can either be data or control packets. Simplifying the reception algorithm:

3. If the destination MAC address corresponds to the node MAC address and the packet (data packet) is unicast, then the packet is copied to the reception buffer and stripped from the ring. If the packet is multicast then is forwarded since multicast packets need to be stripped by the source instead of the destination node.
4. If the destination MAC address match is not made or the packet (data packet) is multicast, the packet is placed into the transit buffer to be forwarded to the next node if the packet passes Time to Live and CRC tests. That means all packets which are not stripped in the node will be forwarded at the SRP-MAC layer directly and do not stress the corresponding IP-router in the node.
5. Control packets are always stripped once the information is extracted and copied to the reception buffer.

A transmission algorithm is needed since the packets can be sent either from the transit buffer (data generated by other nodes) or the transmit buffer (data generated by the own node) and both handling different priorities (high or low). The transmission algorithm is the following:

High priority packets from the transit buffer are always sent first.

High priority packets from the transmit buffer are sent as long as the low priority transit buffer is not full.

Low priority packets from the transmit buffer are sent as long as the low priority transit buffer has not crossed a threshold indicating this situation and the SRP-fa rules allow it.

If nothing else can be sent, low priority packets from the low priority transit buffer are sent.

It is worth noting that the SRP-fa contributes to the transmission algorithm controlling the access to the ring (step 3).

Multicasting

SRP provides direct support for IP multicasting. IP multicast uses class D address space and this class D multicast address is mapped to the appropriate 48-bit MAC address for transport on the ring.

Unlike unicast packets, multicast packets are source stripped. The multicast packets are placed into the transit buffer for continued circulation.

A.1.2.3.2 SRP Fairness Algorithm

The SRP-fa is a distributed algorithm that takes charge of ensuring:

- Global fairness: Each node gets a fair share of the ring bandwidth.
- Local optimization: Each node maximally leverages the spatial reuse properties of the ring
- Scalability: SRP-fa is able to handle efficiently large rings with many nodes. SRP-fa allows maximal 128 nodes and data rates up to STM-64c. This bandwidth is independent from the number of nodes and has to be shared between the nodes.

In addition, the algorithm is a preventive control of congestion. It is worth noting that the SRP-fa only applies to low priority traffic.

Simplifying, the algorithm works as follows:

- A set of usage counters monitor the rate at which low priority transmit data and forwarded data are sent (MY_USAGE and FWD_USAGE respectively).
- There is a congestion threshold in the low priority transit buffer (see Figure 32) used to detect congestion. If the congestion threshold is crossed (congestion detected), the node begins to advertise to upstream nodes the value of its transmit usage counter (MY_USAGE). In order to send this bandwidth information, Usage control packets are used. These usage messages are generated periodically even if there is no new bandwidth information to send (a null value is sent) since these packets also inform to the destination node that a valid data link exists.
- Nodes receiving usage messages adjust their transmit rates so as not to exceed the advertised value and propagate the usage messages received whenever the node is not congested. If the node receiving the usage information is also congested, propagate the minimum value of their transmit usage and the usage message received.

A more detailed explanation of the SRP-fa (included the SRP fa pseudo-code) can be found in the SRP Internet draft [DPT-3].

The SRP protocol utilizes a protocol known as Intelligent Protection Switching to provide the ability of the SRP ring (DPT ring) to recover from events and faults such as fiber cuts or node failures (see [WP2-D7] for details).

A.1.3 Gigabit Ethernet

A.1.3.1 Architecture

In comparison to previous Ethernet architectures, Gigabit Ethernet differs only in the physical layer and resembles them from the data link layer upward. In order to obtain a 1 Gbps speed, IEEE 802.3 Ethernet is enhanced by ANSI X3T11 Fiber Channel in the physical layer. What it does is adopting the high-speed physical interface of Fiber Channel, while maintaining the traditional Ethernet frame format.

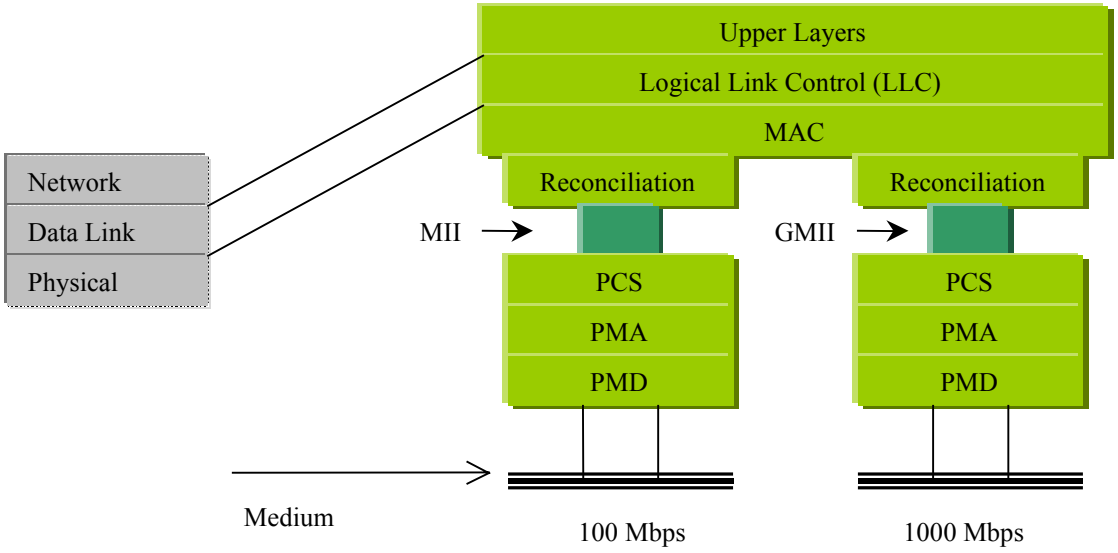


Figure 33: Architecture of IEEE 802.3z Gigabit Ethernet

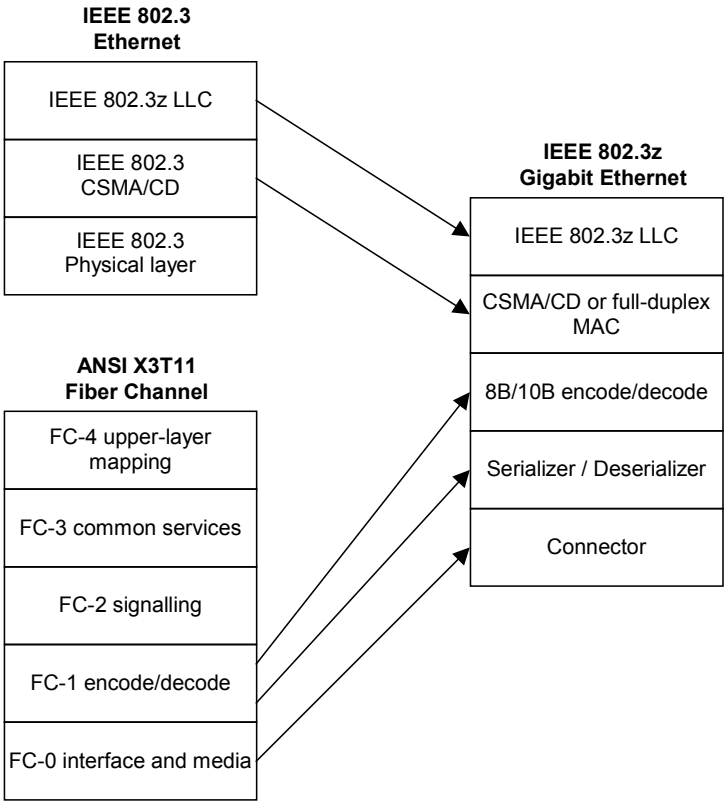


Figure 34: Comparison Ethernet, GbE and FC layered architecture.

A.1.3.1.1 Physical Layer

Gigabit Ethernet supports 4 physical media types. These are defined in 802.3z (1000Base-X) and 802.3ab (1000Base-T).

1000Base-X

The 1000Base-X standard is based on the Fiber Channel Physical Layer. Fiber Channel is an interconnection technology for connecting workstations, supercomputers, storage devices and peripherals. Fiber Channel has a 4-layer architecture. The lowest two layers FC-0 (Interface and media) and FC-1 (Encode/Decode) are used in Gigabit Ethernet. Since Fiber Channel is a proven technology, re-using it greatly reduced the Gigabit Ethernet standard development time. Three types of media are included in the 1000Base-X standard:

- 1000Base-SX 850 nm laser on multi mode fiber.
- 1000Base-LX 1300 nm laser on single mode and multi mode fiber.
- 1000Base-CX Short haul copper "twinax" STP (Shielded Twisted Pair) cable.

1000Base-T

1000Base-T is a standard for Gigabit Ethernet over long haul copper UTP. It provides 1 Gbps Ethernet signal transmission over four pairs of category 5 UTP cable, covering cabling distances of up to 100 meters or networks with a diameter of 200 meters. This standard will outline communications used for horizontal copper runs on a floor within a building using structured generic cabling, taking advantage of the existing UTP cable already deployed. This effort requires new technology and new coding schemes in order to meet the potentially

difficult and demanding parameters set by the previous Ethernet and Fast Ethernet standards.

A.1.3.1.2 MAC Layer

The MAC Layer of Gigabit Ethernet uses the same CSMA/CD protocol as Ethernet. The maximum length of a cable segment used to connect stations is limited by the CSMA/CD protocol. If two stations simultaneously detect an idle medium and start transmitting, a collision occurs. Ethernet has a minimum frame size of 64 bytes. The reason for having a minimum size frame is to prevent a station from completing the transmission of a frame before the first bit has reached the far end of the cable, where it may collide with another frame. Therefore, the minimum time to detect a collision is the time it takes for the signal to propagate from one end of the cable to the other. This minimum time is called the Slot Time. (A more useful metric is Slot Size, the number of bytes that can be transmitted in one Slot Time. In Ethernet, the slot size is 64 bytes, the minimum frame length.)

The maximum cable length permitted in Ethernet is 2.5 km (with a maximum of four repeaters on any path). As the bit rate increases, the sender transmits the frame faster. As a result, if the same frames sizes and cable lengths are maintained, then a station may transmit a frame too fast and not detect a collision at the other end of the cable. So, one of two things has to be done:

6. Keep the maximum cable length and increase the slot time (and therefore, minimum frame size) or
7. Keep the slot time same and decrease the maximum cable length or
8. both.

In Fast Ethernet, the maximum cable length is reduced to only 100 meters, leaving the minimum frame size and slot time intact. Gigabit Ethernet maintains the minimum and maximum frame sizes of Ethernet. Since, Gigabit Ethernet is 10 times faster than Fast Ethernet, to maintain the same slot size, maximum cable length would have to be reduced to about 10 meters, which is not very useful. Instead, Gigabit Ethernet uses a bigger slot size of 512 bytes. To maintain compatibility with Ethernet, the minimum frame size is not increased, but the "carrier event" is extended. If the frame is shorter than 512 bytes, then it is padded with extension symbols. These are special symbols, which cannot occur in the payload. This process is called Carrier Extension.

A.1.3.1.3 GMII (Gigabit Media Independent Interface)

The GMII is the interface between the MAC layer and the Physical layer. It allows any physical layer to be used with the MAC layer. It is an extension of the MII (Media Independent Interface) used in Fast Ethernet. It uses the same management interface as MII. It supports 10, 100 and 1000 Mbps data rates. It provides separate 8-bit wide receive and transmit data paths, so it can support both full-duplex as well as half-duplex operation.

The GMII provides 2 media status signals : one indicates presence of the carrier, and the other indicates absence of collision. The Reconciliation Sublayer (RS) maps these signals to Physical Signaling (PLS) primitives understood by the existing MAC sublayer. With the GMII, it is possible to connect various media types such as shielded and unshielded twisted pair, and single-mode and multi mode optical fiber, while using the same MAC controller. The GMII is divided into three sublayers: PCS, PMA and PMD.

PCS (Physical Coding Sublayer)

This is the GMII sublayer which provides a uniform interface to the Reconciliation layer for all physical media. It uses 8B/10B coding like Fiber Channel. In this type of coding, groups of 8

bits are represented by 10 bit "code groups". Some code groups represent 8 bit data symbols. Others are control symbols. The extension symbols used in Carrier Extension are an example of control symbols.

Carrier Sense and Collision Detect indications are generated by this sublayer. It also manages the auto-negotiation process by which the NIC (Network Interface) communicates with the network to determine the network speed (10,100 or 1000 Mbps) and mode of operation (half-duplex or full-duplex).

PMA (Physical Medium Attachment)

This sublayer provides a medium-independent means for the PCS to support various serial bit-oriented physical media. This layer serializes code groups for transmission and deserializes bits received from the medium into code groups.

PMD (Physical Medium Dependent)

This sublayer maps the physical medium to the PCS. This layer defines the physical layer signalling used for various media. The MDI (Medium Dependent Interface), which is a part of PMD is the actual physical layer interface. This layer defines the actual physical attachment, such as connectors, for different media types.

A.1.3.2 Standards

Gigabit Ethernet was standardised in June 1998.

To recap the recent history of the Gigabit Ethernet standards process, in July 1996, after months of initial feasibility studies, the IEEE 802.3 working group created the 802.3z Gigabit Ethernet task force. The key objectives of the 802.3z Gigabit Ethernet task force were to develop a Gigabit Ethernet standard that does the following:

- Allows half- and full-duplex operation at speeds of 1000 Mbps
- Uses the 802.3 Ethernet frame format
- Retains the CSMA/CD access method with support for one repeater per collision domain
- Addresses backward compatibility with 10BASE-T and 100BASE-T technologies

The task force identified three specific objectives for link distances: a multimode fiber-optic link with a maximum length of 550 meters; a single-mode fiber-optic link with a maximum length of 3 kilometers (later extended to 5 kilometers); and a copper based link with a maximum length of at least 25 meters. The IEEE is also actively investigating technology that would support link distances of at least 100 meters over Category 5 unshielded twisted pair (UTP) wiring.

A.1.3.3 Gigabit Ethernet and Asynchronous Transfer Mode (ATM) technologies

When ATM (Asynchronous Transfer Mode) was introduced, it offered 155 Mbps bandwidth, which was 1.5 times faster than Fast Ethernet. ATM was ideal for new applications demanding a lot of bandwidth, especially multimedia. Demand for ATM continues to grow for LAN's as well as WAN's.

ATM was touted to be the seamless and scaleable networking solution - to be used in LANs, backbones and WAN's alike. However, that did not happen. And Ethernet, which was for a long time restricted to LANs alone, evolved into a scalable technology.

As Gigabit Ethernet products enter the market, both sides are gearing up for the battle. Currently, most installed workstations and personal computers do not have the capacity to use these high bandwidth networks. So, the imminent battle is for the backbones,

the network connections between switches and servers in a large network. Both ATM and Gigabit Ethernet solve the issue of bandwidth. ATM provides a migration from 25 Mbps at the desktop, to 155 Mbps from the wiring closet to the core, to 622 Mbps within the core. ATM also promises 10 Gbps of bandwidth via OC-192, which was available and standard at the end of 1997. Ethernet currently provides 10 Mbps to the desktop, with 100 Mbps to the core.

ATM still has some advantages over Gigabit Ethernet:

- ATM is already there. So it has a head start over Gigabit Ethernet. Current products may not support gigabit speeds, but faster versions are in the pipeline.
- ATM is better suited than Ethernet for applications such as video, because ATM has QOS (Quality of Service) and different services available such as CBR (constant bit rate) which are better for such applications. Though the IETF (Internet Engineering Task Force, the standards body for internet protocols) is working on RSVP which aims to provide QOS on Ethernet, RSVP has it's limitations. It is a "best effort" protocol, that is, the network may acknowledge a QOS request but not deliver it. In ATM it is possible to guarantee QOS parameters such as maximum delay in delivery. Ethernet promises to provide Class of Service (CoS) by mapping priority within the network to mechanisms such as Resource Reservation Protocol (RSVP) for IP as well as other mechanisms for Internetwork Packet Exchange (IPX).

Gigabit Ethernet has its own strengths:

- The greatest strength is that it is Ethernet. Upgrading to Gigabit Ethernet is expected to be painless. All applications that work on Ethernet will work on Gigabit Ethernet. This is not the case with ATM. Running current applications on ATM requires some amount of translation between the application and the ATM layer, which means more overhead.
- Currently, the fastest ATM products available run at 622 Mbps. At 1000 Mbps, Gigabit Ethernet is almost twice as fast.

It is not clear whether any one technology will succeed over the other. It appears that sooner or later, ATM and Ethernet will complement each other and not compete.

A.1.3.4 Topology

Gigabit Ethernet supports full duplex operating modes for point-to-point connections, e.g. switch-to-switch and switch-to-server. Further, Gigabit Ethernet offers half duplex modes for shared connections using repeaters and the CSMA/CD scheme.

Table 7: Ethernet topology rules for maximum network distance

	Ethernet	Fast Ethernet	Gigabit Ethernet
Data rate	10 Mbps	100 Mbps	1000 Mbps
Cat 5 UTP	100 m	100 m	25 m
STP/Coax	500 m	100 m	25 m
Multimode Fiber	2 km	2 km	500 m
Single-mode Fiber	25 km	20 km	2 km

Gigabit Ethernet is essentially a "campus technology", that is, for use as a backbone in a campus-wide network. It will be used between routers, switches and hubs. It can also be

used to connect servers, server farms (a number of server machines bundled together), and powerful workstations.

Essentially, four types of hardware are needed to upgrade an exiting Ethernet/Fast Ethernet network to Gigabit Ethernet:

- Gigabit Ethernet Network Interface Cards (NICs)
- Aggregating switches that connect a number of Fast Ethernet segments to Gigabit Ethernet
- Gigabit Ethernet switches
- Gigabit Ethernet repeaters (or Buffered Distributors)

Two likely upgrade scenarios are given below:

A.1.3.4.1 Upgrade a shared FDDI Backbone

Fiber Distributed Data Interface (FDDI) is a common campus or building backbone technology. An FDDI backbone can be upgraded by replacing FDDI concentrators or Ethernet-to-FDDI routers by a Gigabit Ethernet switch or repeater.

A.1.3.4.2 Upgrade a Fast Ethernet Backbone

A Fast Ethernet backbone switch aggregates multiple 10/100 Mbps switches. It can be upgraded to a Gigabit Ethernet switch which supports multiple 100/1000 Mbps switches as well as routers and hubs which have Gigabit Ethernet interfaces. Once the backbone has been upgraded, high performance servers can be connected directly to the backbone. This will substantially increase throughput for applications which require high bandwidth.

A.1.3.5 10Gigabit-Ethernet

The huge of demand for high-speed networking has stirred development of the next generation 10-Gigabit Ethernet. The 10-Gigabit Ethernet standard will support the data rate of 10 Gbps, which is 10 times faster than the current transmission rate of 1-Gigabit Ethernet, but the cost is targeted around 2 to 3 times the cost of the current 1-Gigabit Ethernet technology. It will not support the half-duplex operation, but it will maintain the compatibility with the preceding Ethernet technologies by using the same Ethernet frame format, enabling seamless integration among the existing networks and the new technology. 10-Gigabit Ethernet may be used in LAN, MAN, and WAN, implying technology convergence and faster switching.

In addition to the data rate of 10 Gb/s,10-Gigabit Ethernet shall be able to accommodate slower date rates such as 9.584640 Gbps (OC-192). This is likely to be done by using the word-by-word pacing mechanism via the 10GMII interface. The current proposal for 10GMII uses 32-bit data paths with 4 control bits (one control bit per one data byte) and a TX_hold line. This structure provides scalability and can support various variations of the physical layer implementation. The discussions regarding issues in the physical layer are still ongoing. The standard may provide different specifications for different applications. The main considerations include the implementation architectures (serial, parallel, or WDM), coding techniques, supporting technologies, and media.

The 10 Gbps serial transmission solution appears to be the easiest and lowest-cost option. The recent technology demonstration by Lucent validates the 10 Gbps serial transmission using 850-nm VCSEL sources on the enhanced multi-mode fibers over the link longer than 300 meters. However, it cannot support the existing multi-mode fiber infrastructure. On the other hand, the WWDM solution supports the existing multi-mode fiber infrastructure but the implementation is more complicated and the devices are still more expensive. The 4 x 3.125

GBaud WDM solution using 1300-nm uncooled DFB sources on the existing 62.5-um multi-mode fiber can reach the link distance up to 300 meters. In addition to the WDM solution, the parallel fiber option may be suitable for short-haul applications such as computer rooms. In addition to fibers, copper wires may be used for a very short link (10-20 meters) but may not be considered in the early version of the standard.

The coding issue still remains opened. Several coding techniques are considered. The trend is that the 8B/10B encoding technique may be specified for the LAN rate Ethernet because it is compatible with 1-Gigabit Ethernet, while the SONET-like scrambled encoding techniques may be specified for the WAN rate Ethernet because it is compatible with SONET. These coding techniques are proven and well understood, but the problem with the 8B/10B is the large 20% overhead. Other coding techniques such as 15B/16B, MB810, and PAM-5 are considered as well.

In short, 10-Gigabit Ethernet will be a low-cost solution for high-speed and reliable data networking and it will dominate the LAN, MAN, and WAN markets in the near future.

A.1.3.6 Conclusion

Gigabit Ethernet is a viable technology that allows Ethernet to scale from 10/100 Mbps at the desktop to 100 Mbps up the riser to 1000 Mbps in the data center. By leveraging the current Ethernet standard as well as the installed base of Ethernet and Fast Ethernet switches and routers, network managers do not need to retrain and relearn a new technology in order to provide support for Gigabit Ethernet.

APPENDIX 2
IP Service Level Agreements

In a corporate environment, the IT department takes on the role of network provider, and various internal departments take the position of the customers. Applying certain use policies, so called Service Level Agreements (SLAs), are already common in modern IT strategies. For example, in many enterprises, it is no longer acceptable that employees' personal web traffic takes the same priority as crucial business-critical traffic, such as order-entry systems or SAP traffic. It must be possible to differentiate the various types of traffic on a private network and prioritise them according to some global strategy.

When dealing with public network services, with the increasing importance of IP-based communications for business, guarantees about the behaviour of the service are becoming more and more important. Offering an IP transport service, if an IP network provider is to make some form of guarantee to the customer, a SLA must be formed. In this case a SLA is a contract between the customer and the provider. The customer lays out their requirements for the network, and the provider makes some level of guarantee to provide those requirements. The guarantee may not be 100%, but could take the form of a probability, for example: the requirements will be met 99.9% of the time, over a one-month period.

Based on the parameters qualifying the IP transport services described above Table 8 gives an example of a level of guarantee for a possible set of IP transport service classes that currently could be offered by a Network Operator. Three classes are described, namely GOLD, SILVER and BRONZE:

- The GOLD service represents the transport class providing the best performances for all potential applications. It is foreseen for supporting very time sensitive applications.
- The SILVER service is a transport class for supporting time sensitive applications.
- The BRONZE service is the transport class for applications that require no particular performances.

Table 8: Example of parameters qualifying an IP Transport Service

PARAMETERS	GOLD	SILVER	BRONZE
Availability	99.97	99.95	99.95
End-to-End delay	90 ms	150 ms	
Packet Loss Ratio	2 %	5 %	10 %
Security	-	-	-
Provisioning Time	-	-	-
Access	2 - 34 Mbit/s	N x 64 kbit/s	ISDN

Note that the only network parameter for which a valuable guarantee is offered is the End-to-End delay. The announced Packet Loss Ratio is very poor even for the GOLD service class. Also note that no security guarantee is offered.

The case presented in Table 8 is, at present, the most common situation: Some ISPs are already offering SLAs to their customers, but very limited. One major ISP is offering their customers guarantees on timely installation, and 100% availability. The only network parameter for which they offer a guarantee is latency - "Average network latency of less than

85 ms roundtrip between designated U.S. hubs and an average latency of 120 ms roundtrip or less between Trans-Atlantic gateway hubs located in New York and London". As we can see, only average latency times are guaranteed, as the ISP's connections could still congest for short periods. Simple over-provisioning is used here. If stronger network guarantees could be offered, on parameters such as bandwidth, delay or jitter, the IP provider would have many new services to offer. To illustrate this case Table 9 includes the IP transport services level of guarantee provided by two different USA network operators, UUNET and AT&T WorldNet.

Table 9. Two real live examples of SLAs.

UUNET	Internet Access Guarantee	Internet Access Make Good	IP-VPN Service Guarantee	IP-VPN Service Make Good
Network Availability	100%	1 day credit for up to each hour downtime	99.9% 99.8% 99.6%	25% credit on entire VPN MSC
Maximum Latency	85 ms (US) 120 ms (int)	1 day credit after 30 day grace make good period	100 ms (US) 200 ms (int)	25% credit on entire VPN MSC
Maximum Packet Loss	No			
Local Access included in Guaranties	Yes		Yes	
AT&T WorldNet	Internet Access Guarantee	Internet Access Make Good	IP-VPN Service Guarantee	IP-VPN Service Make Good
Network Availability	99.5%	1/30 of MSC refund per 15 min. downtime per day; Up to 100% MSC refund per quarter	99.7%	5% MSC refund per 10 min downtime per day; up to 25% MSC refund
Maximum Latency	N/O	N/O	100 ms (US) 250 ms (int)	1/30 of MSC credit
Maximum Packet Loss	N/O	N/O	N/O	N/O
Local Access included in Guaranties	Yes	N/A	Yes	N/A

SLA metrics

From above, it can be stated that one important aspect which form part of the SLA is the QoS characteristics required for the service. The QoS requirements of the SLA can be categorised into three main groups:

- Bandwidth parameters, describing the bandwidth requirements of the service. An example of these is average throughput.
- Delay parameters, describing the delay requirements of the service. Examples of these are average delay and jitter.
- Loss parameters, describing the loss requirements of the service. Examples of these are average packet loss and packet loss sensitivity.

- The SLA also covers all aspects of the customer-provider relationship, which are relevant to the service. Some metrics that SLAs may specify include:
- What percentage of the time services will be available
- The number of users that can be served simultaneously
- Specific performance benchmarks to which actual performance will be periodically compared
- The schedule for notification in advance of network changes that may affect users
- Help desk response time for various classes of problems
- Dial-in access availability
- Usage statistics that will be provided

Once a SLA is agreed, the provider has committed to deliver a QoS and must deliver on that commitment. When the customer and provider have agreed a SLA, it is the responsibility of the provider to monitor and maintain that SLA. If the SLA is not met, the customer is entitled to some form of compensation. This is really a form of insurance policy, which - like most insurance policies - is a last resort for the holder - the customer does not want this money back, they want their service to perform as expected.

In order to minimise customer compensation (increasing profitability and increasing customer satisfaction), management applications for SLA and Quality of Service management are required by the provider. Management of the network must attempt to ensure that SLAs are respected at all times. If this is not possible, the violated SLAs must be prioritised.

This requires tools for measuring the performance of the network, analysis of the performance information to identify violations of SLAs, and the provision of decision support capabilities for aiding the provider in the resolution of such problems. Such tools must be multi-vendor (capable of managing network elements from a range of vendors in a consistent manner) and be capable of SLA management in a multi-provider domain.

APPENDIX 3
Enabling technologies

In order to achieve the goal of developing a new transport network, it is necessary that WDM completes its transition towards a fully functional OTN. This is not an unrealistic objectives, but some steps are still to be done.

In this section, a wide meaning will be assigned to the term "enabling technologies", which will be classified in two groups:

- Technologies for the physical realization of the new generation of OTN transport nodes
- Technologies for overcoming the still open issues.

A.3.1 Technologies for OTN transport nodes

Optical network nodes can use space, time and wavelength to transport and switch optical channels. Networks can be categorised depending on the lightpath establishing techniques they use:

- Circuit-switched: the circuit route is fixed and is established before transmission.
- Virtual circuit-switched: the route is fix but statistically allocated
- Datagram-switched: the route is selected per packet during transmission, being statistically allocated

Optical nodes participating in the establishment/release of light-paths can be distinguished depending on the switching technique. Thus, they are called cross-connects when the connection pattern is semi-permanent and light-path set up is done by the network operator. They are referred to as switches (intelligent nodes) when light-paths are established using signalling protocols. Routers are nodes that perform connectionless switching of datagrams.

The core of a switching optical node is the Switch Matrix. It can be optical or electrical. An electronic Switch Matrix uses optical to electronic conversion at the input and electrical to optical conversion at the output. This nodes are known as Opto-Electric Cross-Connects (OEXC) – or Opaque Optical Cross-Connects. The Switch Matrix can also be optical and then, many different options can be used.

The main components of an Optical Node are here briefly recalled, and the consolidated and emerging technologies are mentioned, with a particular focus on the switch matrix.

Table 10: Optical technologies for Optical Nodes

Functional Block	Consolidated Technology	Advanced Technology
WDM Mux/Demux (as for WDM line systems)	micro-optics diffraction gratings cascade of interferential filters cascade of Fibre Bragg Gratings (FBG)	Arrayed WaveGuides (AWG) built as Planar Lightwave Circuits (PLC)
Optical Amplifier	Erbium-Doped Fibre Amplifiers - EDFA (optimized for multi-channel WDM, also through external control gain systems) Semiconductor Optical Amplifiers - SOA (on single channels, can be used as "on/off gates")	Different dopings for extended bandwidth e.g Erbium-Doped Fibre Fluoride Amplifiers - EDFFA the objective is to have two usable bandwidths, C, L, over a whole range between 1530 and 1620 nm SOA for multi-wavelengths channels
Wavelength Converter	Based on O/E/O conversion	All-Optical Wavelength Converter (AOWC) built by SOA-based interferential structures

		(this is not a short term solution)
Splitter / Combiner	Fibre-based splitter/combiner	Planar Lightwave Circuit (PLC) in SiO_2/Si
Tunable Filter (e.g., for split & select nodes)	Mechanic. tunable Fibre Fabry-Perot (FFP) Mechanic. tunable interferometric filter (angle-tilting)	thermo-optically tunable PLC in SiO_2/Si acousto-optically tunable in $LiNbO_3$
Add/Drop Filter	Fibre Bragg Gratings (FBG) based	thermo-optically tunable in SiO_2/Si
Var. Attenuator	thermo-optical PLC in SiO_2/Si	
2x2 switch (e.g. for protection switching)	micro-optics electro-mechanical (MEMS, e.g., based on a movable cantilever) piezo-electric electro-mechanical passive PLC in $LiNbO_3$	thermo-optical PLC in SiO_2/Si active PLC in InP micro-mirror based
1xN selector (N < 8)	micro-optics electro-mechanical (MEMS, e.g., based on a movable cantilever)	polymeric waveguide based
Switch Matrix	Bulk electro-mechanical (not larger than 64x64) Micro-Optics Electro-Mechanical	Micro-Mirror based matrix *(see section A.4.1.2.4)* Intersecting Waveguides + InkJet bubbles Active matrix based on SOA gates

Table 11: Electronic technologies for Optical Nodes

Functional Block	Consolidated Technology	Advanced Technology
Switch Matrix	Cross-point matrix Open issues are: max. frequency that can be switched (suitable to 2.5 Gb/s ch. not to 10 Gb/s ch.) electronic interconnection between matrix units (short reach optics maybe required)	Cross-point matrices able to handle higher frequencies (consistent with 10 Gb/s channels) with optimized backplanes for electronic interconnection (at least up to 2.5 Gb/s channels)

A.3.2 Technologies for advanced network functionalities

From the viewpoint of the technologies which are necessary, two items emerge:

- the Optical Supervisory Channel technology,
- the Digital Wrapper technology.

A.3.2.1 The Optical Supervisory Channel technology

The Optical Supervisory Channel (OSC) is an idea derived from the WDM line systems (e.g., see the products mentioned in section A.4.1.1). The OSC in line systems is a means for remote monitoring of line stations; it can also be used to carry service communication channels. Typically, the OSC is a low bit-rate channel (e.g., 2 Mb/s) modulated onto a wavelength out of the EDFA amplification bandwidth, to avoid the under-exploitation of an in-band wavelength (better used as a payload). The OSC can be in the 1310 nm band, or at 1510/1520 nm, or around 1625 nm. Obviously, it is extracted before and re-inserted after EDFAs, which are not transparent to the OSC wavelengths.

When evolving from point-to-point WDM to OTN, the OSC may play a more important role:

- it keeps the function of remote monitoring of line sites, properly enhanced to exchange also information related to the remote management of sites where no local NE Agent is present
- it supports optical OA&M functions (maintenance messages, protection protocols, etc.)
- more generally, it transports the overhead information related to the OTS and OMS optical network layers, and that part of the OCh layer overhead which is not strictly associated with the channels itself (e.g., OCh OA&M messages) – the part of OCh overhead strictly associated with each optical channel needs another type of support (e.g., the Digital Wrapper)
- in the evolutionary scenario towards the Intelligent OTN the OSC is a possible means to carry optical routing protocol information.

A.3.2.2 The Digital Wrapper technology

A requirement of the WDM-OTN is that client layer provides a continuous bit-stream. Therefore packet based client networks cannot be mapped onto the server layer directly. The trend for OTN server/client network interworking is to define a common frame structure in the optical channel layer to carry clients regardless of their type. With state-of-the-art technology, WDM-OTN networks are unable to provide neither all-optical 3R (re-amplification, reshaping, and re-timing) regeneration nor QoS guarantee using optical protocols to client signals. Therefore, a highly desirable feature of a TDM optical container is to provide somehow means for signal quality monitoring analogously to SONET/SDH.

The Digital Wrapper (DW) is a Lucent's proposal. It comes after results of different studies that have shown the convenience of carrying the associated optical channel overhead in a TDM format. DW structure is based on the ITU-T Rec. G.975 and is aimed at carrying the OCh-OH as well as an FEC (Forward Error Control) and possibly ring protection on a per-wavelength basis. DW is a means to effectively manage wavelengths or Optical Channels (OCh), irrespectively from the type of client. Until now, the only feasible way to support signal regeneration and to monitor, analyze and manage optical channels (wavelengths) was to rely on SONET/SDH signals and equipment throughout the network. However, as already noted several times, it will be required to the new OTN to be a “universal” platform, able to support the multiplicity of "IP over WDM" signal mappings, i.e., to reliably carry a wide variety of client signals-including SONET/SDH, IP, ATM, GbE, and SDL-directly over the optical layer of the network.

The Digital Wrapper technology will provide functionality and reliability similar to SONET/SDH, but potentially at a lower cost and without adding more equipment to the network.

APPENDIX 4
Product overview

A.4.1 IP-over-WDM

A.4.1.1 IP over WDM as transmission technology

Some of the most advanced line system products, by major manufacturers, are summarised in [NR-P&T], as significant examples. Reported there data had been considered just as a qualitative information – drawn from different public sources – and obviously not exhaustive. The exact commercial situation (what is already deployed, what is ready to be sold, what is at prototype level) is not specified in detail, because analytical comparisons between different products is out of the scope of that document. Anyway, that rough overview should be sufficient to have an idea of the state of the art and of the ongoing short-term evolution.

The following advanced line system products had been analysed as examples in [NR-P&T]-p.6.1.1: Lucent: WaveStar OLS 400G, Cisco Photonics: TeraMux HyperDWDM, Nortel: OpTera LH, Ciena: MultiWave Sentry 4000 and MultiWave CoreStream, Siemens: TransXpress Infinity, Alcatel: OPTINEX 1640 WM.

A.4.1.2 IP over OTN as a new transport network

A non-exhaustive, brief overview of the emerging products by major vendors, in the area of optical transport nodes, is provided in [NR-P&T] - Appendix 1 p. 6.1.2.

A.4.1.2.1 Transport Nodes for the Core Backbone Network

In traditional networks, the physical layer is unchanging unless a manual reconfiguring or fiber reconnection is performed. This method is disadvantageous, expensive and error-prone. Moreover, it cannot face the faster and faster network evolution, often hardly predictable. The answer to all of this is to provide the flexibility of switching high-speed optical channels in the core network. This functionality is provided by new kinds of transport nodes i.e. optical cross-connects (OXC) and optical add-drop multiplexers (OADM).

These optical transport nodes provide several benefits. They offer a protected, restorable optical layer transport network of wide flexibility. They allow to achieve lower network cost and higher fiber efficiency. A typical traffic pattern at a node in the core of a network is that between 50 and 75 percent of the traffic is "through" traffic. Such traffic can be routed straight through a cross-connect without a need for any additional processing or switching. Instead of devoting high-cost ports on DXC or IP Switches to such traffic, much lower cost optical cross-connect switches can be used to express route the trough traffic. Cost savings of 30 to 40 percent are projected. Operations costs are also reduced because of the automated and remote control functions available.

Optical transport nodes also enable ring network interworking and the building of networks of complex ring and mesh architectures. Networks can be configured in the same cross-connect on a wavelength-by-wavelength basis: several wavelengths, or channels, can be configured as self-healing rings, while other channels are connected as meshes. Protection and restoration protocols are selectable on a wavelength-by-wavelength basis.

There are different ways to pursue these objectives:

One means is to use large port-count digital cross-connect switches and many time division multiplexers. This is an expensive approach. It also is difficult to perform restoration and

protection switching in reasonable time limits in large networks because the switching action is done on sub-rate channels, sequentially, rather than at the high-speed channel rate. Configuration or provisioning of high-speed interconnection paths requires the coordinated switching of all of the sub-channels and re-assembly.

On the other hand, all-optical switching fabrics are proposed as a means of cross-connecting signals regardless of bit rate or data format, therefore providing a "future proof" design. These would be designed as optically transparent switches. However, the actual state of the art of all-optical switches is that they are expensive, and they are of insufficient port count to be competitive with electronic fabrics.

The alternative is to deploy "electronic" cross-connects at the optical layer, able to switch at the line rate of the input channels, providing restoration and protection switching on a full OC-48 channel in one operation and avoiding the need for large banks of TDM multiplexers. The size of the required cross-connect is also commensurately reduced, both in port count and in physical dimension. Management of the optical layer is done at the optical layer and directly on each of the optical channels.

These optical cross-connects, so called because they operate at the optical line rate and on whole optical channels at a time, require a significant amount of information about the optical channels in order to perform fault management, protection switching and restoration, and to verify connectivity through a switch. This can be done by accessing the various overhead bytes of the SONET carrier, or by introducing a new kind of digital overhead, associated to the Optical Channel (Digital Wrapper). Both solutions lead to the use of electronic switching fabrics as the most cost-effective way of implementing the cross-connect switch function.

Therefore, among the products for the core backbone the following types of nodes are considered:

"Multi-service" nodes, directly evolving from SDH DXC nodes e.g., Lucent WaveStar Bandwidth Manager, Tellabs Titan 6500, Alcatel 1680 OGX

Opaque OXC: wavelength cross-connection, using electrical core switching matrix e.g., Ciena Lightworks CoreDirector, Cisco ONS15900, Lucent Aurora512, Nortel OpteraConnect, Sycamore SN16000

Transparent OXC:

- Automatic large size optical matrices (e.g., Lucent LambdaRouter, Nortel/X-Ros)

- Manually reconfigurable OXC *plus* automated protection, remote monitoring (e.g., Siemens OSN)

A.4.1.2.2 Multi-Service nodes for the Core Backbone Network

A starting point for the discussion can be a look on the classical broadband DXC, optimized for SDH (excellent examples of this category are the ***Siemens SXD*** and the ***Nortel S/DMS*** nodes).

Typical feature of these nodes are:

large-size synchronous switching matrices – optimized for compactness (for example, 2048 – 4096 STM-1 equivalent, in a single rack),

capability of working on high level Virtual Containers (typically VC-4, at a STM-1 line rate of 155 Mb/s),

their role is at the upper level of an SDH-compliant core switching network, whereas a set of different Network Elements (often belonging to the same vendor's product family) are

optimized for different functions (regenerators, terminals, add-drop multiplexers, wideband cross-connects, etc.),

capability of supporting OC-48/STM-16 interfaces towards the high-capacity transmission systems,

often already equipped with (or at least pre-disposed for) OC-192/STM-64 interfaces,

often already equipped with (or at least pre-disposed for) coloured DWDM-compliant interfaces,

operations on concatenated interfaces/channels possible,

support to different protection schemes developed for SDH (linear MSP, SNCP, ring MSP, BSHR, UPSR) – to ensure survivability,

optimized OA&M and Network Management features, through the capability of handling DCC for all the throughput channels.

However, despite their good features, these "pure" SDH nodes cannot keep the pace with the evolving requirements for transport network.

The first category of "new nodes" include the "multi-service" nodes, directly evolving from SDH DXC nodes. The main additional features in these nodes may typically be:

the capability of switching aggregates at large granularities (comparable to the wavelength granularities) although in a synchronous electronic domain,

the capability of switching lower granularities, with respect than the classical VC-4 of broadband DXC,

the possible integration of layer 2 switching and layer 3 routing sub-systems,

the addition of local intelligence for protection/restoration, beyond the network management features.

This is the roadmap preferred by manufacturers already present in the SDH and in the transport networking market. Different evolutionary paths are pursued. The following three examples of different and original trends, (Lucent WaveStar Bandwidth Manager, Tellabs Titan 6500 MTS, Alcatel 1680 OGX) are reported in [NR-P&T]-p. 6.1.2.

Optical Gateways: Where optical layer backbone and access networks interconnect, optical gateways and photonic cross-connects will emerge. The optical gateway is a necessary element of true optical networking, providing a platform with integrated Dense Wave Division Multiplexing (DWDM) optics while internally managing broadband services in the electrical domain. Gateways become the bridge between flexible access networks and highly efficient backbone networks. Optical gateways will displace the broadband cross-connect to manage the transparent to opaque conversion between the access and long haul networks, and to manage the broad range of payloads in wavelength services. They will simultaneously support cell based routing and aggregation of ATM or IP payloads and legacy STM services. In this way, lower rate wavelengths from the access will be aggregated to high speed ITU compliant channels for the backbone network. And as DWDM systems grow in channel count (currently projected up to 240 channels), optical gateways will emerge as the central wavelength service management device.

A.4.1.2.3 Opaque Optical Cross-Connects for the Core Backbone Network

The second category include the "Opaque" OXC, i.e., cross-connect nodes able to perform "wavelength management" and switching on wavelength granularity, able to handle optical channel overhead (pre-disposed for Digital Wrapper techniques), which have however

an internal electronic core. It must be highlighted that, although electronic, the switching matrix is an asynchronous space switch, and has the only limitation of the maximum bit-rate that can be handled electronically (today, up to 2.5 Gb/s).

This category is today the most important one, since it allows the realization of large-size matrices, and is consistent with the necessity of 3R regeneration, digital performance monitoring, channel-associated digital overhead.

The principles behind this choice are:

Networks are Becoming Data-Centric,

- High- speed Data Signals do not need STS- 1 grooming,

- WDM has made bandwidth plentiful,

- managing OC- N is still difficult & expensive.

Mesh Protocols will dominate the Optical Network Core.

OC- 48 is the "brick" for building Optical Networks.

SONET monitoring is essential for intelligent layer 1.

Opaque will continue to win over Transparent.

Network Management is key to survivable Optical Layer.

Many vendors think that the architecture of preference for an optical cross-connect is that of an electronic switching fabric, with input and output interface boards that provide optical detection and regeneration, and overhead processing. All of these functions are performed at the line rate of the optical channel, most commonly at OC-48 (SDH-16) rates. Extensions to OC-192 interfaces and switching will be developed soon. The following product lines are presented as example in [NR-P&T]-p.6.1.2: Ciena: MultiWave CoreDirector, Cisco: ONS15900, Lucent: Aurora 512, Nortel: OpTera Connect, Sycamore: SN16000.

A.4.1.2.4 Transparent Optical Cross-Connects for the Core Backbone Network

The third category include the "Transparent" OXC, i.e., cross-connect nodes able to perform "wavelength management" and switching on wavelength granularity, through optical core matrices.

Two different levels of transparency can be identified:

Complete transparency – avoid 3R regenerators at all, save costs, allow end-to-end transparency for each type of client – but problems related to transmission performances and overhead remain

Core switching transparency – 3R Reg. and electronic overhead processing are still present, but the internal optical core allow to properly switch 10 Gb/s channels, differently from the electronic core.

In any case, the optical switching matrix of sufficient dimensions is required.

No consolidated products are available in this category. The focus is still on the availability of a core optical matrix characterised by proper dimensions.

Two different approaches are considered in the following examples:

Focus on technology, to achieve the objective of a fully functional large size Optical Cross-Connect (a "true" OXC) – this is pursued by Lucent (LambdaRouter), Nortel (former X-Ros technology), Agilent; what is offered is not yet a complete product but a solution for the optical matrix, that is the main bottleneck for the realisation of the whole system..

Focus on a smooth evolution from what is possible right now – this is pursued e.g., by Siemens– "optical" features are progressively added to existing nodes, accepting intermediate steps in the short term (e.g., manual cross-connection or wavelength Add/Drop - rather than full cross-connect - capability).

[NR-P&T]-p.6.1.2 briefly describes following technology for transparent optical cross-connects solutions: Lucent: LambdaRouter, Nortel/X-Ros: X-1000, Agilent Photonic Switching technology, Siemens: OSN.

A.4.1.2.5 Nodes for the Metropolitan Transport Network

The argument for WDM as simply a network infrastructure tool is not as compelling for metropolitan or other short haul networks. It must have more benefit to the operator to gain acceptance into the network. However, the appearance of wavelengths in both long haul and access networks opens new revenue generating service opportunities.

While it is sometimes assumed that the widespread deployment of DWDM in core networks will inevitably spread into short-haul optical transport networks, very different issues come into play if DWDM is to benefit from widespread application beyond fibre exhaust situations. The cost savings that result in core networks from replacing electrical regenerators with optical amplifiers are largely irrelevant in short-haul optical transport networks, especially metropolitan areas, due to the limited distances involved.

The need to add/drop traffic from the metro network at many locations places a premium on cost-effective optical add/drop rather than maximum channel density.

Another driver for metro DWDM deployment may be the provisioning of dedicated wavelengths for high-speed access. For customers who require very high access bandwidths, above OC-12/STM-4 for example, providing a dedicated wavelength over a DWDM ring may become an attractive alternative. Today these requirements are often addressed by point-to-point fibre between customer and central office, by two node SONET/SDH rings, or by allocating a large portion of a SONET/SDH ring. DWDM access rings may provide a cost-effective answer to this challenge as the requirement for reliable gigabit access grows.

Standalone metro DWDM systems are already available for ring and point-to-point applications. However, these systems are offered at a high cost-per-wavelength and require additional platforms to deliver services at DWDM nodes. Integration of DWDM into optical transport equipment provides a cost-effective way for service providers to scale beyond the speed limitations of serial optics and meet increasing bandwidth requirements.

Metropolitan, interoffice and access networks consist of a diverse range of architectures, bit rates, traffic patterns and protocols, in contrast to the long distance environment that is evolving from an established, standardized synchronous optical network (SONET) base. The economics of WDM compared to alternative solutions are, therefore, much more complex.

Cost is clearly a major issue. Unlike the long distance environment, there has been no obvious "killer application"--an economically compelling reason to deploy WDM in the majority of cases, over all other solutions.

In the short term, therefore, the decision to deploy WDM in the local environment appears to be driven by simple network congestion and cost considerations, where WDM is cheaper and/or more readily available than the alternatives for increasing capacity, then it will be deployed. This is occurring in those networks that are "fibre poor", i.e. where existing fibre capacity is running out and there is a lot of traffic on interoffice facilities. On the contrary, in "fibre rich" networks, where there are spare conduits for laying new fibre and/or network congestion is not so acute, WDM may not be deployed widely in the near term.

However, despite the relative lack of success thus far for metro WDM, simply looking at the technology in terms of point-to-point capacity increase is misleading. Optical networking, deploying WDM ring architectures as opposed to simple point-to-point WDM, potentially will provide a common method for transport regardless of signal format in the metro environment, giving Local Exchange Carriers (LEC) the ability to deliver new services more quickly while reducing the costs associated with delivery.

For example, service providers will be able to assign the end customer a set of wavelengths and route those wavelengths through the network independent of whatever service is being carried. The marketing of transport on a wavelength basis will enable operators to offer end users a way to enter the high-speed backbone in native formats, avoiding the costs of aggregating multiple service types.

Because of the reasons expressed in the previous rationale, the commercial availability of optical nodes for metropolitan systems is one step behind with respect to the situation valid for the core network. Nonetheless, some examples can be mentioned, by looking at medium capacity / medium throughput products, optimized for ring topologies (typical of metro networks).

Two lines of evolution can be identified:

Smooth evolution from DWDM line systems, which gradually offer linear and then ring wavelength add/drop capabilities (some of these products may be found in section A.4.1.1, where add/drop capabilities offered by commercial DWDM line systems have been highlighted, e.g., in Lucent OLS 400G, Nortel OpTera LH, Ciena MultiWave CoreStream, Alcatel 1640 WM)

Development of "optical nodes" – in parallel to the ones proposed for the core transport network – whose features are optimized to the metro transport environment (smaller size but also lower cost, ring instead of mesh topology, wavelength add/drop multiplexing rather than full wavelength cross-connection/routing). The optical nodes are in this case Optical Add Drop Multiplexers (OADM), Four examples of these systems will be briefly considered in [NR-P&T]-p.6.1.2 (Marconi: Smart PhotoniX PMA8, Ciena: MultiWave Metro, Nortel: OpTera Metro, Siemens: WaveLine).

A.4.1.3 IP over "intelligent" OTN

An overview of the products in the area of Optical Transport Nodes has already been done in previous section, where the availability of intelligent protocols has been highlighted among the node properties. Existing intelligent nodes mainly belong to the category of Opaque OXC.

The emerging "Intelligent Optical Nodes" are just recalled in the following list:

Ciena MultiWave CoreDirector with Optical Signaling and Routing Protocol (OSRP),

Cisco ONS 15900 Wavelength Router with Wavelength Routing Protocol (WaRP),

Lucent 512 Aurora with StarNet (Signaling Algorithm for Restoring Networks),

Sycamore SN16000 with automatic topology discovery and optical routing algorithms (e.g., OSPF), LDP for circuit set-up.

A.4.2 A survey of optical networking products for the transport network

The results deriving from survey of optical networking products for transport network are presented in [NR-P&T]-p.6.2. This section lists manufacturers in alphabetical order and gives short descriptions of the main products pertaining to the optical networks area of each company.

Below we are listing main manufacturers of products and solutions for the long haul transport network. [NR-P&T]-p.6.2.3. contains short descriptions of the main products or families of products offered by each vendor.

Main manufacturers:

ALCATEL (http://www.alcatel.com, http://www.alcatel.europe.fr)
AVICI (http://www.avici.com)
Charlotte's Web Networks (http://www.cwnt.com)
CIENA (http://www.ciena.com)
CISCO (http://www.cisco.com)
CORVIS (http://www.corvis.com)
ECI Telecom Ltd. (http://www.ecitele.com)
ERICSSON (http://www.ericsson.com)
FOUNDRY Networks (http://www.foundrynet.com)
FUJITSU Network Communications (http://www.fnc.fujitsu.com)
HITACHI Telecom (http://www.hitel.com)
JUNIPER Networks (http://www.juniper.net)
LUCENT (http://www.lucent-optical.com)
MARCONI Communications (http://www.marconi.com)
NEC (http://www.necpng.com)
NOKIA (http://www.nokia.com)
NORTEL Networks (http://www.nortelnetworks.com)
PIRELLI TELECOM SYSTEMS Division (http://www.pirelli.com)(now acquired by CISCO)
PLURIS Inc. (http://www.pluris.com)
SIEMENS (http://www.siemens.com)
SYCAMORE Networks (http://www.sycamorenet.com)
TELLABS (http://www.tellabs.com)
TELLIUM (http://www.tellium.com)
Other manufacturers:
ALIDIAN Networks (ex TERABIT Net.) (http://www.alidian.com)
ARGON (purch. by Unisphere Solutions) (http://www.argon.com)
ASTRAL POINT Communications (http://www.astralpoint.com)
ATMOSPHERE NETWORKS (http://www.atmospherenet.com)
MAYAN Networks (http://www.mayannetworks.com)
OSICOM TECHNOLOGIES (http://www.osicom.com)
OPTICAL NETWORKS Inc. (http://www.opticalnetworks.com)
QEYTON Systems (http://www.qeyton.com)
SILK ROAD Corp. (http://www.silkroadcorp.com)
Zaffire (http://www.new-access.com) (previously New-Access Communications)

A.4.3 DPT Applications and Products

The DPT system can be used in LAN, MAN and WAN ring applications due to its new MAC layer protocol (SRP protocol). But it is in the local and the metropolitan environment where the DPT technology is more interesting.

The metropolitan area market, concretely the metro DWDM market, can be divided at least in three different segments [DPT-4]:

Metro Core: Connections are between carriers PoPs and do not directly interface with end users.

Metro Access: Consists of the rather nebulous segment between carriers' PoPs and access facilities.

Enterprise: Includes multichannel systems used to create high-bandwidth, multiprotocol links between sites within an enterprise.

DPT can be used in all these three segments and an enumeration of some DPT applications is listed below:

IntraPOP connectivity for leased-line services aggregation and IntraPOP connectivity via high-speed regional rings (Metro Core segment). The DPT provides an ideal solution since uses efficiently the bandwidth and eliminates the complexity associated with intermediate layer 2 switching solutions, among other benefits.

Metropolitan area IP rings for shared business and residential access loop services (Enterprise and Metro Access segment) allowing robust, self-healing transport for premium services including VoIP, video over IP, VPNs, and protected IP managed bandwidth services.

Campus rings for enterprise services for large businesses covering MANs and WANs (Enterprise segment).

DPT products can be currently found on several Cisco router series, Cisco 12000 series Gigabit Switch Routers (GSRs) and Cisco 7500 and 7200 series routers [DPT-5].

PentaCom, recently acquired by Cisco, is another provider of products implementing the SRP technology. Its RingStar8000 is a layer one concentrator conceived to complement the Cisco’s Dynamic Packet Transport family of products [DPT-6].

Upcoming DPT developments include the following:

DPT running at speeds ranging from OC-3c/STM-1c to OC-192c/STM-64c with speed mismatch adaptation.

To provide tools for planning, monitoring, provisioning and configuration, and trouble resolution.

To provide architectural extensions including routing and QoS enhancements.

To extend DPT technology to the edge of the network using DPT access multiplexor products.

APPENDIX 5
Networks deployment

The Section on network deployment is divided in three Subsections. One concerning Metropolitan Area Networks (MAN), another concerning Wide Are Networks (WAN), and the third one concerning Global Area Networks (GAN).

A table template was proposed within the project in order to simplify the task of classification and comparison of actually deployed networks. Partners filled out this table with information they had available on current MAN, WAN and GAN networks deployment. The set of tables collected from the partners of the project is included in the appendix of the Milestone WP1M2 [WP1-M2].

A.5.1 Metropolitan Area Networks

A.5.1.1 Task

TASK is one of the biggest metropolitan area networks in Poland. It is administrated by Academic Computer Centre in Gdansk (CI TASK). The State Committee for Scientific Research (KBN) and Foundation for Polish Science (FNP) are main sponsors of TASK. Connections in the ATM backbone are based on PVCs, SVCs as well as SoftPVCs.

TASK is connected to POL-34 network over POLPAK–T network (TP S.A.) and NASK (NASK - Research and Academic Computer Network). On the base of ATM backbone plain telephony network service was set up in the separated 2 Mbps bandwidth. It was created mainly for internal use of AMG (Medical University of Gdansk) connecting its switchboards, for voice transmission (VoATM) and connectivity with public POTS network. Transmission works on the basis of CBR switched SoftPVC circuits.

A.5.1.2 Anella Científica

The ANELLA CIENTÍFICA (ANELLA) is a Metropolitan Area Network deployed in the area of Barcelona. It has up to 90 Km of fibber rings; SDH based transport and IP/ATM networking. The "Fundació Catalana per a la Recerca" funds the ANELLA. The ANELLA connects 17 entities (universities, hospitals, research centres and school networks), giving 34/155 Mbps access points and 2 Mbps FR/ATM access points. The ANELLA is operated by Catalana de Telecomunicacions S.A. (AL-Pi), and is managed by the Centre de Supercomputació de Catalunya (CESCA).

The ANELLA is connected to the Spanish National Research Network (RedIris) and it is connected to the Neutral Internet Exchange Point in Catalonia, both at 155Mbps.

A.5.2 Wide Area Networks

A.5.2.1 TP S.A.

TP S.A. is one of the largest and dynamic enterprises in Poland. As a national operator TPSA co-operates with international telecom organisations, thus gaining access to the state-of-the-art telecommunication technology and co-creating new solutions.

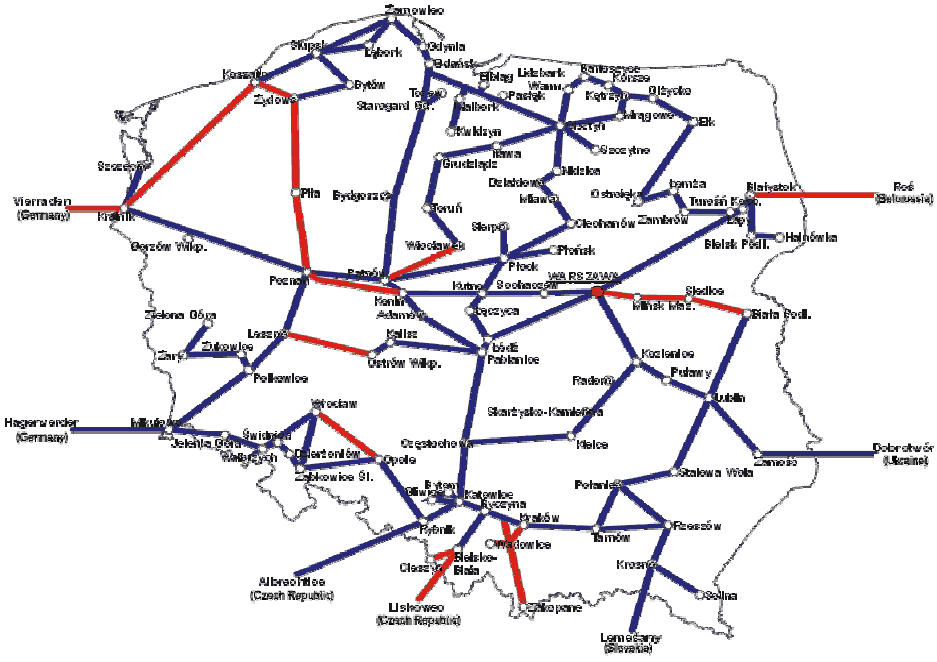
The major source of income of TP S.A. are fixed telephony services in public network-local, long-distance and international for both voice and data transmission.

In April 1996, TP S.A. launched a high-speed data transmission network called POLPAK-T, in addition to low-speed transmission network based on X.25 protocol called POLPAK. The network is based on the Frame Relay protocol and ATM. POLPAK-T provides transmission rates of up to 155 Mb/s. POLPAK-T network allows virtual private networks (VPNs) to be established, and offers permanent virtual circuits (PVCs). POLPAK-T network charges are based on the transmission capacity granted to customers. The customers include banks and companies with a geographically extensive network.

Since 1995, the Company has been developing its fibre-optic cable in the long-distance transmission network utilising SDH technology. By the end of year 2000, the Company plans to install, as a second stage of the development of the long distance network, a further SDH loops, and to introduce 80 colours DWDM systems. SDH technology has also been applied in the development of the local network. From the end of 1995, all inter-metropolitan fibre-optic connections are utilising SDH technology.

A.5.2.2 Tel–Energo

Tel-Energo SA is operator of the Polish nation-wide fibre optical network of the power industry. It manages the network of total length about 8000 kilometres The fibre optical network consists of the backbone network and regional networks constructed in OPGW (Optical Ground Wire) and ADSS (All-Dielectric Self Supporting) technologies. Tel-Energo SA offers the leased digital line service based on fibre optical SDH network. In the near future Tel-Energo SA plans to implement the WDM technology on the most traffic intensive routes, allowing increase in transmission capacity of the backbone network. Tel-Energo SA's plans also include construction of a dedicated SDH ring for handling the international connections. All SDH equipment is monitored and configured from the Management Centre located in Warsaw.



Actual connections - blue colour
Future connections - red colour

Figure 35: Tel-Energo S.A.'s network map

A.5.2.3 RedIris

On 1988, the National Plan for R&D started a horizontal special programme -IRIS- for the Interconnection of Computer Resources (Interconexión de los Recursos InformáticoS) of universities and research centres, and until the end of 1993 this programme was managed by Fundesco. From 1991, when the first promotion stage was finished, IRIS became what RedIRIS is nowadays: the national academic and research network, still funded by National Plan for R&D and at present managed by the Scientific Research Council (Consejo Superior de Investigaciones Científicas). Telefónica Data S.A provides the basic infrastructure.

RedIRIS is the main tool of the National Programme of Applications and Telematic Services and will assume the liability of providing the required network services and actual and future support to the infrastructure, according to the main objectives of the Programme...

About 250 institutions are nowadays connected to RedIRIS, mainly universities and R&D Centres.

A.5.2.4 Telecom Italia

The purpose of this contribution is to describe the architecture of the Telecom Italia transport network.

The reference architecture is based on the deployment of SDH and WDM systems for the transmission of PSTN/ISDN, Leased Lines (CDN) and ATM/IP networks. A simplified model of the network is shown in Figure 39.

The architecture is based on three hierarchical levels: local, regional (metropolitan) and national (backbone) level.

The nodes of the local and regional networks contain different equipment (accordingly to the geographical location):

- Local Exchanges (SL/SGU) for the PSTN/ISDN networks;
- DXC 1/0 for the Leased Lines;
- ATM switches and IP Edge routers for data services;
- ADSL MUX.

SDH rings at different rates (STM-1 to STM16) interconnect local and regional nodes.

The national nodes include the Transit Exchanges (SGT) and the backbone DXC, ATM switches and IP routers.

The interconnection between the regional and national network is based on two nodes (for protection) via DXC 4/1.

The national network interconnects couples of regional nodes with a meshed topology. The transmission network is based on about 50 nodes DXC 4/4 connected via STM-16 line terminals (LT).

Currently, the national network is evolving to a ring architecture based on a number of 4-fibre MS-SPRings interconnected by DXC 4/1. WDM point-to-point systems are also introduced on some links of the national network.

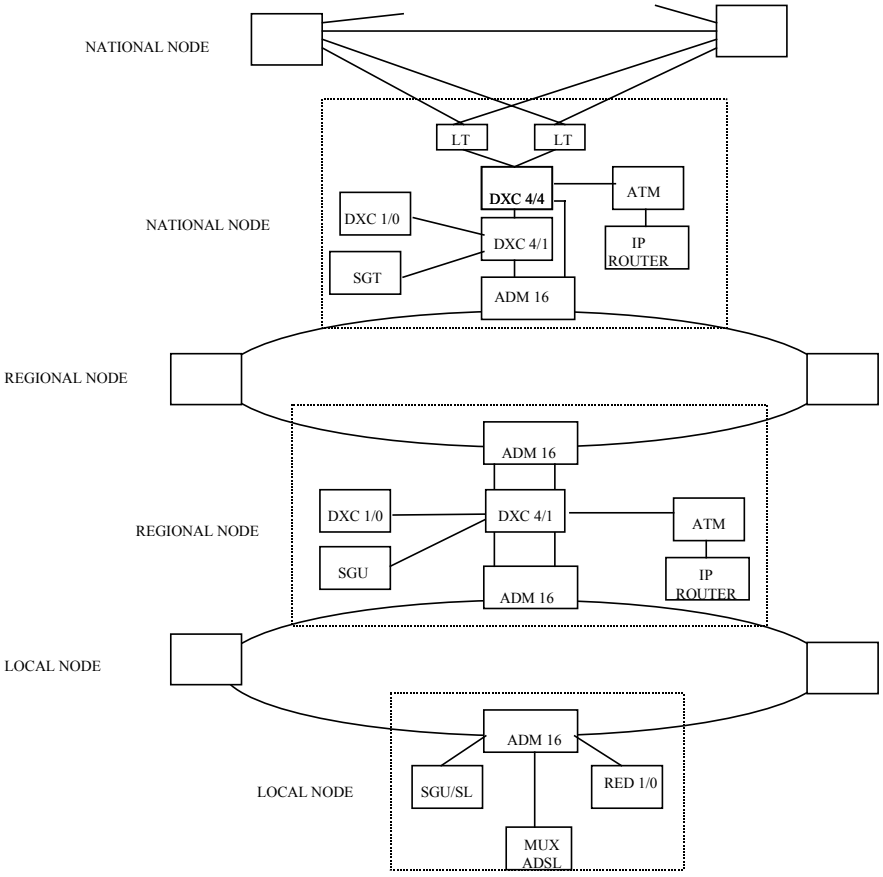


Figure 36: Architecture of the transport network

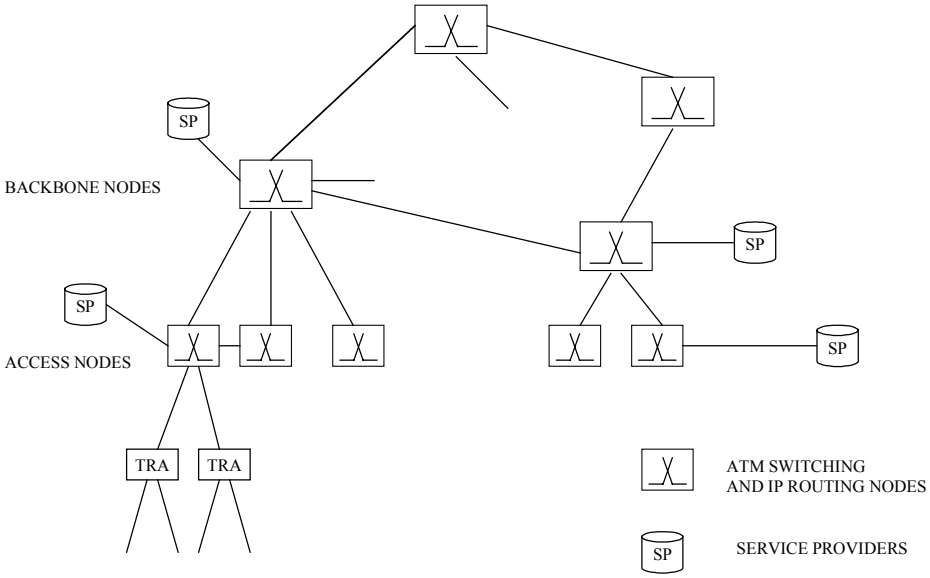


Figure 37: Architecture of the ATM/IP network

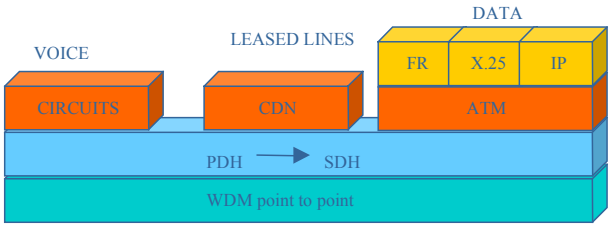


Figure 38: Services and Protocols: current situation

A.5.2.5 Gigabit-Wissenschaftsnetz (G-WiN)

In 1984 the DFN-Verein was founded as a non-profit making body grouping together scientific and research interests in Germany for the promotion of computer-based communication and information services. Now it supplies the scientific community with a high performance information and communication system - the Deutsches Forschungsnetz, DFN. Around 400 institutions from science, research and education make up the DFN-Verein membership.

The current network – the Breitband-Wissenschaftsnetz (B-WiN) - is a virtual private ATM Network and has been operated since spring 1996 with lines up to 155 Mbps. The volume of data traffic on B-WiN as of spring 2000 was more than 200 TBytes per month.

In summer 2000 the B-WiN is scheduled to be replaced by the new Gigabit-Wissenschaftsnetz (G-WiN), which is engineered by Deutsche Telekom Systemlösungen GmbH, a subsidiary of Deutsche Telekom AG. Based on the SDH/WDM infrastructure of Deutsche Telekom AG, G-WiN provides for the following services:

- DFNIP - the global high performance Internet service, managed by the DFN-Verein (IP).
- DFNATM - for flexible broadband usage using Asynchronous Transfer Modus (ATM) with Quality-of-Service.
- DFNCONNECT - the SDH-point-to-point service between various user locations for flexible real-time usage of high quality guaranteed band width ideal for temporary, high service quality computer link-ups for video conferences etc.
- DFNMBONE - the Multicast Service enabling reciprocal communication between a large number of partners.
- DFNWiNShuttle - dial-in access to the Wissenschaftsnetz

A.5.2.6 NTT

The following figures presents NNT offering of ATM services. Summarising table of NTT network deployment is included in the Milestone WP1M2 Appendix. Architecture of the pilot photonic network intallation is depicted in the figure 46.

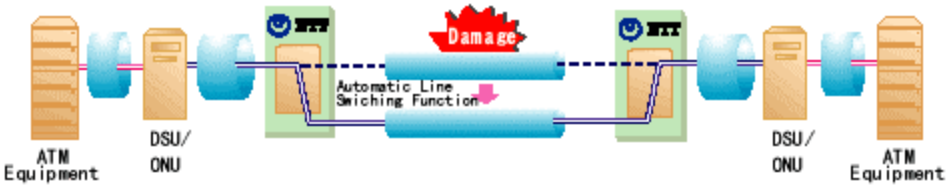


Figure 39: ATM service (dual class)

Figure 40: ATM service (extra class)

Figure 41: ATM service (single class)

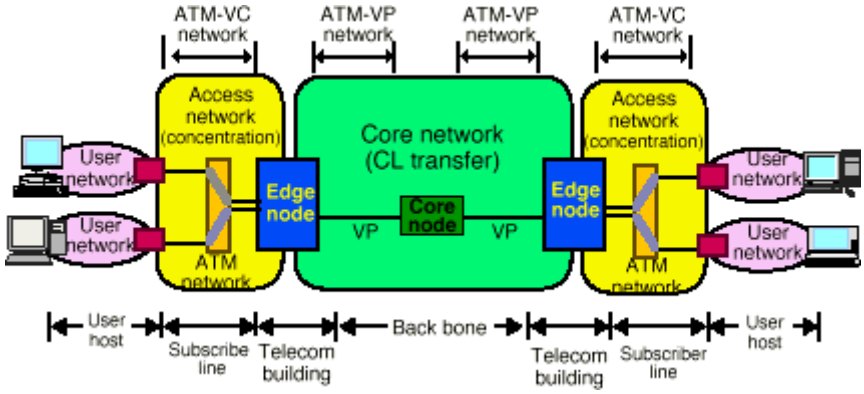


Figure 42: GMN-CL architecture

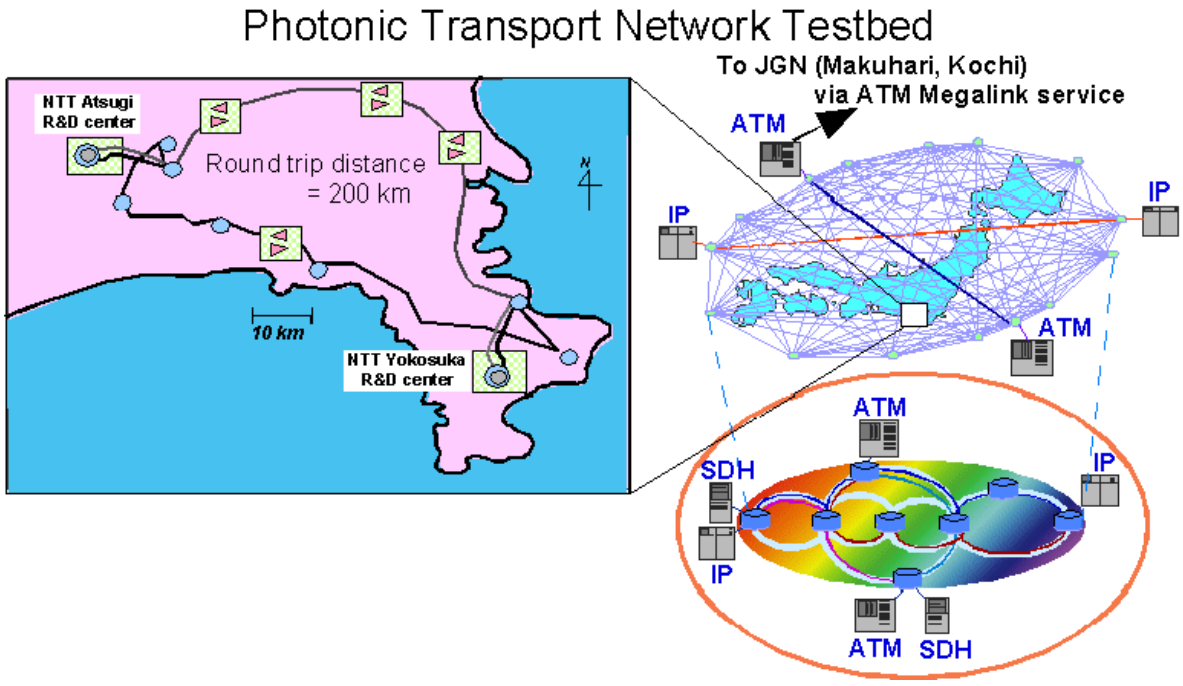


Figure 43: Photonic network testbed

A.5.2.7 GRNET

The GR-NET backbone consists of network nodes in 7 major greek cities, that is, Athens, Thessaloniki, Patras, Ioannina, Xanthi, Heraklion and Larisa. All 7 nodes are co-located in the Greek Public Network Operator's (OTE) central offices under a leasing agreement. GRnet as a network started operating in 1995. It currently provides advanced networking services to more than 50 institutions (all Greek Universities, Research Centers and Technological Institutes) with 120.000 Internet users (approximately 50% of the Greek Internet user community). GRNET's deployment and operation is co-funded by the European Union and the Greek State, since more than 80% of the traffic involves Universities and other Educational Institutions. In addition, it supports advanced pilots of academic and industrial partnerships in the European Union Information Society Technologies (IST) initiative, as an integral part of the Pan-European Research Network TEN-155.

A.5.3 Global area networks

A.5.3.1 Ebone (GTS)

Ebone provides high-quality Internet transit services throughout Europe which are soon to be extended to Eastern Europe and the USA.

Today, it operates a multi-megabit backbone with interconnection points in Amsterdam, Barcelona, Brussels, Bratislava, Copenhagen, Pennsauken, Frankfurt, Geneva, London, Madrid, Milan, Munich, New York, Paris, Prague, Stockholm, Vienna and Zurich.

Ebone's Internet traffic is carried over GTS's dedicated 2.5 Gbps network. This is designed to achieve zero packet loss so the full channel bandwidth is always available for customer traffic. Backbone traffic is exchanged with other major IP networks at appropriate interconnection points.

Ebone's network is specifically designed to meet the needs of:

- High volume Internet Access providers
- Internet Service providers
- Web-hosting providers
- Large multinational corporations

The company is an active partner in the further development of the Internet.

Ebone contributed to the development of the Internet by providing, for the first time, a high-capacity, congestion-free, carrier services to the Internet community. This is based on the use of the GTS trans-European fibre-optic network.

As customers of Ebone's IP service you too can enjoy the complementary benefits of GTS transmission and Ebone Internet services.

A.5.3.2 TEN–155 (DANTE)

DANTE is a non-profit, limited liability company set up in 1993 by European National Research Network organisations, with research association status registered in the UK. Its role is to provide international services to national networks. DANTE is coordinator of Quantum project, which calls for experimentation of new IP and ATM technology using a Wide Area and international test network. TEN-155 is the operational network built as a result of the Quantum project.

The TEN-155 network combines the best of both IP and ATM technology. The network is based on SDH circuits with an ATM overlay, which allows for bandwidth management for optimal loading of the network.

The participating national research networks have the choice of ATM or Packet over SONET access to the TEN-155 network. If the access is done using ATM, different "channels" can be established and can be used for other purposes that the main IP interconnection.

The activities relevant to Quantum Test Project are as follows:

- Multicasting (IP and ATM)
- IP QoS (DiffServ, RSVP, RSVP to ATM SVC mapping)
- IP over ATM
- ATM SVCs
- IPv6
- MPLS
- Route monitoring
- QoS and Flow based monitoring

The European part of the multicast service is currently technically operational and stable. There are still operational issues to be solved regarding the US connectivity, which is not as stable as required.

Managed Bandwidth Service (MBS) allows the definition of Virtual Private Networks linking members of a project and supplying them with network resources defined as bandwidth requirements, lifetime of the established connections, traffic profile and a complete set of network parameters.

The proposal, known as GÉANT, advocates an evolutionary approach to the development of TEN-155 based on the existing structure to create a shared multi gigabit core network available to all of the national research networks across Western, Central, and Eastern Europe. This will be complemented by a continuation of the Managed Bandwidth Service and the technology and service test programme, which are currently part of TEN-155.

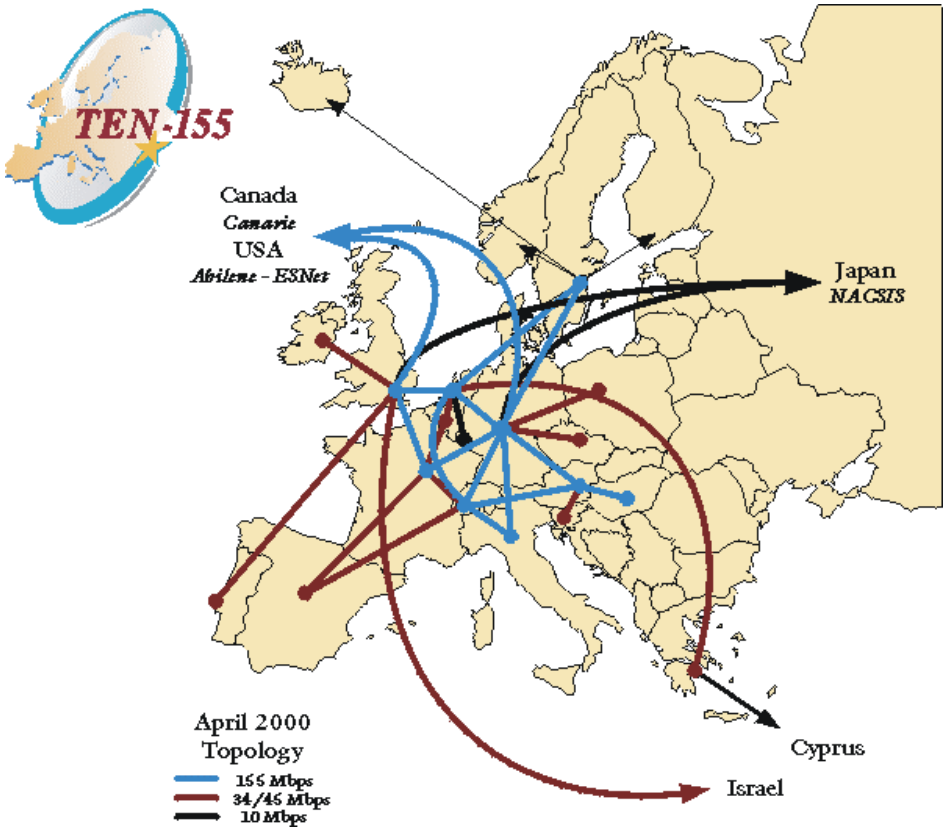


Figure 44: TEN-155 coverage

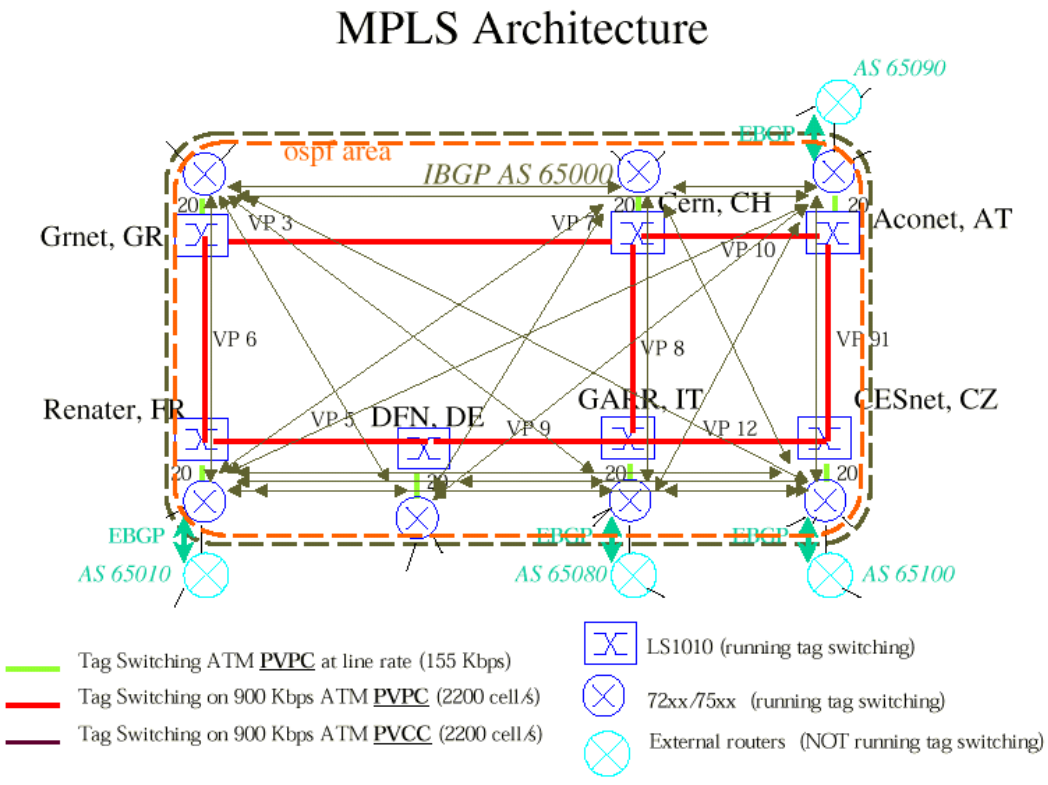


Figure 45: MPLS testbed architecture

A.5.4 Summary

The section on deployed transport networks presents 11 networks: 2 Metropolitan Area Networks, 7 Wide Area Networks and 2 Global Area Networks. All networks with exception of EBONE are situated in Europe and nearly all of them were deployed in the 1990's. Only EBONE was deployed in 2000. MANs are represented by two example networks: TASK and ANELLA. TASK is based on the fibre optic infrastructure of the total length of 140 km. There are about 40 main network nodes. ANELLA MAN uses 90 km of fibre with 21 nodes. In both Metropolitan Area Networks WDM is still not deployed. SDH/SONET technology is used only in ANELLA network with APS (1+1) protection and STM1 interfaces in double ring topology. Both of the networks offers ATM services based on PVC, SVC (also Soft-PVC in the TASK) connections with bit rates starting from 34 to 622 Mbps. IP is the common layer of both described MAN networks.

Contributions from following WAN network operators have been received and included in the document: TPS.A., Tel-Energo S.A., RedIris, G-WiN, Telecom Italia. TPS.A. is a national operator in Poland, Tel-Energo S.A. is the operator of the Polish, nation-wide fibre optic network of the power industry. RedIris (Spain) is the national academic and research network. The Deutsches Forschungsnetz is the computer-based communications infrastructure for research, science, education and culture in Germany. This infrastructure consists the broadband research network: B-WiN, in summer 2000 the B-WiN is scheduled to be replaced by the new Gigabit-Wissenschaftsnetz (G-WiN). CSELT and NTT contributions on its networks are also included.

Described WAN networks are based on G.652 and G.653 type of fibres. The total fibre length vary from 400km to 8000km. In all networks with the exception of RedIris, WDM is used or planned, TP S.A. plans introduction of the WDM systems with 1+1 protection. Operators of two networks already decided to introduce 80 colour system (TPS.A., G-WiN).

At the SDH layer all presented WANs uses interfaces of bit rate varying from STM1 to STM16, deployed protection mechanisms include 1+1, 1:1 and 1:n. G-WiN operator is the only one who now planned introduction of the STM64c in 2002. Reported SDH topology varies from ring/mesh (TPS.A.), star (RedIris) or ring (Telecom Italia).

The use the ATM technology is reported only by TPS.A., RedIris, Telecom Italia and NTT. It is reported that the most of the equipment is provided by Nortel and Cisco. The typical data rate of the ATM interfaces is 155 Mbps. G-WiN operator plans deployment of ATM but final decision has not been pronounced. In the networks of TP S.A. and Telecom Italia services based on Frame Relay are available.

IP layer services are offered up till now by RedIris, G-WiN and Telecom Italia. G-WiN's 27 core IP nodes work with Packet over SDH interfaces. IP service charging in G-WiN is volume based. Status of all presented networks is operational. TPS.A., Tel-Energo S.A. and Telecom Italia plan introduce or expand WDM in their networks.

Two of the included in the document networks are characterised as a global ones: EBONE – European Internet backbone is placed in 12 countries with direct connection to New York. Ten-155 (Dante) is a non-profit, limited liability company set up by European National Research Network organisations. Its role is to provide international services to national networks. DANTE is coordinator of Quantum project, which calls for experimentation of new IP and ATM technology using a wide area and international test network. TEN-155 is the operational network built as a result of the Quantum project. Network interconnects 23 European countries in Western, Central and Eastern Europe. EBONE uses 16,964 km of fibre with POPs in 28 cities. In both of the networks SDH technology is deployed.

IP services are offered in the both networks. IP4/IP6 are supported by TEN-155. In TEN-155 network MPLS is planned and being tested. The network supports following services in the backbone: Deterministic Bit Ratio (DBR) with QoS-1, traffic parameter: Peak Cell Rate (PCR)

Statistical Bit Rate (SBR2 or SBR3) with QoS-3, traffic parameters: Peak Cell Rate (PCR), Sustained Cell Rate (SCR) and Available Bit Rate (ABR) with QoS-3, traffic parameters: Peak Cell Rate (PCR) and Minimum Cell Rate (MCR).

There is available SLA prepared for EBONE, key parameters reported: delay 100ms (average) [Hessing, 1999], Packet Loss Ratio: 0% [Hessing, 1999], [URL GTS] Availability: 99.5% [Hessing, 1999], 99.9% [URL GTS]. Status of EBONE is operational. TEN-155’s status is operational/pilot and the main purpose indicated is service and research. All details of the report on status of deployed transport networks are included in the Milestone document WP1M2 [WP1-M2].