Kungl Tekniska Högskolan
Institutionen för teleinformatik
CCSlab

2G1305 Internetworking/Internetteknik
Winter 2002, Period 3

Module 1: Introduction

Lecture notes of G. Q. Maguire Jr.

For use in conjunction with *TCP/IP Illustrated*, volumes 1 and 3, by W. Richard Stevens and *TCP/IP: Principles, Protocols, and Architectures*, *Vol. 1*, by Douglas E. Comer, Prentice Hall, 4th ed. 2000.

Last modified: 2002.01.13:17:36

Welcome to the Internetworking course!

The course should be fun.

We will dig deeper into the TCP/IP protocols and protocols built upon them.

Information about the course is available from the course web page:

http://www.it.kth.se/edu/gru/Internet

Staff Associated with the Course

Instructor (Kursansvarig)

prof. Gerald Q. Maguire Jr. <maguire@it.kth.se>

Assistants for Recitation Sessions (Övningar)

Dr. Johan Montelius <jm@it.kth.se>

Associate Professor Dr. Robert Rönngren <robertr@it.kth.se>

Administrative Assistant: recording of grades, registration, etc.

Rita Johnsson <ritaj@it.kth.se>

Goals, Scope and Method

Goals of the Course

- To give deep knowledge and competence (*designing*, *analyzing,* and *developing*) of Internet protocols and architecture, both practical and analytical.
- To be able to read and understand the Internet standardization documents (IETF RFCs and Internet Drafts) and current Internet literature.
- You should have the knowledge and competence to do exciting Internet related research and development.

Scope and Method

- Dig deeper into the TCP/IP protocol suite by using diagnostic tools to examine, observe, and analyze these protocols in action. Understanding the details!
- Demonstrate this by writing a written report (and passing the exam).

Aim

After this course you should be able to read the current internet literature at the level of IEEE Communications Magazine, IEEE Network, IEEE Transactions on Communications, IEEE Transactions on Communications, IEEE Journal on Selected Areas in Communications, IEEE/ACM Transactions on Networking, IEEE Communications Surveys (On-line Journal), … . See the IEEE Communication Society's list of publications.

While you may not be able to understand all of the articles in the above journals and magazines, you should be able to read 90% or more of the articles and have good comprehension. *You should develop a habit of reading the journals, trade papers, etc.*

You should be able to write internetworking articles at the level of Miller Freeman's *Network Magazine* or IEEE Internet Computing. In subsequent courses you will also develop you ability to orally present your ideas.

Prerequisites

- Telesys, gk **or**
- Datorkommunikation och datornät/Data and computer communication **or**
- Equivalent knowledge in Computer Communications (this requires permission of the instructor)

Contents

This course will focus on the **protocols** that are the fundaments of the Internet. We will explore what internetworking means and what it requires. We will give both practical and more general knowledge concerning the Internet network architecture.

The course consists of 14 hours of lectures and 14 hours of recitations (övningar).

Topics

- What an internet is and what is required of protocols to allow internetworking
- details of routing and routing protocols (RIP, BGP, OSPF, …)
- multicasting
- Domain Name System (DNS, Dynamic DNS)
- what happens from the time a machine boots until the applications are running (RARP, BOOTP, DHCP, TFTP)
- details of the TCP protocols and some performance issues
- details of a number of application protocols (especially with respect to distributed file systems)
- network security (including firewalls, AAA, IPSec, SOCKs, …)
- differences between IPv6 and IPv4
- network management (SNMP) and
- We will also examine some emerging topics:
  - cut-through routing, tag switching, flow switching, QoS, Mobile IP, Voice over IP, SIP, NAT, VPN, Diffserv, … .

Examination requirements

- Written examination (3 p)
  - based on literature, lectures, and recitations
- Written assignments (1 p)
  - based on lectures, recitations, and your references

Grades: U, 3, 4, 5

Written Assignment

Goal: to gain analytical or practical experience and to show that you have mastered some Internetworking knowledge (in addition to what you show on the written examination).

- Can be done in a group of **1 to 3** students (formed by yourself). Each student must contribute to the final report.
- There will be one or more suggested topics, additional topics are possible (discuss this with one of the teachers **before** starting).

Assignment Registration and Report

- Registration: 1 March 2002, to <maguire@it.kth.se>
  - Group members, leader.
  - Topic selected.
- For Analytical Assignments
  - The length of the final report should be 7-8 pages (roughly 3,000 words) for each student.
  - Contribution by each member of the group - must be clear
- For Practical Assignments
  - A short technical document describing: 1) what you have done; 2) who did what; 3) methods and tools used; 3) the test or implementation results.

Final Report: **March 11, 2002**

- Send email with URL link to <maguire@it.kth.se>
- Late assignments will <u>not</u> be accepted

Note that it is pemissible to start working *well in advance* of the deadlines!

Literature

The course will mainly be based on the book: *TCP/IP Illustrated*, Volume 1: The Protocols by W. Richard Stevens, Addison-Wesley, 1994, ISBN 0-201-63346-9 and *Internetworking with TCP/IP: Principles, Protocols, and Architectures, Vol. 1*, by Douglas E. Comer, Prentice Hall, 4th edt. 2000, ISBN 0-13-018380-6.

We will also frequently refer to the commented source code in *TCP/IP Illustrated*, Volume 2: The Implementation by Gary R. Wright and W. Richard Stevens, Addison-Wesley, 1995, ISBN 0-201-63354-X and *IPv6: The New Internet Protocol*, by Christian Huitema, Prentice-Hall, 1996, ISBN 0-13-241936-X. Copies of these two reference books is on reserve in the library (both on main campus and at Electrum).

You may also find the brand new book *TCP/IP and Linux Protocol Implementation: Systems Code for the Linux Internet* by John Crowcroft and Iain Phillips), Wiley, November 2001, ISBN: 0-471-40882-4 to be a very useful book; in the future it will probably displacing Vol. 2 of Wright and W. Richard Stevens.

When discussing the performance of HTTP we will refer to *TCP/IP Illustrated*, *Volume 3: TCP for Transactions, HTTP, NNTP, and the UNIX Domain Protocols*, Addison-Wesley, 1996, ISBN 0-201-63495-3.

With regard to Mobile IP the following two books are useful sources:

- *Mobile IP: Design Principles and Practices* by Charles E. Perkins, Addison-Wesley, 1998, ISDN 0-201-63469-4.
- *Mobile IP: the Internet Unplugged* by James D. Solomon, Prentice Hall, 1998, ISBN 0-13-856246-6.

Internetworking Technologies Handbook by Kevin Downes (Editor), H. Kim Lew, Steve Spanier, Tim Stevenson (Online: *http://www-fr.cisco.com/univercd/cc/td/doc/cisintwk/ito_doc/index.htm*)

We will refer to other books, articles, and RFCs as necessary. In addition, there will be **compulsory** written exercises.

Lecture Plan

- Lecture 1: Introduction
  - Course arrangement
  - Internet Basics, IP addressing and subnetting.
- Lecture 2: IP Routing
  - Routing principle and protocols (RIP, OSPF, BGP, CIDR)
  - ICMP, ARP/RARP, Ping, Traceroute
- Lecture 3: User Datagram Protocol
  - UDP
  - Multicasting, IGMP, DNS, TFTP, DHCP/BOOTP
- Lecture 4: TCP
  - TCP connections, Interactive and bulk data flows
  - TCP timers, retransmission and performance
- Lecture 5: Network Management and Applications
  - SNMP, FTP, SMTP, NFS, HTTP, VoIP
- Lecture 6: IP Mobility and IP Security
  - Mobile IP, IPsec., AAA.
- Lecture 7: Future Issues

Context of the course

“The network called the Internet is the single most important development in the communications industry since the public switched voice network was constructed…”

-- John Sidgmore
CEO, UUNET Technologies
and COO, WorldCom[1]

1. http://www.lucent.com/enterprise/sig/exchange/present/slide2.html

Architecture

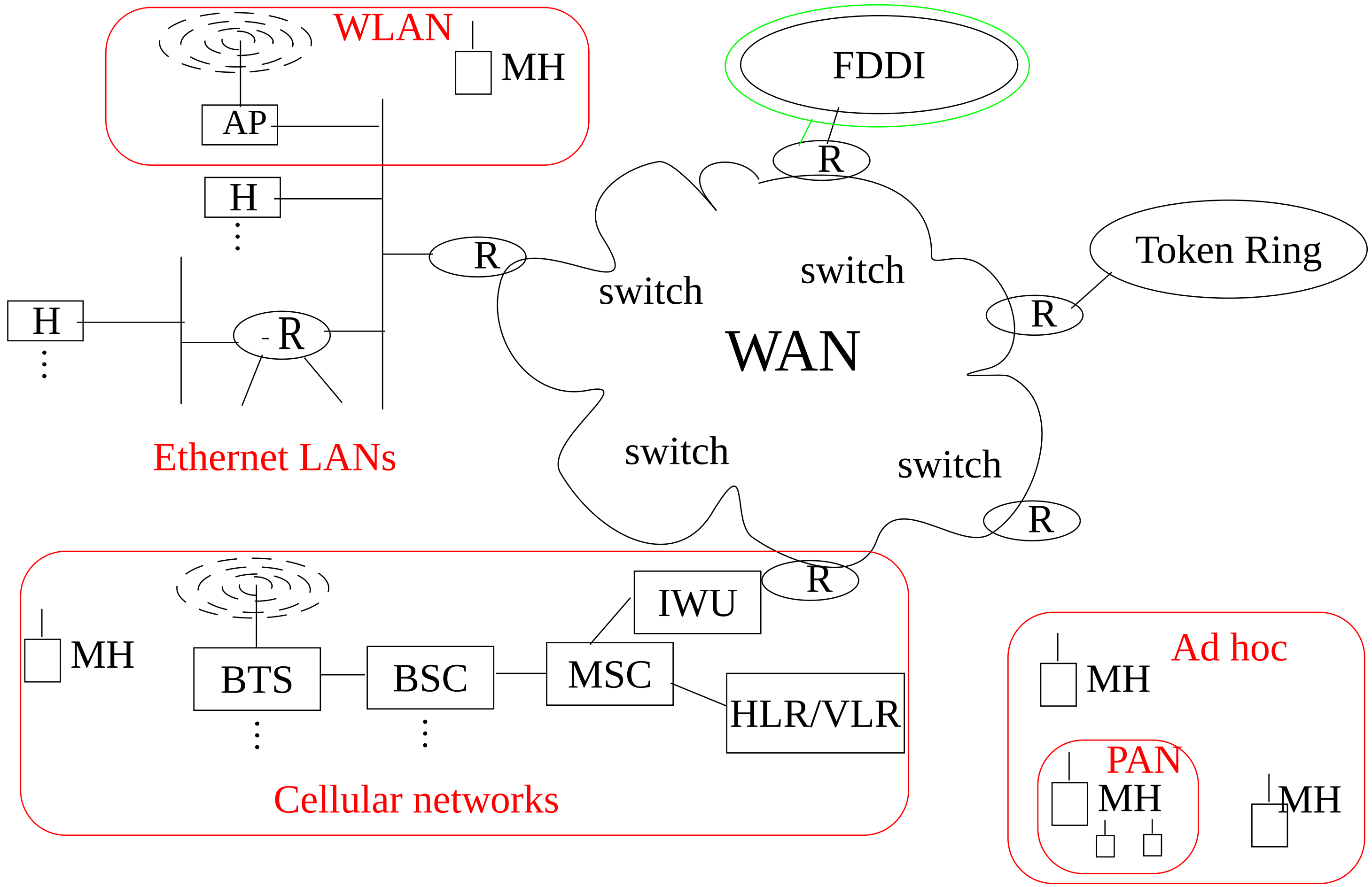

Figure 1: Multiple network technologies - internetworked together

Maguire
maguire@it.kth.se

Power of the Internet (chaos)

'"Historically, the Internet has been an environment in which to experiment. There have been a few basic rules. The most important is the standard for IP and TCP.

The are other important standards for promulgating routing information and the like, but the real power of the Internet idea is that there are not mandidated stanards for what can run over the 'Net.

Anyone who adheres to TCP/IP standards can create applications and run them without getting anyone's permission. No ISP even has to know you are experimenting (or playing, which is also OK). This freedom produces unpredictable results. New industries can be created almost overnight and existing industries severely affected. …"

-- Scott O. Bradner, "The Importance of Being a Dynamist", Network World, December 13, 1999, p. 48 (www.nwfusion.com)

Internet Trends

- Number of users and internet devices increases very rapidly
  - with projections of 25%/year according to the Gartner Group
  - Network Wizards' Internet Domain Survey - 125,888,197 as of July 2001 *http://www.isc.org/ds/WWW-200107/index.html*
  - RIPE's survey *European hosts*: 33,866,411 of which 816,961 are in Sweden - Nov. 2001 {Note that the above estimates are based on DNS information}
  - Matrix Information and Directory Services' *world map of the Internet*
  - NUA Internet Surveys estimates by region

http://www.nua.ie/surveys/how_many_online/index.html

	Sept. 1999	August 2001
World Total	201. million	513.41 million
Africa	1.72 million	4.15 million
Asia/Pacific	33.61 million	143.99 million
Europe	47.15 million	154.63 million
Middle East	0.88 million	4.65 million
Canada & USA	112.4 million	180.68 million
Latin America	5.29 million	25.33 million

- QoS: Demand for integrating many different types of traffic, such as video, audio, and data traffic, into one network ⇒ **Multicast**, **IPv6**, **RSVP**, **DiffServ**, emphasis on **high performance**, and TCP **extensions**
  - we will examine a number of these protocols in this course
- Mobility: both users and devices are mobile
  - There is a difference between portable (bärbar) vs. mobile (mobil).
  - IP is used in wireless systems (for example 3G cellular).
  - Increasing use of wireless in the last hop (WLAN, PAN, Wireless MAN, …)
- Security:
  - Wireless mobile Internet - initial concern driven by wireless link
  - Fixed Internet - distributed denial of service attacks, increasing telecommuting, …

IP traffic growing exponentially!

Traffic increasing (but not due to voice)

- IP traffic between US and Sweden many times the total voice+FAX traffic
- 60Gbit/s transatlantic fiber

Fixed Links - arbitrarily fast:

- LANs: 10Mbits/s, 100Mbits/s, 1Gbits/s, 10Gbits/s, …
- Backbones: 45 Mbits/s or 34 Mbits/s ⇒ 155 Mbit/s, 662 Mbit/s, and Gigabits/s
  Transoceanic fibers between continents ⇒ Gbit/s
- Major sites link to backbones: T1 (1.5 Mbits/sec) or E1 rate links (2 Mbits/sec)
  ⇒ increasingly 34 Mbits/s or 45 Mbits/s - much greater in some metropolitan areas
- Individual users links: 28.8 Kbits/s and ISDN (128Kbits/s)
  ⇒ ethernet and xDSL (Mbits/s .. ~51 Mbits/s in the fast direction)

Points of Presence (PoPs) + FIX/CIX/GIX/MAE[1] ⇒ GigaPoPs

(George) Guilder's Law states that network speeds will triple every year for the next 25 years. This dwarfs Moore's law that predicts CPU processor speed will double every 18 months.

1. Federal Internet eXchange (FIX), Commercial Internet eXchange (CIX), Global Internet eXchange (GIX), Metropolitan Access Exchange (MAE)

Speed

"... The Internet world moves fast. The integration of voice and data onto a single network is not being lead by the International Telecommunications Union or by Bellcore. Rather, its being lead by entrepreneurs like …. Until now, the voice networks dominated. Data could ride on top of the phone network -- when it was convenient. The explosion of data networking and Internet telephony technology is making the opposite true. Now voice can ride on data networks -- when it is convenient."[1]

Because of bandwidth constraints, Internet telephony would not be a major factor "**for a long time -- maybe nine to twelve months.**"

-- president of a major ISP[2]

Internet time - 7x real time

-- Ira Goldstein, HP

1. from *http://www.dialogic.com/solution/internet/apps.htm*
2. from *http://www.dialogic.com/solution/internet/apps.htm*

Growth rates

> Some people think the Internet bandwidth explosion is relatively recent, but right from the beginning it's been a race against an ever-expanding load. It isn't something you can plan for. In fact, the notion of long-range planning like the telcos do is almost comical. Just last month, a local carrier asked us why we didn't do five-year plans, and we said, "We do-about once a month!"

-- Mike O'Dell[1] VP and Chief Technologist UUNET

Mike points out that the growth rate of the Internet is driven by the increasing speed of computers, while telcos have traffic which was proportional to the growth in numbers of people (each of whom could only use a very small amount of bandwidth).

- by 1997 UUNET was adding at least one T3/day to their backbone

1. from *http://www.data.com/25years/mike_odell.html*

¿Question?

"Which would you rather have twice as fast:
your computer's processor or modem?

After 30 years of semiconductor doublings under Moore's Law, processor speed are measured in megahertz. On the other hand, after 60 years of telco's snoozing under monopoly law, modem speeds are measure in kilobits. Modems are way too slow for Internet access, but you knew that."[1]

-- Bob Metcalfe, inventor of Ethernet in 1973

1. "From the Ether: Moving intelligence and Java Packets into the Net will conserve bandwidth", by Bob Metcalfe, Inforworld, Oct., 6, 1997, pg. 171.

Increasing Data Rates

“Ethernet’

- 3 Mbps Ethernet (actually 2.944 Mbits/sec)
- 10 Mbps Ethernet (which became 802.3)
- 100 Mbps Ethernet (100Tx)
- Gigabit Ethernet (802.3z, 802.3ab)
- 10 Gbps Ethernet (IEEE 802.3ae)

Optical

- Dense Wavelength Division Multiplexing (DWDM) - allowing 1000s of multi-Gbits/s channels to be carried on existing fibers

Wireless

- 802.11 Wireless LAN
- 802.15 Wireless Personal Area Network (WPAN)
- 802.16 Metropolitan Area Networks - Fixed Broadband Wireless (10 .. 66 GHz)

Internetworking

Internetworking is

- based on the interconnection (concatenation) of multiple networks
- accommodates multiple underlying hardware technologies by providing a way to interconnect **heterogeneous** networks and makes them inter-operate.

We will concern ourselves with one of the most common internetworking protocols IP (there *are* other internetworking protocols, such as Novell's Internetwork Packet Exchange (IPX), Xerox Network Systems (XNS), IBM's Systems Network Architecture (SNA), OSI's ISO-IP).

We will examine both IP:

- version 4 - which is in wide use
- version 6 - which is coming into use

Internet: the worldwide internet

Review of Layering

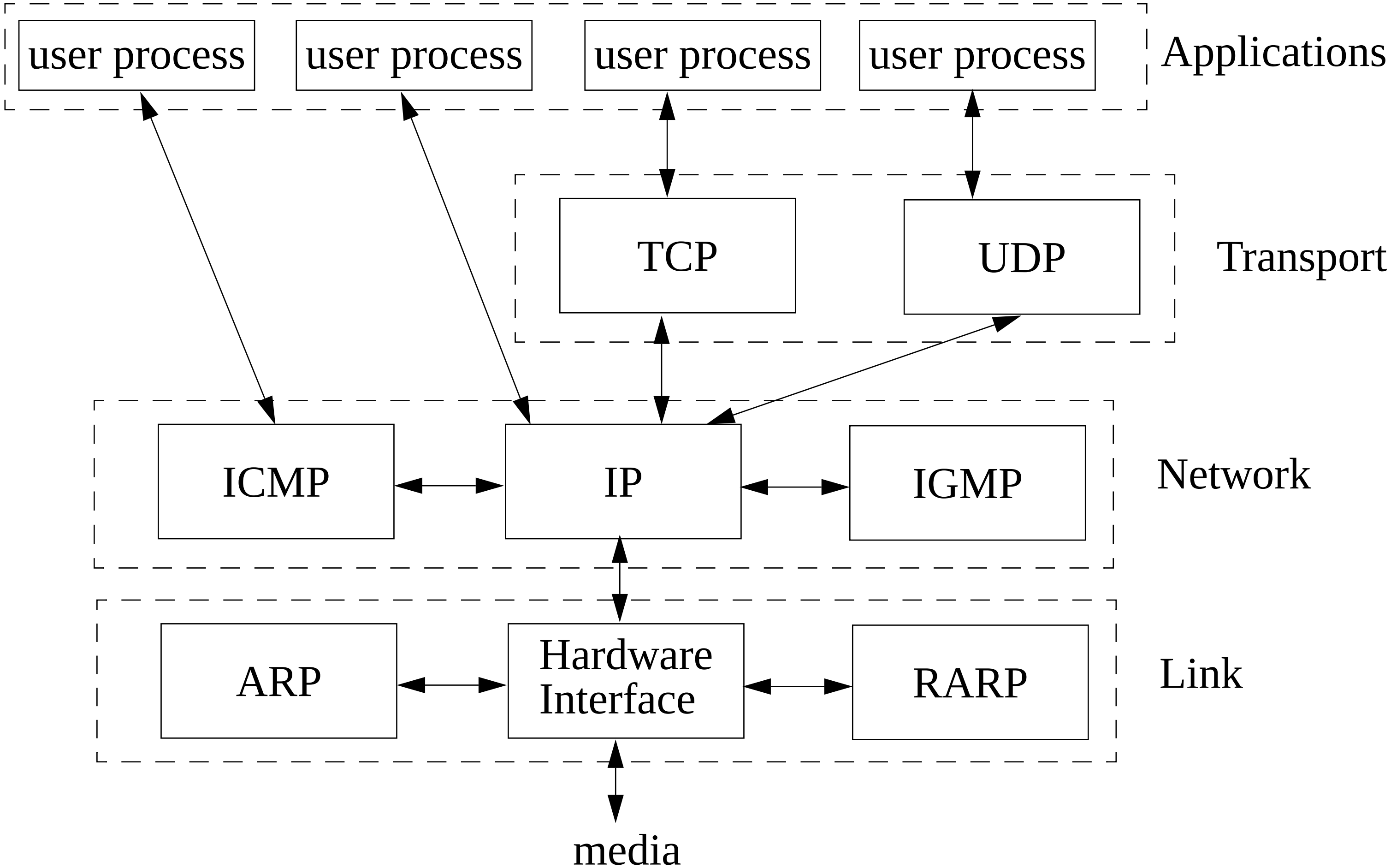

Figure 2: protocol layers in the TCP/IP protocol suite
(see Stevens, Volume 1, figure 1.4, pg. 6)

Encapsulation

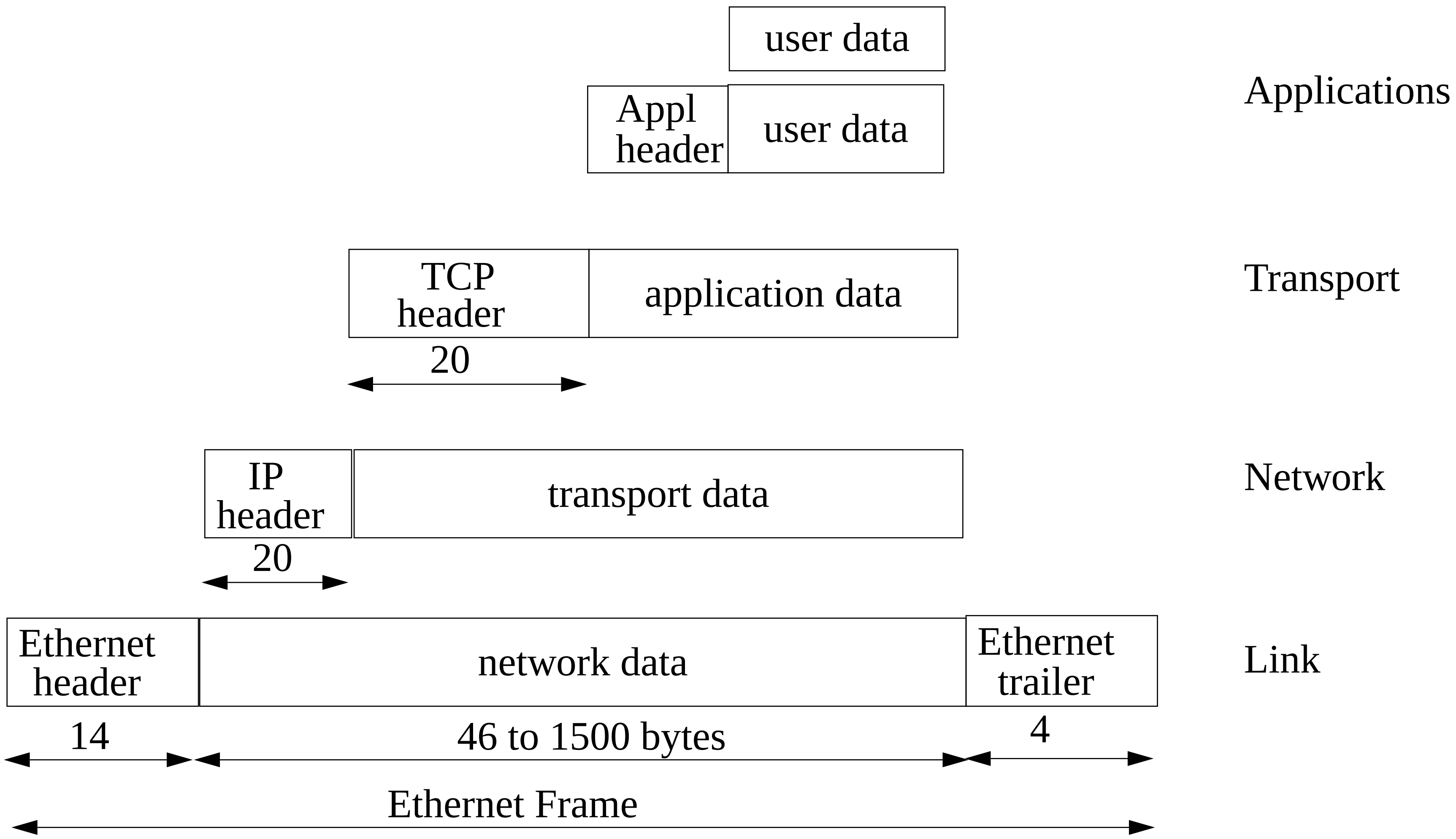

Figure 3: Encapsulation of data
(see Stevens, Volume 1, figure 1.7, pg. 10)

Demultiplexing

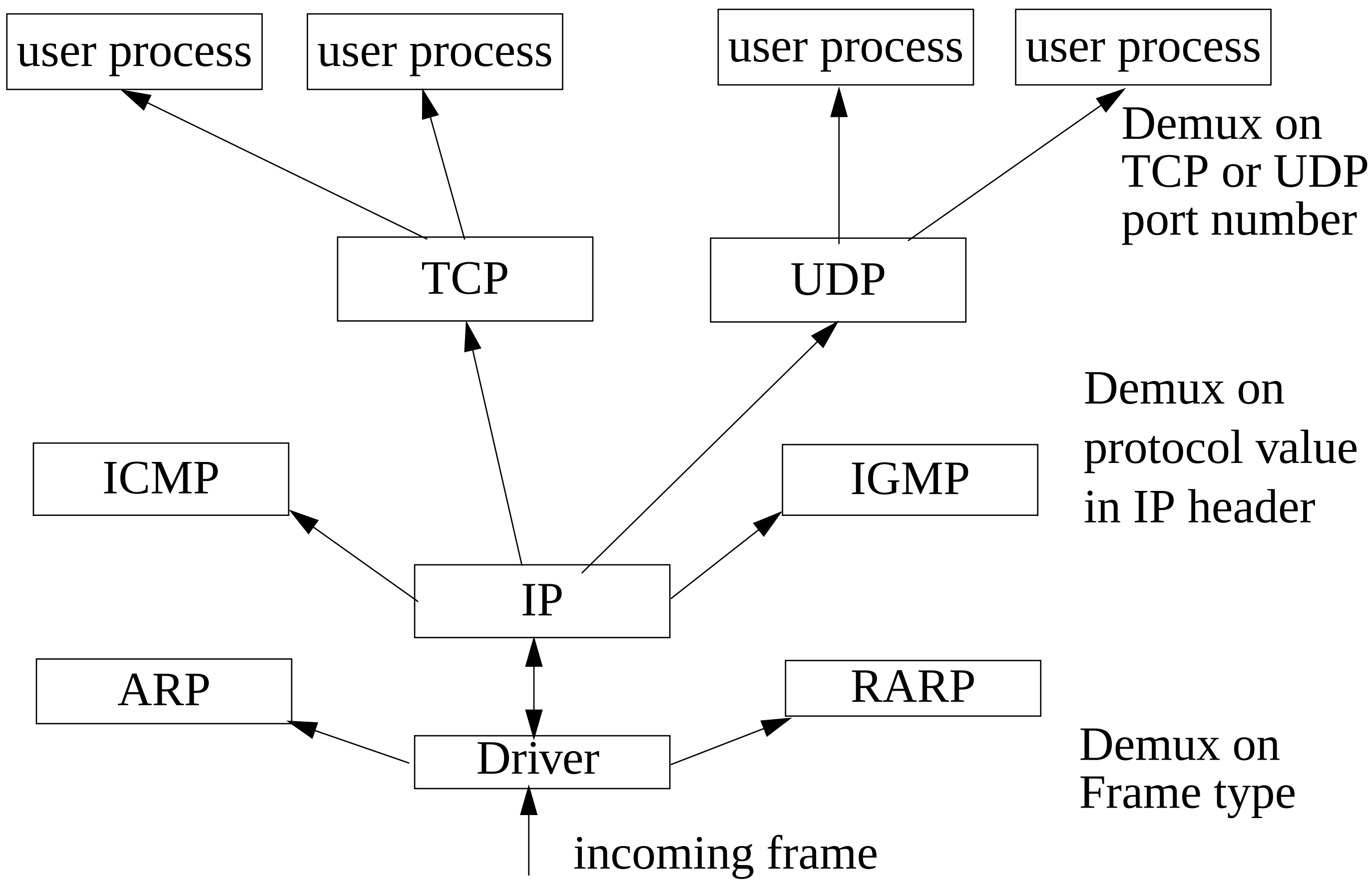

Figure 4: Demultiplexing
(see Stevens, Volume 1, figure 1.8, pg. 11)

IP "Protocol" field (RFC 1700)

In the Internet Protocol (IP) [DDN], [RFC791] there is a field, called Protocol, to identify the next level protocol. This is an 8 bit field.

Assigned Internet Protocol Numbers (assigned by *Internet Assigned Numbers Authority* (IANA) *http://www.iana.org/assignments/protocol-numbers*

Decimal	Keyword	Protocol	References
0	HOPOPT	IPv6 Hop-by-Hop Option	[RFC1883]
1	ICMP	Internet Control Message	[RFC792]
2	IGMP	Internet Group Management	[RFC1112]
3	GGP	Gateway-to-Gateway	[RFC823]
4	IP	IP in IP (encapsulation)	[RFC2003]
5	ST	Stream	[RFC1190,RFC1819]
6	TCP	Transmission Control	[RFC793]
7	CBT	CBT	[Ballardie]
8	EGP	Exterior Gateway Protocol	[RFC888,DLM1]
9	IGP	any private interior (used by Cisco for their IGRP)	[IANA]
10	BBN-RCC-MON	BBN RCC Monitoring	[SGC]
11	NVP-II	Network Voice Protocol	[RFC741,SC3]
12	PUP	PUP	[PUP,XEROX]

Decimal	Keyword	Protocol	References
13	ARGUS	ARGUS	[RWS4]
14	EMCON	EMCON	[BN7]
15	XNET	Cross Net Debugger	[IEN158,JFH2]
16	CHAOS	Chaos	[NC3]
17	UDP	User Datagram	[RFC768,JBP]
18	MUX	Multiplexing	[IEN90,JBP]
19	DCN-MEAS	DCN Measurement Subsystems	[DLM1]
20	HMP	Host Monitoring	[RFC869,RH6]
21	PRM	Packet Radio Measurement	[ZSU]
22	XNS-IDP	XEROX NS IDP	[ETHERNET,XEROX]
23	TRUNK-1	Trunk-1	[BWB6]
24	TRUNK-2	Trunk-2	[BWB6]
25	LEAF-1	Leaf-1	[BWB6]
26	LEAF-2	Leaf-2	[BWB6]
27	RDP	Reliable Data Protocol	[RFC908,RH6]
28	IRTP	Internet Reliable Transaction	[RFC938,TXM]
29	ISO-TP4	ISO Transport Protocol Class 4	[RFC905,RC77]
30	NETBLT	Bulk Data Transfer Protocol	[RFC969,DDC1]
31	MFE-NSP	MFE Network Services Protocol	[MFENET,BCH2]
32	MERIT-INP	MERIT Internodal Protocol	[HWB]
33	SEP	Sequential Exchange Protocol	[JC120]
34	3PC	Third Party Connect Protocol	[SAF3]
35	IDPR	Inter-Domain Policy Routing Protocol	[MXS1]

Decimal	Keyword	Protocol	References
36	XTP	XTP	[GXC]
37	DDP	Datagram Delivery Protocol	[WXC]
38	IDPR-CMTP	IDPR Control Message Transport Proto	[MXS1]
39	TP++	TP++ Transport Protocol	[DXF]
40	IL	IL Transport Protocol	[Presotto]
41	IPv6	Ipv6	[Deering]
42	SDRP	Source Demand Routing Protocol	[DXE1]
43	IPv6-Route	Routing Header for IPv6	[Deering]
44	IPv6-Frag	Fragment Header for IPv6	[Deering]
45	IDRP	Inter-Domain Routing Protocol	[Sue Hares]
46	RSVP	Reservation Protocol	[Bob Braden]
47	GRE	General Routing Encapsulation	[Tony Li]
48	MHRP	Mobile Host Routing Protoco	[David Johnson]
49	BNA	BNA	[Gary Salamon]
50	ESP	Encap Security Payload for IPv6	[RFC1827]
51	AH	Authentication Header for IPv6	[RFC1826]
52	I-NLSP	Integrated Net Layer Security TUBA	[GLENN]
53	SWIPE	IP with Encryption	[JI6]
54	NARP	NBMA Address Resolution Protocol	[RFC1735]
55	MOBILE	IP Mobility	[Perkins]
56	TLSP	Transport Layer SecurityProtocol (using Kryptonet key management)	[Oberg]
57	SKIP	SKIP	[Markson]

Decimal	Keyword	Protocol	References
58	IPv6-ICMP	ICMP for IPv6	[RFC1883]
59	IPv6-NoNxt	No Next Header for IPv6	[RFC1883]
60	IPv6-Opts	Destination Options for IPv6	[RFC1883]
61		any host internal protocol	[IANA]
62	CFTP	CFTP	[CFTP,HCF2]
63		any local network	[IANA]
64	SAT-EXPAK	SATNET and Backroom EXPAK	[SHB]
65	KRYPTOLAN	Kryptolan	[PXL1]
66	RVD	MIT Remote Virtual Disk Protocol	[MBG]
67	IPPC	Internet Pluribus Packet Core	[SHB]
68		any distributed file system	[IANA]
69	SAT-MON	SATNET Monitoring	[SHB]
70	VISA	VISA Protocol	[GXT1]
71	IPCV	Internet Packet Core Utility	[SHB]
72	CPNX	Computer Protocol Network Executive	[DXM2]
73	CPHB	Computer Protocol Heart Beat	[DXM2]
74	WSN	Wang Span Network	[VXD]
75	PVP	Packet Video Protocol	[SC3]
76	BR-SAT-MON	Backroom SATNET Monitoring	[SHB]
77	SUN-ND	SUN ND PROTOCOL-Temporary	[WM3]
78	WB-MON	WIDEBAND Monitoring	[SHB]
79	WB-EXPAK	WIDEBAND EXPAK	[SHB]
80	ISO-IP	ISO Internet Protocol	[MTR]

Decimal	Keyword	Protocol	References
81	VMTP	VMTP	[DRC3]
82	SECURE-VMTP	SECURE-VMTP	[DRC3]
83	VINES	VINES	[BXH]
84	TTP	TTP	[JXS]
85	NSFNET-IGP	NSFNET-IGP	[HWB]
86	DGP	Dissimilar Gateway Protocol	[DGP,ML109]
87	TCF	TCF	[GAL5]
88	EIGRP	EIGRP	[CISCO,GXS]
89	OSPFIGP	OSPFIGP	[RFC1583,JTM4]
90	Sprite-RPC	Sprite RPC Protocol	[SPRITE,BXW]
91	LARP	Locus Address Resolution Protocol	[BXH]
92	MTP	Multicast Transport Protocol	[SXA]
93	AX.25	AX.25 Frames	[BK29]
94	IPIP	IP-within-IP Encapsulation Protocol	[JI6]
95	MICP	Mobile Internetworking Control Pro.	[JI6]
96	SCC-SP	Semaphore Communications Sec. Pro.	[HXH]
97	ETHERIP	Ethernet-within-IP Encapsulation	[RXH1]
98	ENCAP	Encapsulation Header	[RFC1241,RXB3]
99		any private encryption scheme	[IANA]
100	GMTP	GMTP	[RXB5]
101	IFMP	Ipsilon Flow Management Protocol	[Hinden]
102	PNNI	PNNI over IP	[Callon]
103	PIM	Protocol Independent Multicast	[Farinacci]

Decimal	Keyword	Protocol	References
104	ARIS	ARIS	[Feldman]
105	SCPS	SCPS	[Durst]
106	QNX	QNX	[Hunter]
107	A/N	Active Networks	[Braden]
108	IPComp	IP Payload Compression Protocol	[RFC2393]
109	SNP	Sitara Networks Protocol	[Sridhar]
110	Compaq-Peer	Compaq Peer Protocol	[Volpe]
111	IPX-in-IP	IPX in IP	[Lee]
112	VRRP	Virtual Router Redundancy Protocol	[Hinden]
113	PGM	PGM Reliable Transport Protocol	[Speakman]
114		any 0-hop protocol	[IANA]
115	L2TP	Layer Two Tunneling Protocol	[Aboba]
116	DDX	D-II Data Exchange (DDX)	[Worley]
117	IATP	Interactive Agent Transfer Protocol	[Murphy]
118	STP	Schedule Transfer Protocol	[JMP]
119	SRP	SpectraLink Radio Protocol	[Hamilton]
120	UTI	UTI	[Lothberg]
121	SMP	Simple Message Protocol	[Ekblad]
122	SM	SM	[Crowcroft]
123	PTP	Performance Transparency Protocol	[Welzl]
124	ISIS	over IPv4	[Przygienda]
125	FIRE		[Partridge]
126	CRTP	Combat Radio Transport Protocol	[Sautter]

Decimal	Keyword	Protocol	References
127	CRUDP	Combat Radio User Datagram	[Sautter]
128	SSCOPMCE		[Waber]
129	IPLT		[Hollbach]
130	SPS	Secure Packet Shield	[McIntosh]
131	PIPE	Private IP Encapsulation within IP	[Petri]
132	SCTP	Stream Control Transmission Protocol	[Stewart]
133	FC	Fibre Channel	[Rajagopal]
134	FRSVP-E2E-IGNORE		[RFC3175]
135-254	Unassigned	[IANA]	
255	Reserved	[IANA]	

There are 37 fewer available protocol numbers since the course in 1999.

Basic communication mechanism: datagram

Properties of datagrams:

- Best effort
- Each message handled independently — global addressing.
- IP packets (datagrams) are forwarded according to the network address (which is in each datagram) by routers.

Basic Ethernet + IP Software Architecture

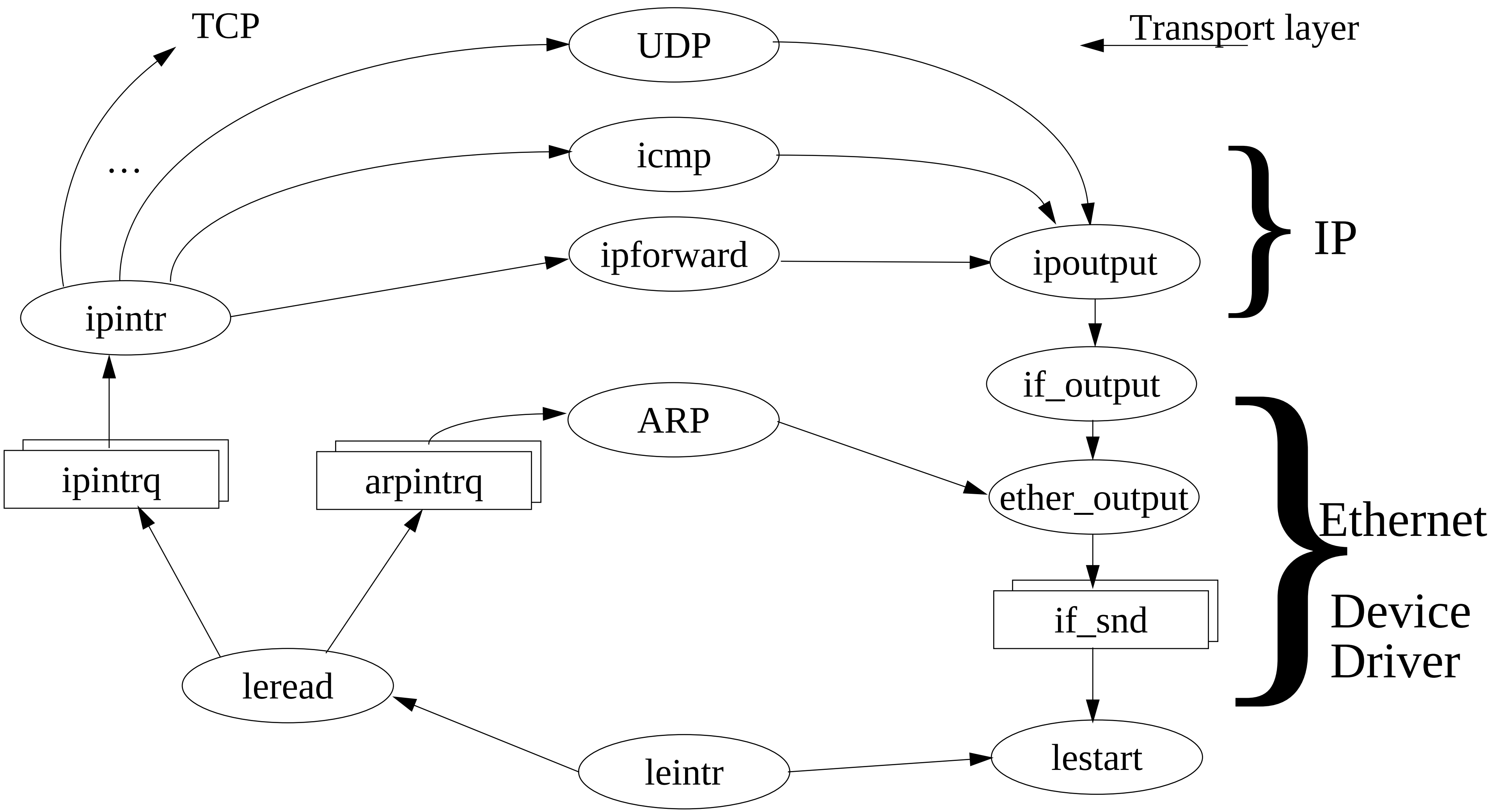

Figure 5: Basic Ethernet + IP Architecture - based on Stevens, TCP/IP Illustrated, Volume 2

Common Used Simple Services

Name	TCP port	UDP port	RFC	Description
echo	7	7	862	server returns what the client sends
discard	9	9	863	server discards what the client sends
daytime	13	13	867	Server returns the time and date in a human readable format
chargen	19	19	864	TCP server sends a continual stream of character, until the connection is terminated by the client. UDP server sends a datagradm containing a random number of characters each time the client sends a datagram.
ftp-data	20			File Transfer Protocol (Data)
ftp	21			File Transfer Protocol (Control)
telnet	23			Virtual Terminal Protocol
smtp	25			Simple Mail Transfer Protocol
time	37	37	868	Server returns the time as a 32-bit binary number. This number is the time in seconds since 1 Jan. 1990, UTC

Link Layer

Possible link layers include:

- Ethernet and IEEE 802.3 Encapsulation
  - with possible Trailer Encapsulation
- SLIP: Serial Line IP
- CSLIP: Compress SLIP
- PPP: Point to Point Protocol
- Loopback Interface
- Virtual Interface
- …
- carrier pigeons - CPIP (Carrier Pigeon Internet Protocol) April 1st 1990, RFC 1149 was written. A protocol for IP over avian carriers. Implementation (April 28 2001): *http://www.blug.linux.no/rfc1149/*

Some of the issues concerning links are:

- MTU and Path MTU
- Serial line throughput

LAN Protocols

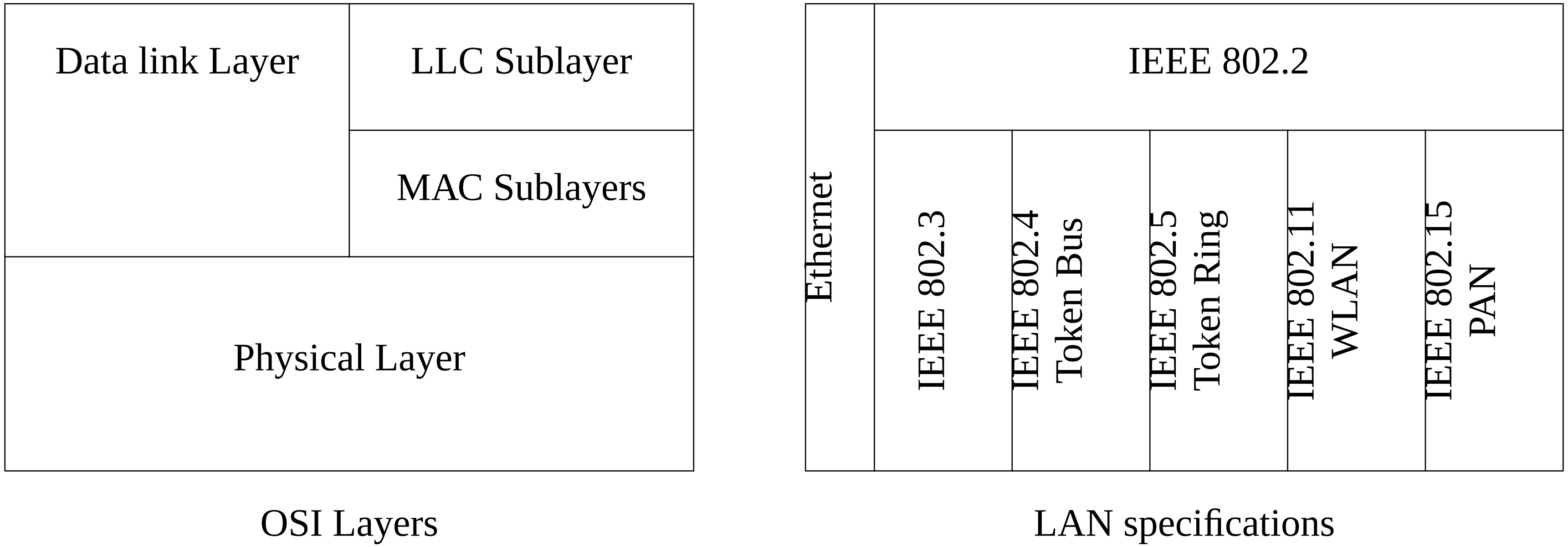

Figure 6: Physical and Link layer protocols used for LANs

Ethernet Encapsulation

Ethernet Encapsulation (RFC 894)

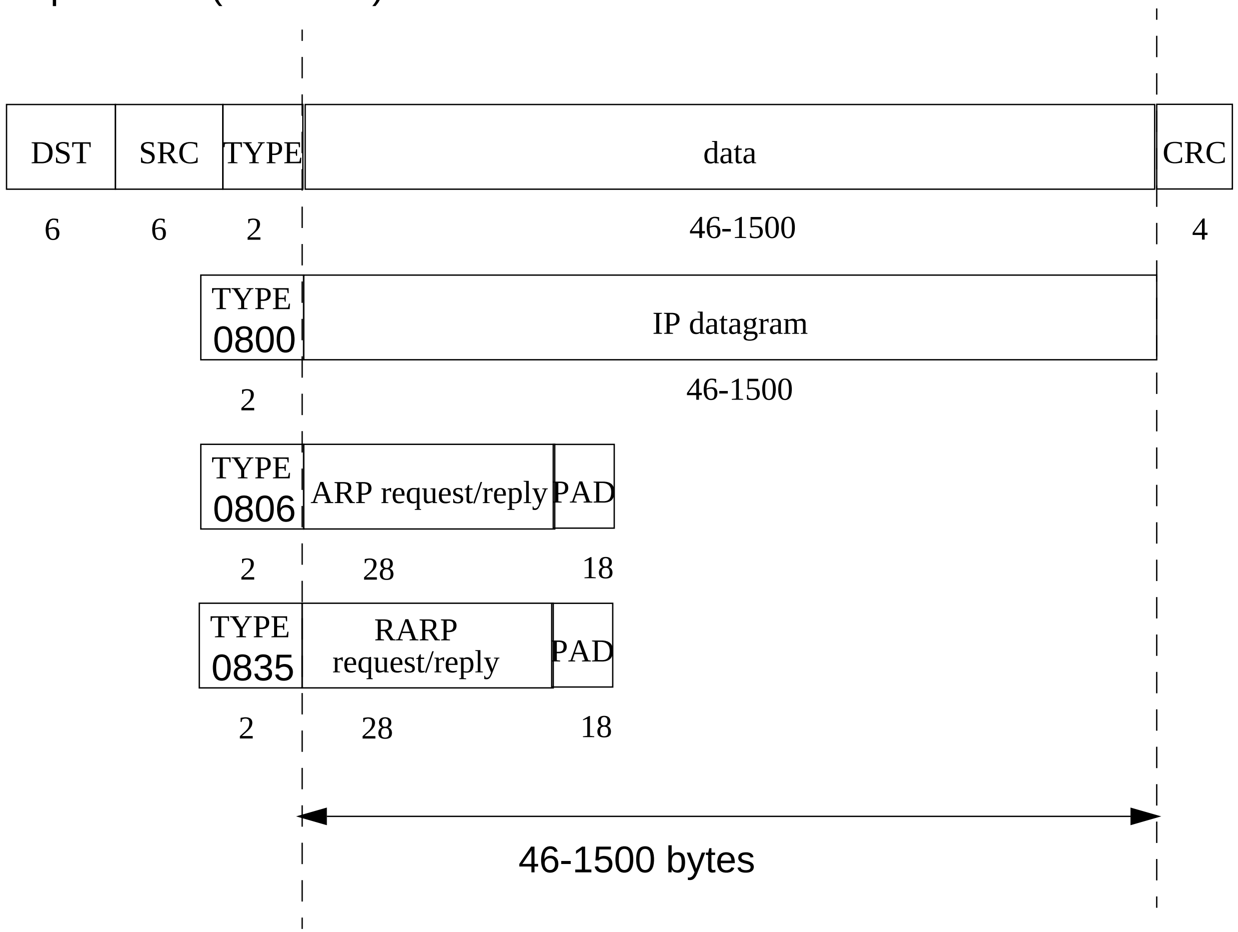

Figure 7: Ethernet encapsulation
(see Stevens, Volume 1, figure 2.1, pg. 23)

IEEE 802.3 Encapsulation

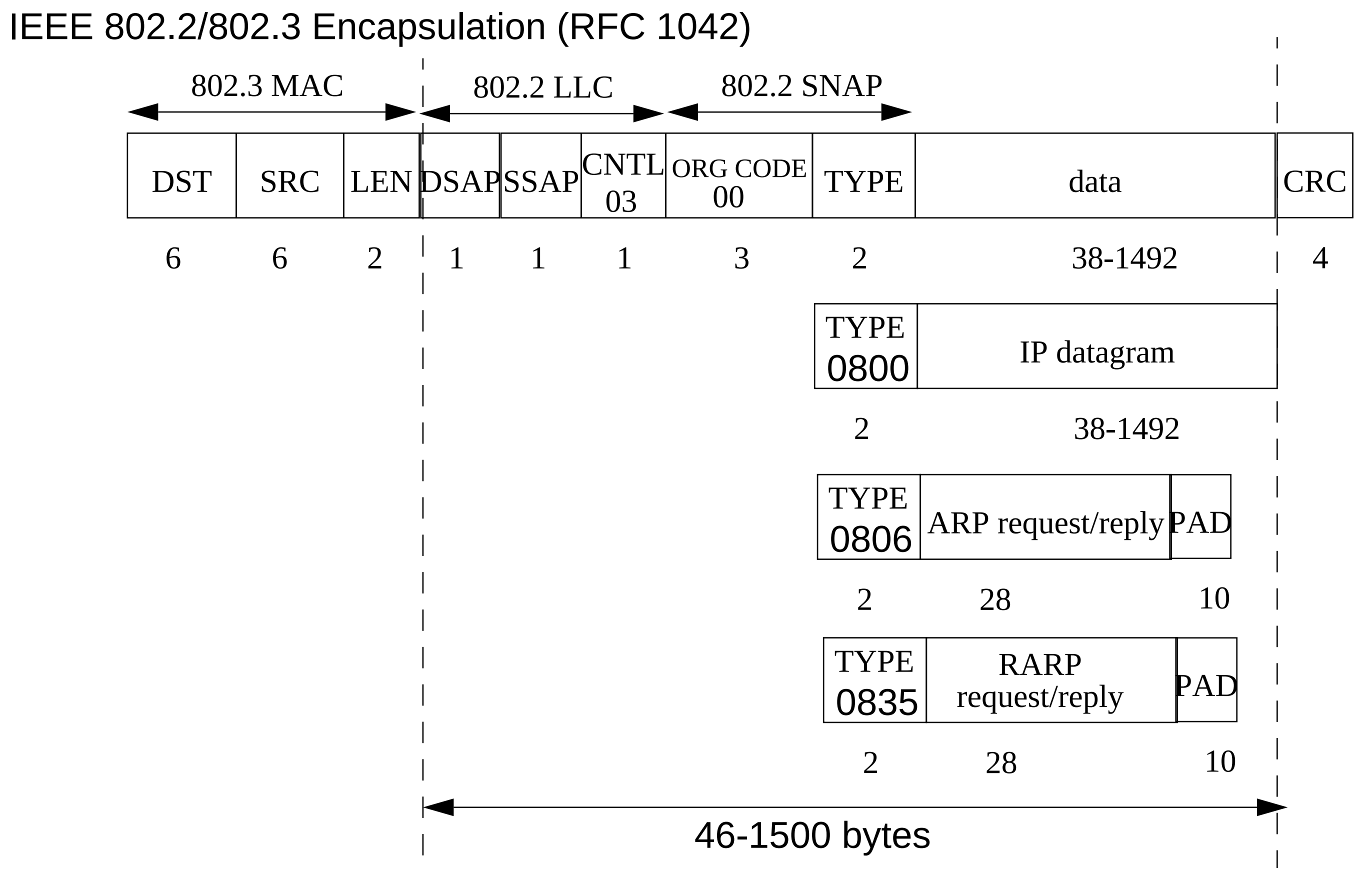

Figure 8: IEEE802.2/802.3 (see Stevens, Volume 1, figure 2.1, pg. 23)

DSAP ≡ Destination Service Access Point; SSAP ≡ Source Service Access Point; SNAP ≡ Sub-Network Access Protocol; for other TYPE values see RFC1700.

IEEE 802 Numbers of Interest

"… IEEE 802 Networks. These systems may use a Link Service Access Point (LSAP) field in much the same way the MILNET uses the "link" field. Further, there is an extension of the LSAP header called the Sub-Network Access Protocol (SNAP).

The IEEE likes to describe numbers in binary in bit transmission order, which is the opposite of the big-endian order used throughout the Internet protocol documentation."

Assignments from RFC1700

Link Service Access Point			**Description**	**References**
IEEE binary	Internet binary	decimal		
00000000	00000000	0	Null LSAP	[IEEE]
01000000	00000010	2	Individual LLC Sublayer Mgt	[IEEE]
11000000	00000011	3	Group LLC Sublayer Mgt	[IEEE]
00100000	00000100	4	SNA Path Control	[IEEE]
01100000	00000110	6	Reserved (DOD IP)	[RFC76]
01110000	00001110	14	PROWAY-LAN	[IEEE]
01110010	01001110	78	EIA-RS 511	[IEEE]
01111010	01011110	94	ISI IP	[JBP]
01110001	10001110	142	PROWAY-LAN	[IEEE]
01010101	10101010	170	SNAP	[IEEE]
01111111	11111110	254	ISO CLNS IS 8473	[RFC926]
11111111	11111111	255	Global DSAP	[IEEE]

SLIP (RFC 1055)

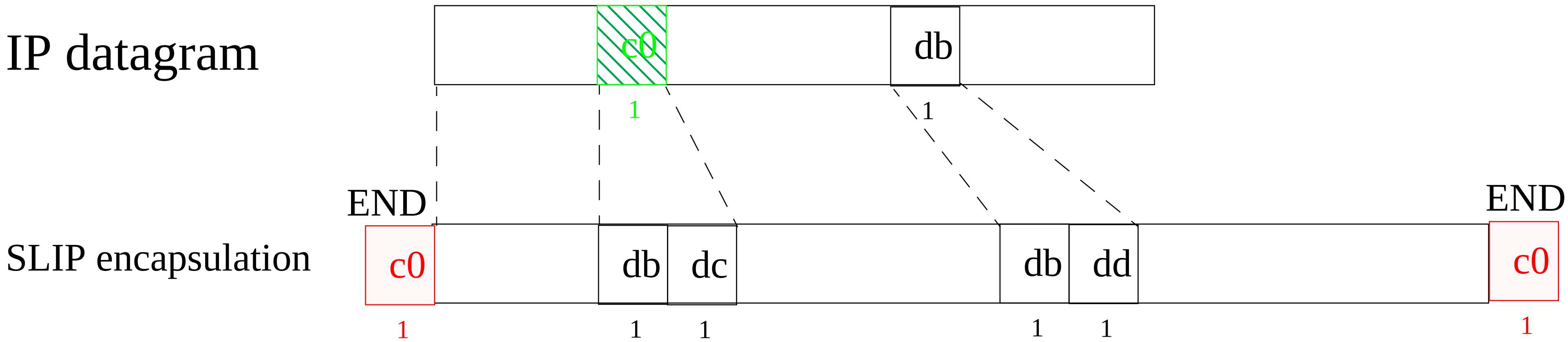

Figure 9: SLIP Encapsulation (see Stevens, Volume 1, figure 2.2, pg. 25)

RFC 1055: Nonstandard *for transmission of IP datagrams over serial lines: SLIP*

SLIP uses character stuffing, SLIP ESC character ≡ 0xdb

SLIP END character ≡ 0xc0

- point to point link, ⇒ no IP addresses need to be sent
- there is no TYPE field, ⇒ you can only be sending IP, i.e., can't mix protocols
- there is no CHECKSUM, ⇒ error detection has to be done by higher layers

SLIP Problems ⇒CSLIP ≡ Compressed SLIP

- because many users running SLIP over lines at 19.2 kbits/s or slower
- lots of interactive traffic (telnet, rlogin, …) which uses TCP
  - many small packets
  - each of which needs a TCP header (20 bytes) + IP header (20 bytes) ⇒ overhead 40 bytes
  - Send 1 user character requires sending a minimum of: 1 + 40 + END, i.e., 42 bytes
  - most of the header is **predictable**

CSLIP (RFC 1144: Compressing TCP/IP headers for low-speed serial links, by Van Jacobson) reduces the header to 3-5 bytes, by:

- trying to keep response time under 100-200ms
- keeping state about ~16 TCP connections at each end of the link
  - the 96-bit tuple <src address, dst address, src port, dst port> reduced to 4 bits
- many header fields rarely change - so don't transmit them
- some header fields change by a small amount - just send the delta
- no compression is attempted for UDP/IP
- a 5 byte compressed header on 100-200 bytes ⇒ 95-98% line efficiency

Robust Header Compression (rohc)

Header compression schemes that perform well over links with high error rates and long roundtrip times.

http://www.ietf.org/html.charters/rohc-charter.html

PPP: Point to Point Protocol (RFC 1331, 1332)

PPP corrects the deficiencies in SLIP. PPP consists of:

- encapsulation for either async or synchronous links,
  - HDLC (see RFC 1549)
  - X.25 (see RFC 1598)
  - ISDN (see RFC 1618)
  - SONET/SDH (see RFC 1619)
- Link Control Protocol
  - establish, configure, and test data-links [includes option negociation]
  - authentication (see RFC 1334)
- Family of Network Control Protocols (NCPs) - specific to different network protocols, currently:
  - IP (see RFC 1332)
  - DECnet (see RFC 1376)
  - OSI network layer (see RFC 1377)
  - AppleTalk (see RFC 1378)
  - XNS (see RFC 1764)

See "PPP Design, Implementation, and Debugging", by James D. Carlson, Second edition, Addison-Wesley,2000, ISBN 0-201-70053-0.

PPP frames

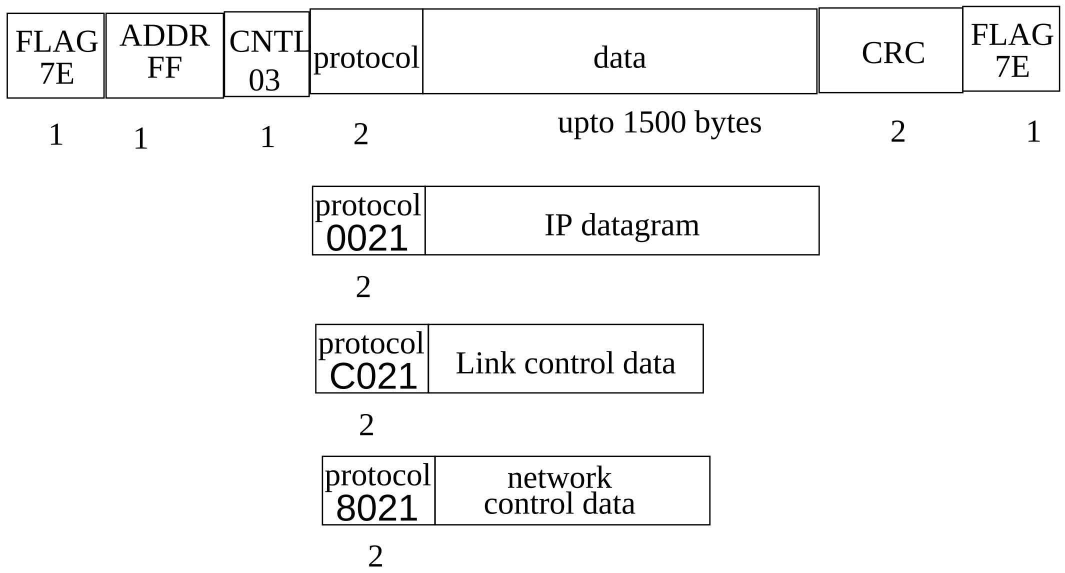

Figure 10: Format of PPP frame (see Stevens, Volume 1, figure 2.3, pg. 26)

- The protocol field behaves like the Ethernet TYPE field.
- CRC can be used to detect errors in the frame.
- Either character or bit stuffing is done depending on the link.
- you can negotiate away the CNTL and ADDRESS fields, and reduce the protocol field to 1 byte ⇒ minimum overhead of 3 bytes
- Van Jacobson header compression for IP and TCP

PPP summary

- support for multiple protocols on a link
- CRC check on every frame
- dynamic negociation of IP address of each end
- header compression (similar to CSLIP)
- link control with facilities for negotiating lots of data-link options

All at a price averaging 3 bytes of overhead per frame.

Loopback interface

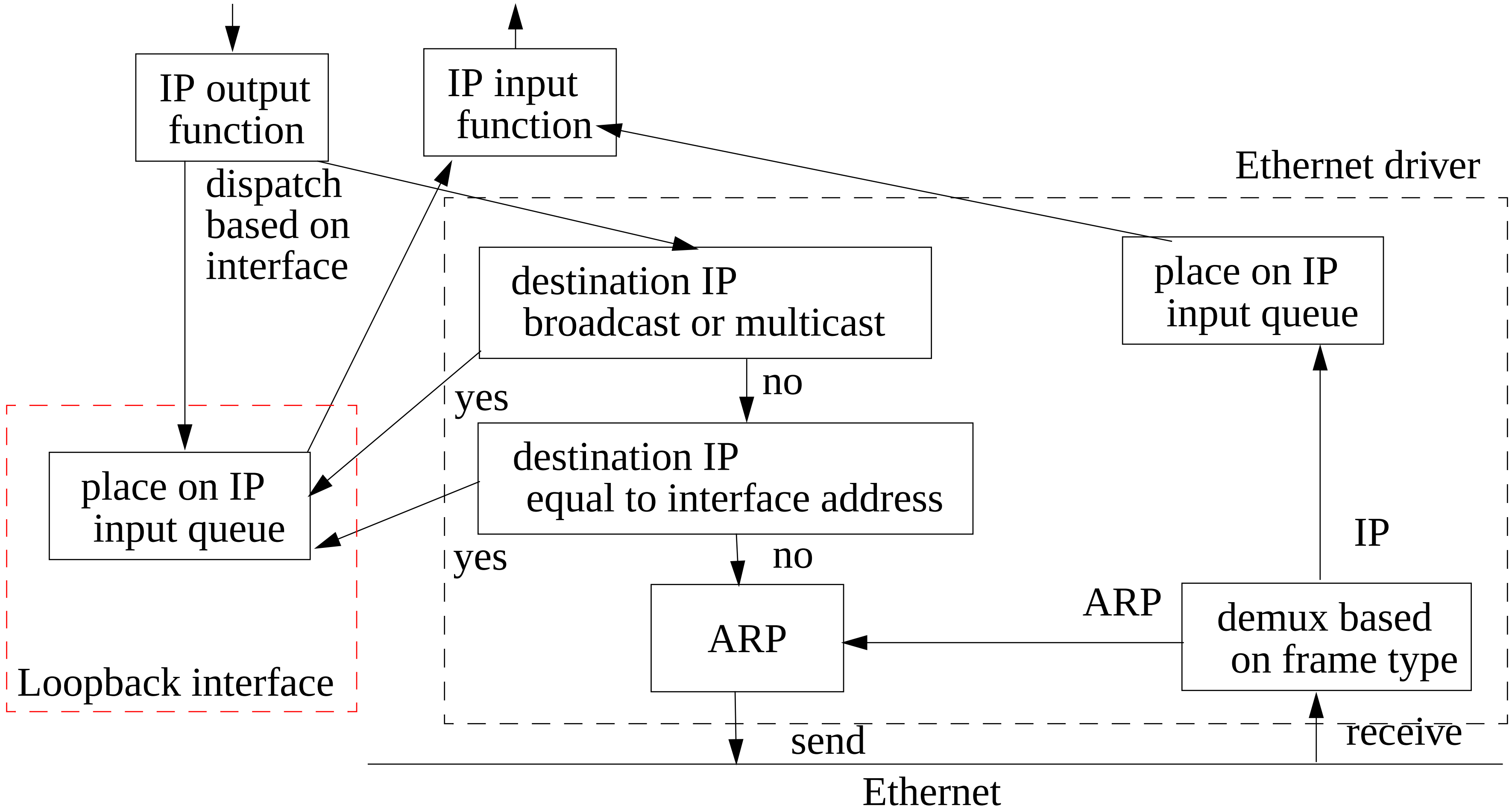

Figure 11: Processing of IP Datagrams (adapted from Stevens, Volume 1, figure 2.4, pg. 28)

Loopback interface summary

- loopback address ≡ 127.0.0.1 generally called “localhost”
- all broadcasts and multicasts get sent to the loopback - because the sender gets a copy too!
- everything sent to the host’s **own** IP address is sent to the loopback interface

Virtual Interface (VIF)

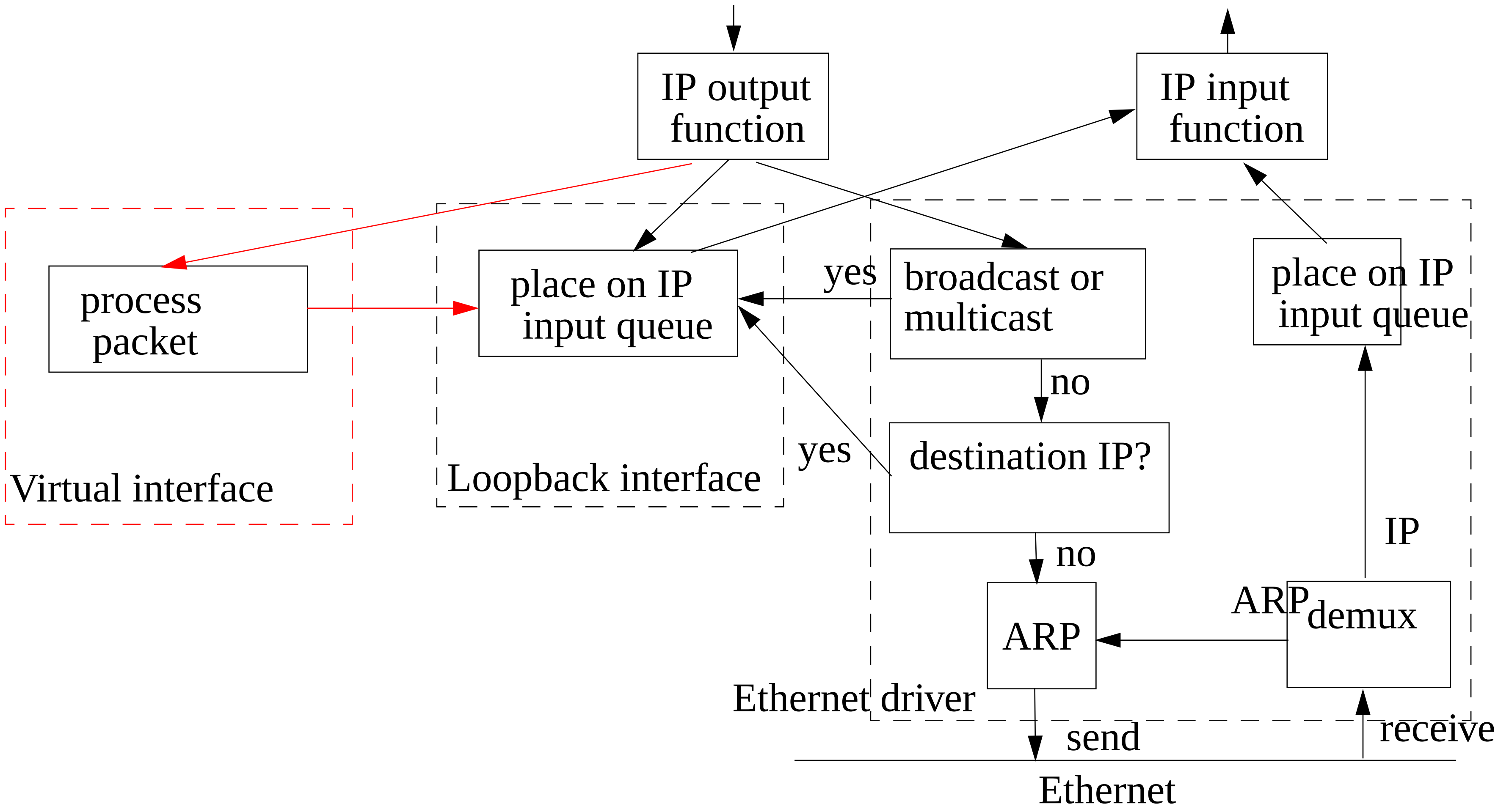

Figure 12: Processing of network packets via a Virtual Interface

Using VIF for tunneling

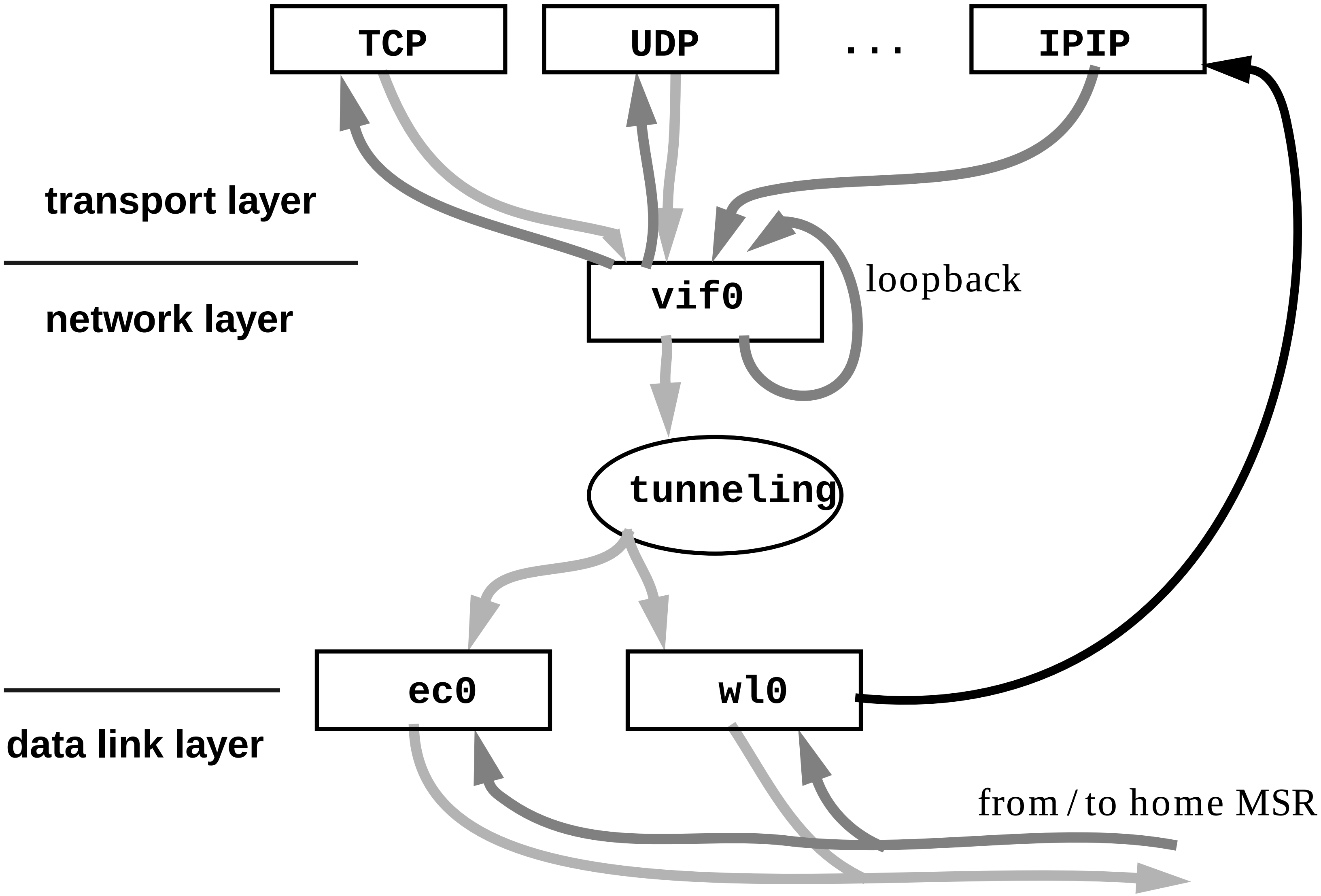

Figure 13: Using a Virtual Interface for Tunneling (IP in IP) adapted from John Ioannidis's thesis

MTU≡Maximum Transmission Unit

MTU is a characteristic of the link layer

Typical MTUS

MTU	Network	
65535	Official maximum MTU	
17914	16Mb IBM Token Ring	
8166	IEEE 802.4	
4464	IEEE 802.5 (4Mb max)	
4352	FDDI (Revised)	
2048	Wideband Network	
2002	IEEE 802.5 (4Mb recommended)	
1536	Experimental Ethernet Nets	
1500	Ethernet Networks	
1500	Point-to-Point (default)	
1492	IEEE 802.3	
1006	SLIP	simply a logical limit for interactive response
1006	ARPANET	
576	X.25 Networks	
544	DEC IP Portal	
512	NETBIOS	
508	IEEE 802/Source-Route Bridge	
296	Point-to-Point (low delay)	
68	Official minimum MTU	

Path MTU

- Each link in path from source to destination can have a different MTU
- to avoid fragmentation you have to find the minimum of these
- RFC 1191: Path MTU discovery uses:
  - “good” guesses (i.e., likely values)
  - By setting Don’t Fragment (DF) bit in IP datagram $\Rightarrow$ change size while you get ICMP messages saying “Destination Unreachable” with a code saying fragmentation needed

Serial line throughput

At 9,000 bits/sec, 8 bits per byte, plus 1 start and 1 stop bit, i.e., 960 bytes/sec, then transferring 1024 byte packets would take 1066 ms

- too long for interactive limits; since the average wait would be 533 ms

∴ shorten the MTU to 296 bytes ⇒ 266 ms/frame or ~133 ms average wait

With 5 bytes of CSLIP header and 256 bytes of data (in the 261 byte frame) ⇒

- 98.1% utilization of link for data and
- 1.9% for header

For single bytes of interactive traffic, the round trip-time is 12.5 ms

Caveats:

- assumes that you give interactive traffic priority
- error correcting and compression in the modem can complicate the calculations - since the modem has to delay traffic to have more to compress and compression takes time

IPv4 (RFC 791)

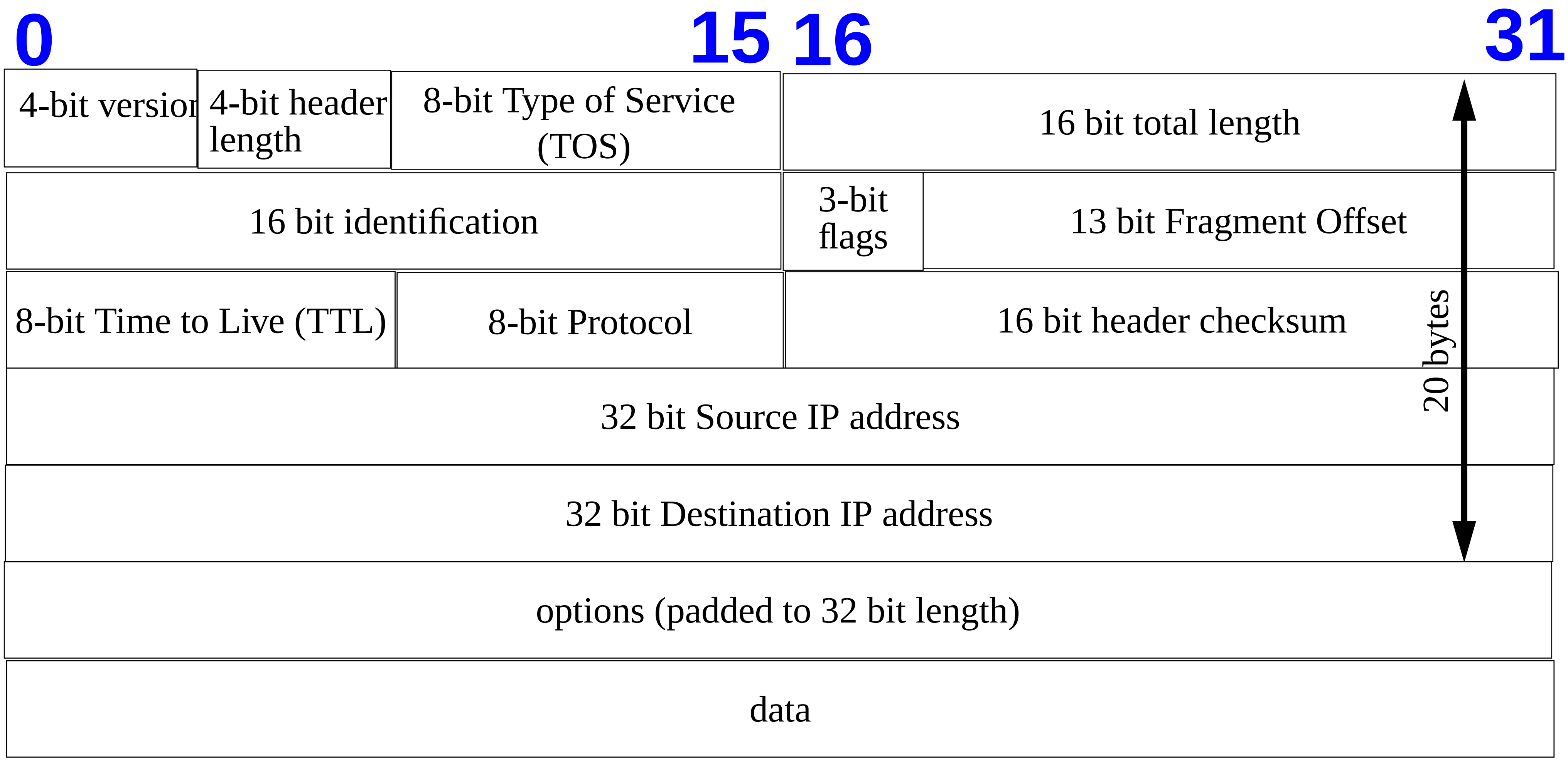

Figure 14: IP header (see Stevens, Vol. 1, figure 3.1, pg. 34)

Type of Service (TOS)

Type of Service: 8 bits

Bits 0-2: Precedence
Bit 3: 0 = Normal Delay, 1 = Low Delay
Bits 4: 0 = Normal Throughput, 1 = High Throughput
Bits 5: 0 = Normal Relibility, 1 = High Relibility.

0	1	2	3	4	5	6	7
Precedence			DELAY	T	R	M	Reserved

Few applications set the TOS field (in fact most implementations will not let you set these bits!) However, 4.3BSD Reno and later - do support these bits.

Differentiated Services (diffserv) proposes to use 6 of these bits to provide 64 priority levels - calling it the Differentiated Service (DS) field [RFC2474]

SLIP guesses by looking at the **protocol** field and then checks the source and destination **port** numbers.

Recommended Value for TOS Field

Application	Minimum delay	Maximize throughput	Maximize reliability	Minimize monitary cost	Hex value
telnet/rlogin	1	0	0	0	0x10
FTP					
control	1	0	0	0	0x10
data	0	1	0	0	0x08
any bulk data	0	1	0	0	0x08
TFTP	1	0	0	0	0x10
SMTP					
command phase	1	0	0	0	0x10
data phase	0	1	0	0	0x08
DNS					
UDP query	1	0	0	0	0x10
TCP query	0	0	0	0	0x00
zone transfer	0	1	0	0	0x08
ICMP					
error	0	0	0	0	0x00
query	0	0	0	0	0x00
any IGP	0	0	1	0	0x04
SNMP	0	0	1	0	0x04
BOOTP	0	0	0	0	0x00
NNTP	0	0	0	1	0x02

Precedence

Precedence values are defined but are largely ignored, few applications use them.

111	Network Control
110	Internetwork Control
101	CRITIC/ECP
100	Flash Override
011	Flash
010	Immediate
001	Priority
000	Routine

In the original ARPANET there were two priority levels defined (in order to support low delay services and regular traffic).

Problems with precedence

- As soon as people found that high priority meant something $\Rightarrow$ all traffic was sent with this bit set!

So unless there is a added cost/policy check/… associated with usage of a precedence level - it is very likely going to be abused.

Precendence and telephony systems

Similar precedence systems exist in most national telephony systems.

Q: What are the A, B, C and D touch tone keys used for? …

A: These are extensions to the standard touch-tones (0-9, *, #) which originated with the U.S. military's Autovon phone network. The original names of these keys were FO (Flash Override), F (Flash), I (Immediate), and P (Priority). The various priority levels established calls with varying degrees of immediacy, terminating other conversations on the network if necessary. FO was the greatest priority, normally reserved for the President or very high ranking officials. P had a lesser priority, but still took precedence over calls that were placed without any priority established.

-- from TELECOM Digest - Frequently Asked Questions - v.8, 8 February 1997

Packet Size

Hosts **only** have to accept packets up to 576 bytes in size.

Most modern systems accept slightly larger than 8,196 + header bytes (to provide efficient file service for 8 Kbyte blocks).

IP addresses

32 bit address divided into two parts:

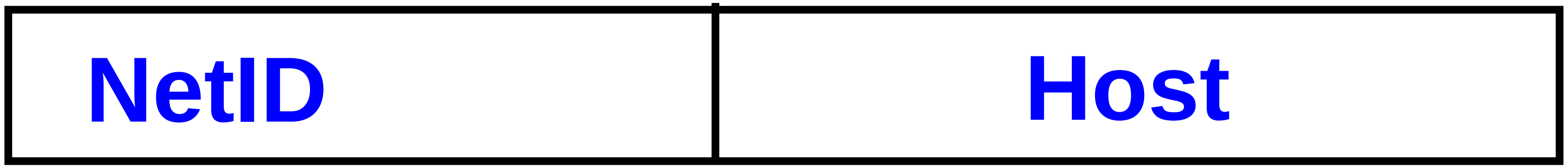

Figure 15: IP address format

Generally divide the address range into classes:

Class	NetID		Range (dotted decimal notation)	host ID
A	0	+ 7- bit NetID	0.0.0.0 to **127**.255.255.255	24 bits of host ID
B	1 0	+ 14-bit NetID	**128**.0.0.0 to **191**.255.255.255	16 bits of host ID
C	1 1 0	+ 21-bit NetID	**192**.0.0.0 to **223**.255.255.255	8 bits of host ID
D	1 1 1 0		**224**.0.0.0 to **239**.255.255.255	28 bits of Multicast address
E	1 1 1 1 0		**240**.0.0.0 to **247**.255.255.255	Reserved for future use

Subneting IP networks

Often we want to “subnet” - i.e., divide the network up into multiple networks:

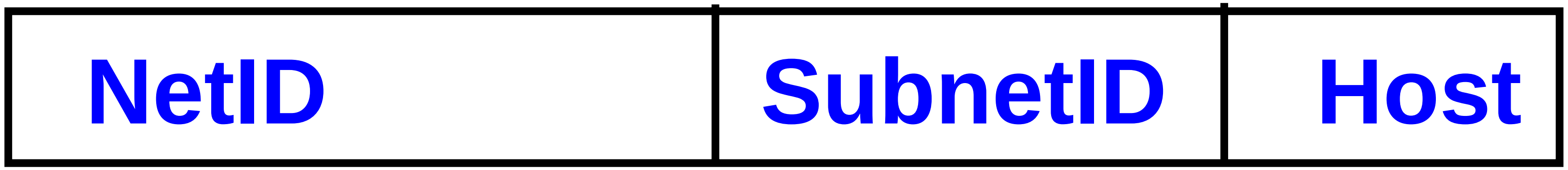

Figure 16: IP addresses and subnetting

Although the Subnet field is shown as a field which is separate from the Host field, it could actually be divided on a bit by bit basis; this is done by a **Subnet Mask.**

A common practice to avoid wasting large amounts of address space is to use Classless Interdomain Routing (CIDR) also called “supernetting” {see §10.8 of Steven’s Vol. 1 and RFCs 1518 and 1519}.

Special Case IP Addresses

IP Address			Can appear as		Description
net ID	**subnet ID**	**host ID**	**source?**	**destination**	
0		0	OK	never	**this** host on **this** net
0		**hostid**	OK	never	specified host on **this** net
127		**any**	OK	OK	loopback address
-1		-1	never	OK	limited broadcast (never forwarded)
netid		-1	never	OK	net-directed broadcast to **netid**
netid	**subnetid**	-1	never	OK	subnet-directed broadcast to **netid**, **subnetid**
netid	-1	-1	never	OK	all-subnets-directed broadcast to **netid**

Subnet mask

32 bit value with a 1 for NetID + subnetID, 0 for HostID

16 bits	8 bits	8 bits
NetID	SubnetID	HostID
1111 1111 1111 1111	1111 1111	0000 0000

Figure 17: Class B address with a 8 bit subnet address

2 different class B subnet arrangements

16 bits	8 bits	8 bits	
NetID	SubnetID	HostID	0xffffff00
1111 1111 1111 1111	1111 1111	0000 0000	255.255.255.0
			/24

16 bits	10 bits	6 bits	
NetID	SubnetID	HostID	0xffffffc0
1111 1111 1111 1111	1111 1111	1100 0000	255.255.255.192
			/26

Classless Inter-Domain Routing (CIDR)

Slash notation

Length (CIDR)	Address mask	Notes	Length (CIDR)	Address mask	Notes
/0	0.0.0.0	All 0's ≡ no mask	/8	255.0.0.0	≡ Class A
/1	128.0.0.0		/9	255.128.0.0	
/2	192.0.0.0		/10	255.192.0.0	
/3	224.0.0.0		/11	255.224.0.0	
/4	240.0.0.0		/12	255.240.0.0	
/5	248.0.0.0		/13	255.248.0.0	
/6	252.0.0.0		/14	255.252.0.0	
/7	254.0.0.0		/15	255.254.0.0	

Slash notation continued

Length (CIDR)	Address mask	Notes	Length (CIDR)	Address mask	Notes
/16	255.255.0.0	≡ Class B	/24	255.255.255.0	≡ Class C
/17	255.255.128.0		/25	255.255.255.128	
/18	255.255.192.0		/26	255.255.255.192	
/19	255.255.224.0		/27	255.255.255.224	
/20	255.255.240.0		/28	255.255.255.240	
/21	255.255.248.0		/29	255.255.255.248	
/22	255.255.252.0		/30	255.255.255.252	
/23	255.255.254.0		/31	255.255.255.254	
			/32	255.255.255.255	All 1’s (host specific mask)

IP address assignments

Internet Service Providers (ISPs) should contact their upstream registry or their appropriate Regional Internet Registries (RIR) at one of the following addresses:

Region	
APNIC (Asia-Pacific Network Information Center)	*http://www.apnic.net*
ARIN (American Registry for Internet Numbers)	*http://www.arin.net*
RIPE NCC (Reseau IP Europeens)	*http://www.ripe.net*

ifconfig, route and netstat Commands

- ifconfig: to configure interface.
- route: to update routing table.
- netstat: to get interface and routing information.

For example: to configure interface, add an network and add a gateway:

root# ifconfig eth0 192.71.20.115 netmask 255.255.255.0 up

root# route add -net 192. 71.20.0 netmask 255.255.255.0 eth0

root# route add default gw 192.71.20.1 eth0

We will discuss these commands in more detail in following lectures and in the recitations.

Summary

- Course Introduction
- Internet Basics
  - Internet trends
  - Internet standards, RFCs
- Link Layer Protocols for the Internet
  - Ethernet
  - SLIP, PPP
  - MTU, path MTU
- IP: Internet Protocol
  - IP addressing
  - Subnetting

W. Richard Stevens

- Born in Luanshya, Northern Rhodesia (now Zambia) in 1951
- Died on September 1, 1999
- He studied Aerospace Engineering, Systems Engineering (image processing major, physiology minor)
- flight instructor and programmer
- His many books helped many people to understand and use TCP/IP
  - UNIX Network Programming, Prentice Hall, 1990.
  - Advanced Programming in the UNIX Environment, Addison-Wesley, 1992.
  - TCP/IP Illustrated, Volume 1: The Protocols, Addison-Wesley, 1994.
  - TCP/IP Illustrated, Volume 2: The Implementation, Addison-Wesley, 1995.
  - TCP/IP Illustrated, Volume 3: TCP for Transactions, HTTP, NNTP, and the UNIX Domain Protocols, Addison-Wesley, 1996.
  - UNIX Network Programming, Volume 1, Second Edition: Networking APIs: Sockets and XTI, Prentice Hall, 1998.
  - UNIX Network Programming, Volume 2, Second Edition: Interprocess Communications, Prentice Hall, 1999.