High Performance Networking

HIPPI NETWORK

CERN 1998

Arie Van Praag CERN IT PDP 1211 Geneva 23 Switzerland
E-Mail: a.van.praag@cern.ch

CERN

High Performance Networking

TO-DAY

- 1 High Performance Networking as sign of its time.
  - A Historical Overview of Hardware and Protocols
- 2 Yesterdays High Performance Networks
  - Ultranet, HIPPI, SCI, Fibre Channel, Myrinet, Gigabit Ethernet
- 3 GSN (the first 10 Gbit/s network and the first secure data network)
  - Physical Layer, Error Correction, ST Protocol, SCSI-ST
- 4 Infiniband (the imitating 2.5 – 30 Gbit/s interconnect)
  - Physical Layer, Protocols, Network Management
- 5 SONET and some facts about DWDM,
  - 10 Gigabit Ethernet, Physical Layers, Coupling to the WAN

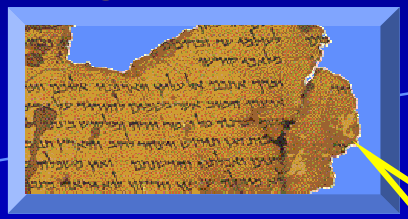
Arie Van Praag CERN IT/ADC
1211 Geneva 23 Switzerland
E-mail a.van.praag@cern.ch

High Performance Networking

Running as Fast as you Can

Every New Network has been High Performance in its own Time

Running Slaves ± 5 Km/h



1000 - 300 B Chr.

Horse Riding ± 20 Km/h

One of the Dead Sea scrolls that is clearly a message from a general to its officers in the Field.

Shrine of the Book,
Israel Museum, Jerusalem

ARIE VAN PRAAG — CERN IT PDP — E-Mail: a.van.praag@cern.ch

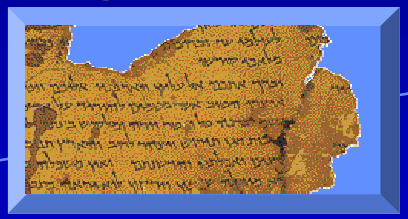
CERN

High Performance Networking

Running as Fast as you Can

Every New Network has been High Performance in its own Time

Running Slaves ± 5 Km/h



300 B Chr.

Horse Riding ± 20 Km/h

Using Birds ± 60 Km/h

ARIE VAN PRAAG — CERN IT PDP — E-Mail: a.van.praag@cern.ch

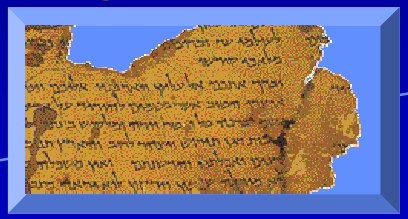
CERN

High Performance Networking

Running as Fast as you Can

Every New Network has been High Performance in its own Time

Running Slaves ± 5 Km/h



300 B Chr.

Horse Riding + 20 Km/h

Using Bi + 60 Km/h

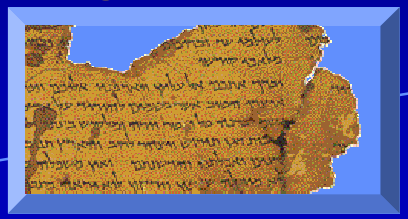
ARIE VAN PRAAG — CERN IT PDP — E-Mail: a.van.praag@cern.ch

High Performance Networking

Running as Fast as you Can

Every New Network has been High Performance in its own Time

Running Slaves ± 5 Km/h



300 B Chr.

Hors Riding + 20 Km/h

Using Bi + 60 Km/h

Anyhow we are far from this

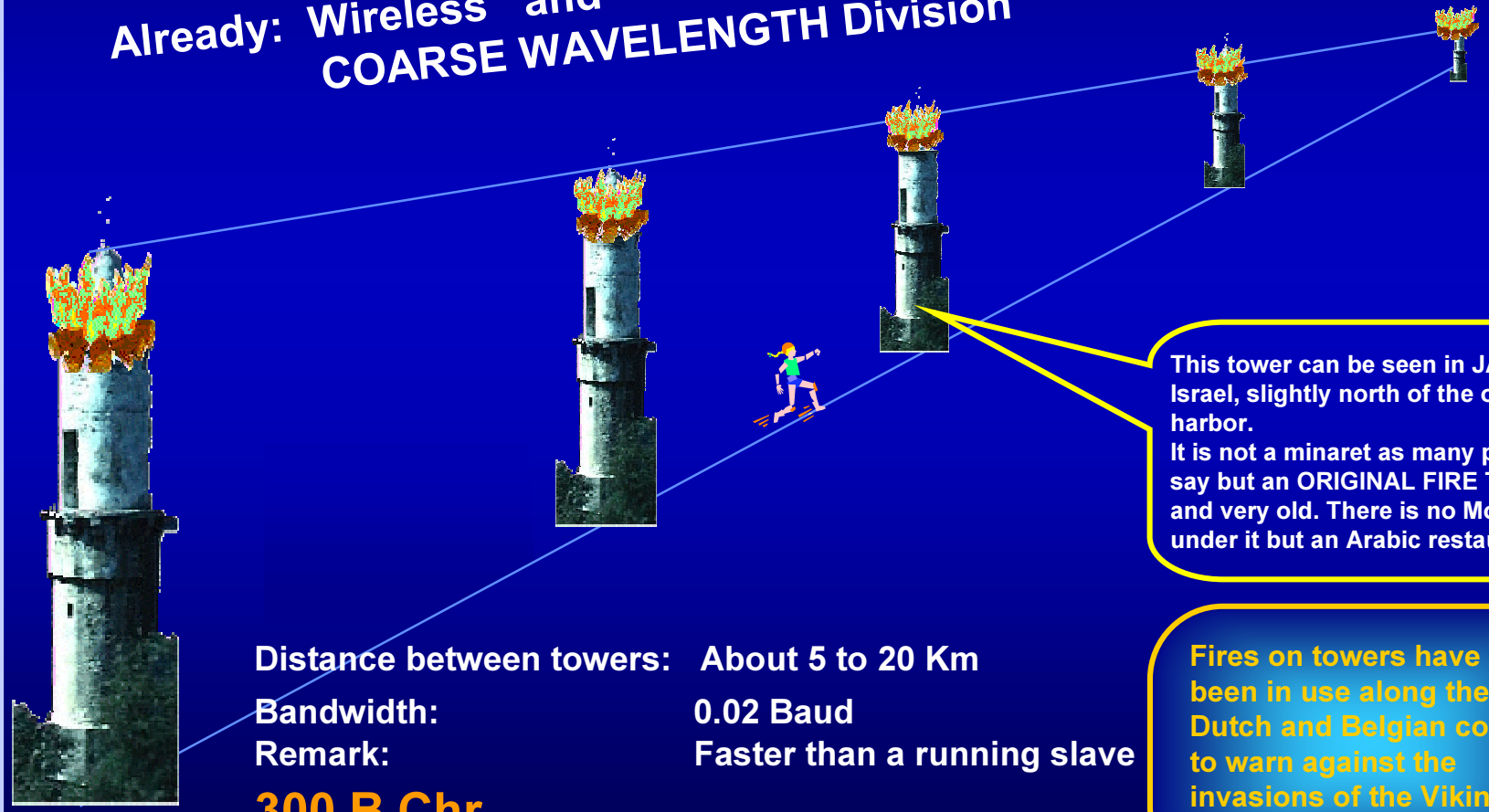
ARIE VAN PRAAG — CERN IT PDP — E-Mail: a.van.praag@cern.ch

CERN

High Performance Networking

The Fire Connection

Already: Wireless and
COARSE WAVELENGTH Division



This tower can be seen in JAFFA, Israel, slightly north of the old harbor.
It is not a minaret as many people say but an ORIGINAL FIRE TOWER and very old. There is no Mosque under it but an Arabic restaurant

Distance between towers: About 5 to 20 Km

Bandwidth: 0.02 Baud

Remark: Faster than a running slave

300 B Chr.

Fires on towers have still been in use along the Dutch and Belgian coast to warn against the invasions of the Vikings
± 800 - 1000

ARIE VAN PRAAG — *CERN IT PDP* — *E-Mail: a.van.praag@cern.ch*

High Performance Networking

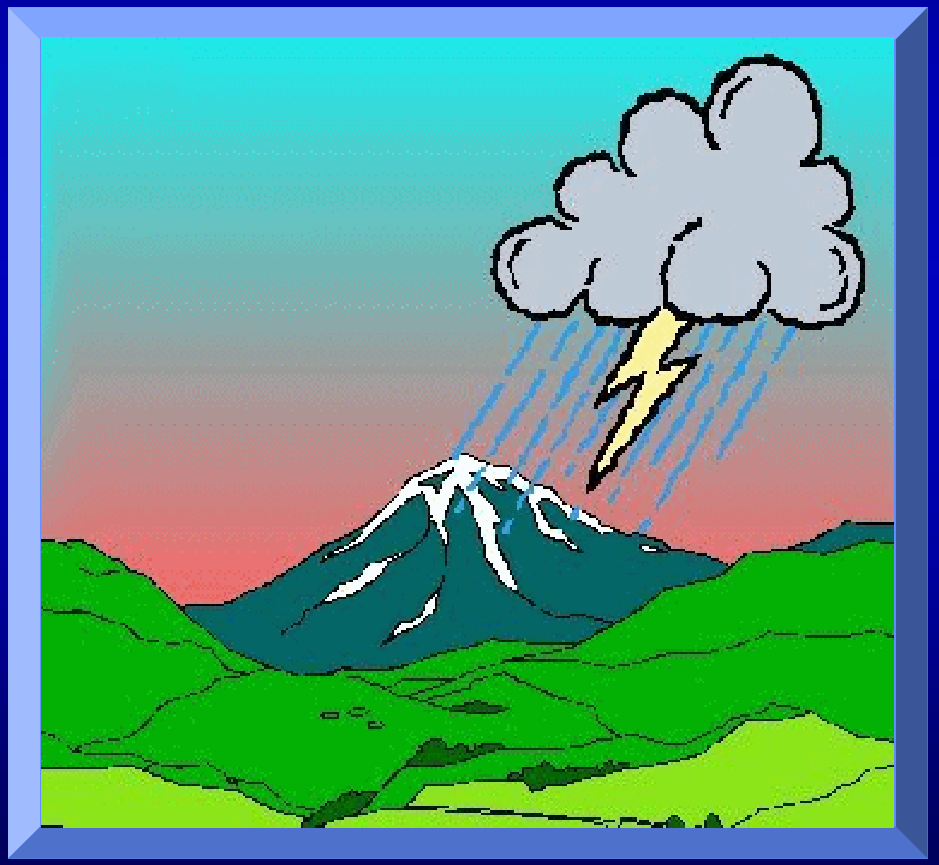
Smoke Signals

Some Clever People Invented **BROADCASTING**

Estimated Distance: 5 - 30 Km
Bandwidth: 0.2 Baud

HOW ? ?

>> 1850



ARIE VAN PRAAG — CERN IT PDP — E-Mail: a.van.praag@cern.ch

High Performance Networking

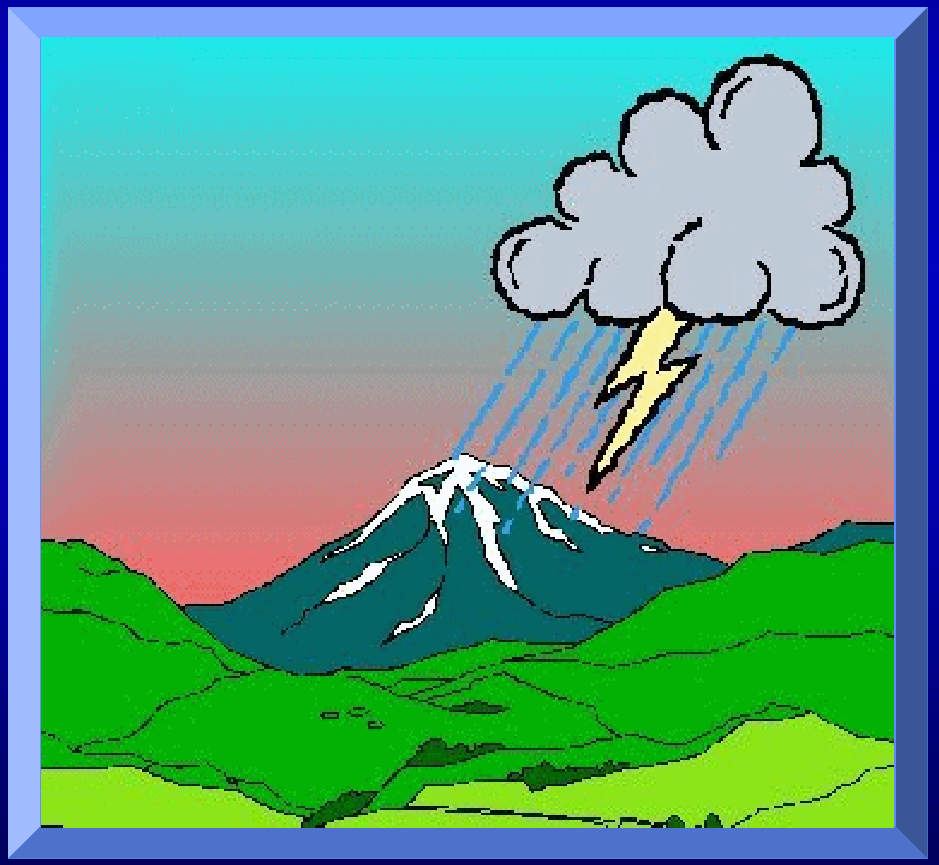
Smoke Signals

Some Clever People Invented **BROADCASTING**

Estimated Distance: 5 - 30 Km
Bandwidth: 0.2 Baud

HOW ? ?

>> 1850

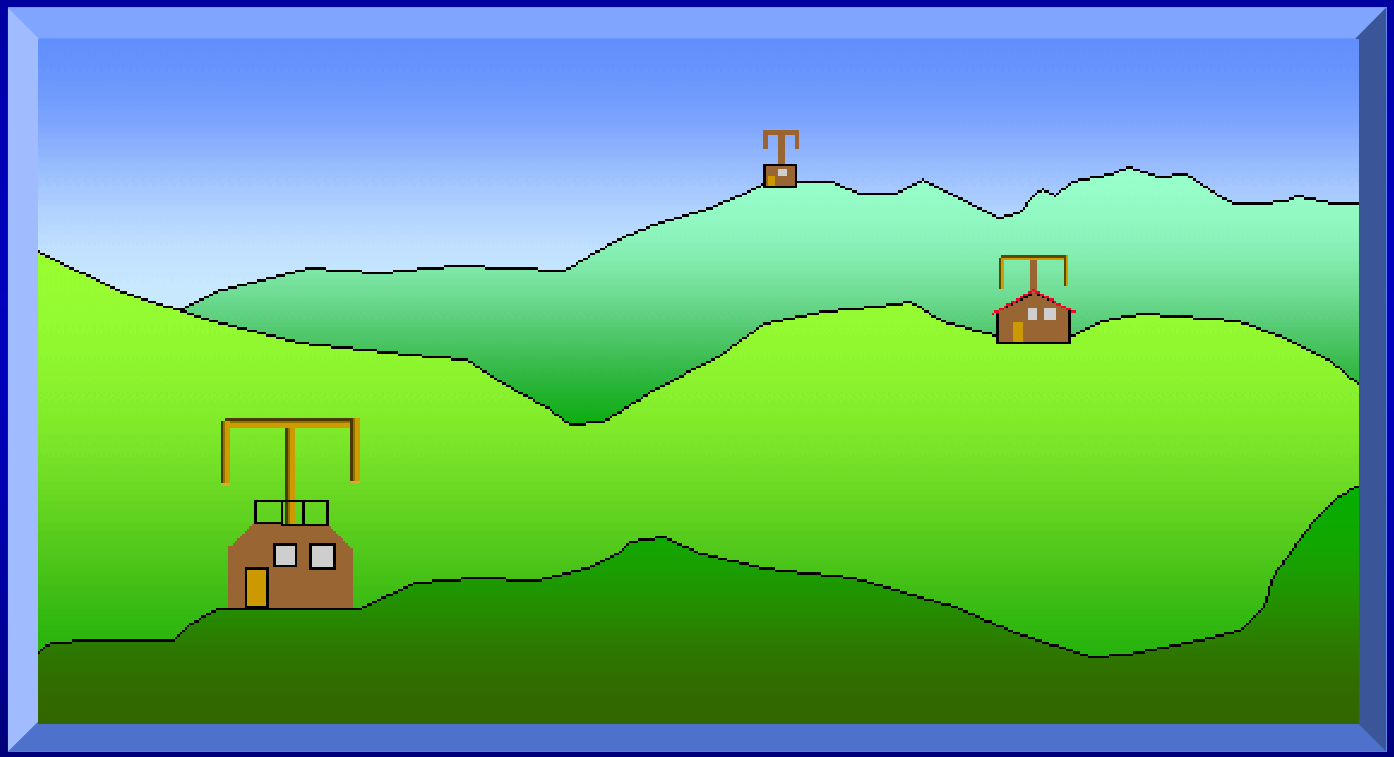


Now to day still used by the Vatican to announce a newly elected Pope

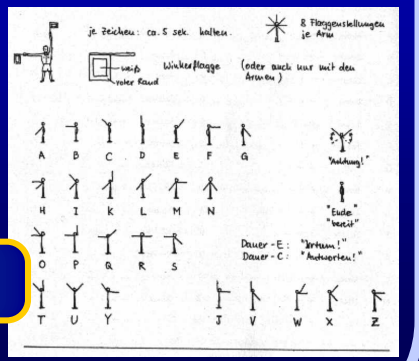
ARIE VAN PRAAG — CERN IT PDP — E-Mail: a.van.praag@cern.ch

High Performance Networking
Semaphores

Semaphore type of Networks came in use around **1783** 1 Byte/s



Invented in France under the King Sun (Louise XIV) and used during his war against the state of Holland



A living language that is still learned and used by scouts now to day.

ARIE VAN PRAAG — *CERN IT PDP* — *E-Mail: a.van.praag@cern.ch*

High Performance Networking
Other Semaphores

SEMAPHORES were in use until the late 50^{s}
to Indicate Water Level or Wind and Other Static Messages.

They still exist
as monuments:
Bremer Harbor, Scheveningen,

What About: Data Security

ARIE VAN PRAAG — CERN IT PDP — E-Mail: a.van.praag@cern.ch

High Performance Networking

There once was a Painter

A	. -	T	-
B	- . . .	U	. . -
C	- . - .	V	. . . -
D	- . .	W	. - -
E	.	X	- . . -
F	. . - .	Y	- . - -
G	- - .	Z	- - . .
H	. . . .	0	- - - - -
I	. .	1	. - - - -
J	. - - -	2	. . - - -
K	- . -	3	. . . - -
L	. - . .	4	. . . . -
M	- -	5	. . . . .
N	- .	6	- . . . .
O	- - -	7	- - . . .
P	. - - .	8	- - - . .
Q	- - . -	9	- - - - .
R	. - .	/	- . . - .
S	. . .	?	. . - - . .
= (BT break)			- . . . -
+ (AR fin du message)			. - . - .
. (point)			. - . - . -
, (virgule)			- - . . - -
@ (SK fin de transmission)			. . . - . -

Son of a Reverend

SAMUEL MORSE

He invented the first Electric Network in **1845** and the corresponding machine language:

MORSE

First demonstrated between Washington and Baltimore

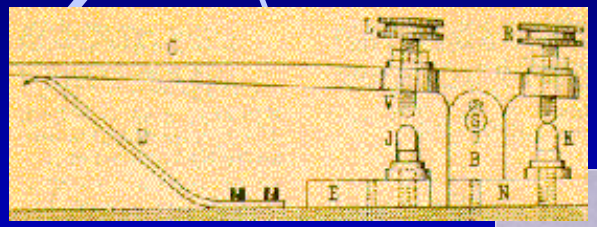
Officially abandoned in 2001, but still used today.

Bandwidth: ± 30 Bytes/s

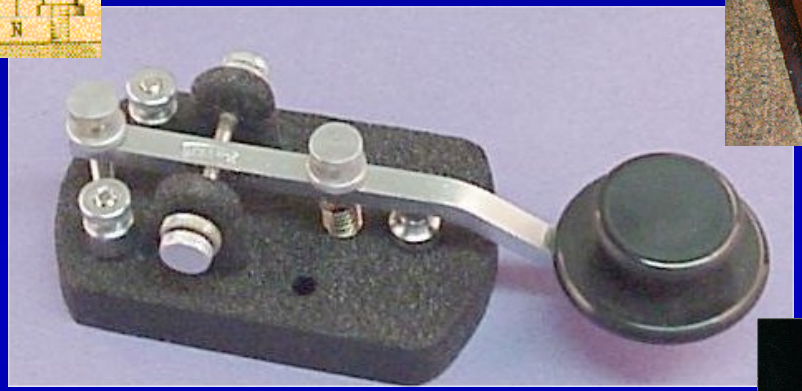
ARIE VAN PRAAG — *CERN IT PDP* — *E-Mail: a.van.praag@cern.ch*

High Performance Networking

From the Original drawing to Equipment

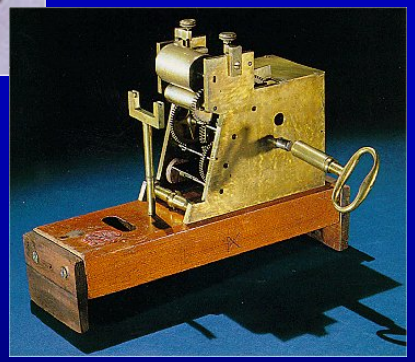


The original drawing
from Samuel Morse



A Sounder

and a Printer



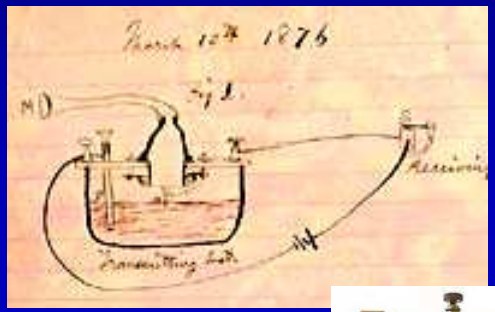
1870

Pulling the cables for the first WAN

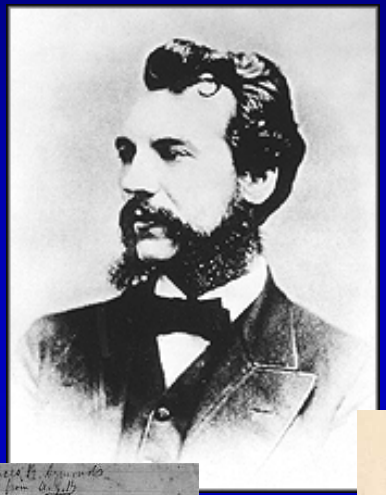
ARIE VAN PRAAG — *CERN IT PDP* — *E-Mail: a.van.praag@cern.ch*

High Performance Networking

Inventing Speech Transmission over a Wire Pair

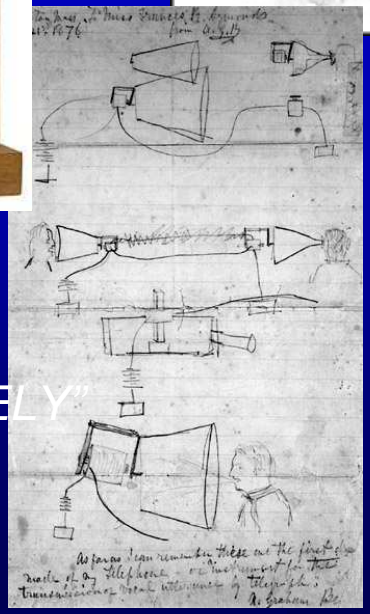


1876



Graham Alexander Bell

But some French people clamed they ware earlier



WATSON COME HERE IMMEDIATELY"

ARIE VAN PRAAG — *CERN IT PDP* — *E-Mail: a.van.praag@cern.ch*

High Performance Networking

The Telephone:

It is a Speech handling Media, not a Data Network.

Well is it ?

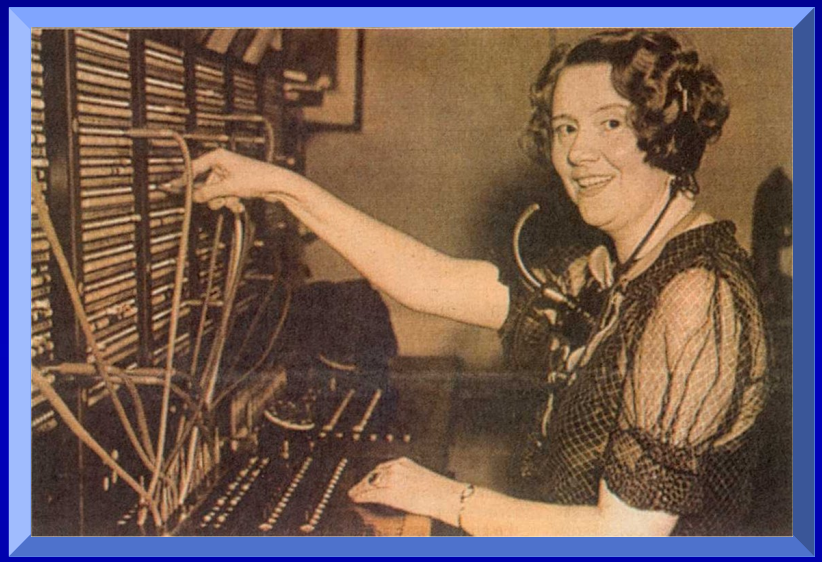
1876

1960

ARIE VAN PRAAG — CERN IT PDP — E-Mail: a.van.praag@cern.ch

High Performance Networking

Interconnect technology



1936

This kind of manual telephone connections was still seen in 1996 in a village in Bulgaria

NOT ALLOWED TO MAKE PICTURES
MILITAIRY OBJECT

With Electromechanical Standards a fully automatic network for voice data was born

ARIE VAN PRAAG — *CERN IT PDP* — *E-Mail: a.van.praag@cern.ch*

High Performance Networking

Using automatic connections

WW II

A Flexowriter modified with a relay banc and a Serial Interconnect brings a new communication standard:

RS 232

Flexowriter 10 Byte/s

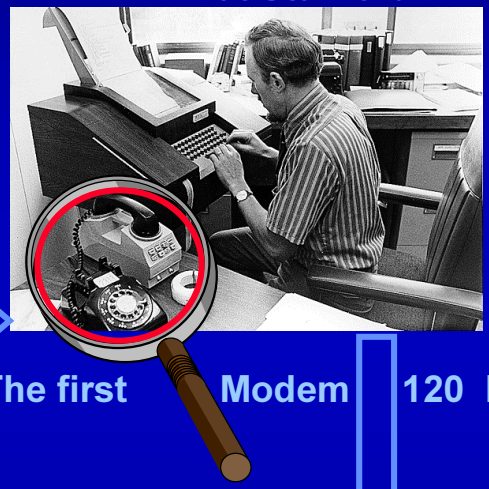
The Flexowriter interconnect made a standard character-set necessary:

ASCII

1962

Teletype 30 Byte/s

1971 at Stanford



The first Modem 120 Byte/s

The first commercial Modem 120 Byte/s

ARIE VAN PRAAG — CERN IT PDP — E-Mail: a.van.praag@cern.ch

High Performance Networking

ARPANET

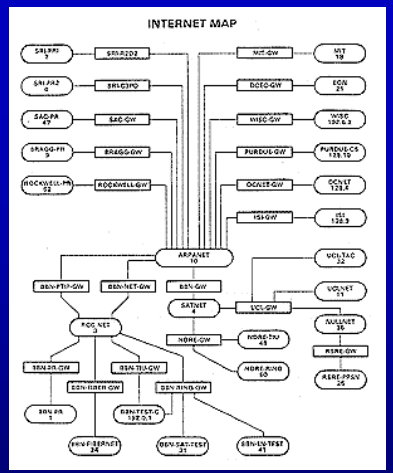
1966 Start of ARPANET in the USA.

Larry Roberts
Designs and oversees ARPAnet, which evolves into the Internet

Robert Taylor
starts ARPAnet project; organizes computer group at Xerox PARC

ARPAnet first connection		1969	
	connected	in 1971	13 machines
	connected	in 1977	60 machines
	connected	in 1980	10 000 machines

Initial speed	2.4 Kbit/s
Incremented later to	50 Kbit/s



ARIE VAN PRAAG — CERN IT PDP — E-Mail: a.van.praag@cern.ch

High Performance Networking

IMP's and Protocols

RFC 0001

Host Software. S. Crocker. Apr-07-1969. (Format: TXT=21088 bytes)

Information is transmitted from HOST to HOST in bundles called messages. A message is any stream of not more than 8080 bits, together with its header. The header is 16 bits and contains the following information:

Destination	5 bits	(Nodes)
Link	8 bits	(Final Addresses = People)
Trace	1 bit	
Spare	2 bits	

The destination is the numerical code for the HOST to which the message should be sent. The trace bit signals the IMPs to record status information about the message and send the information back to the NMC (Network Measurement Center, i.e., UCLA). The spare bits are unused.

IMP
the first message communication engine

The first connection was between UCLA to Stanford University in 1969

RFC = Request For Comments.

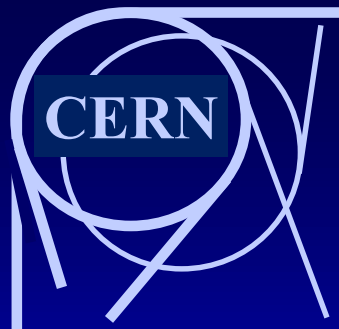
All the Internet conventions are defined in a continuously growing quantity of RFCs.

There are now to day over 3000

RFCs are archived and maintained at the Information Sciences Institute of the University of Southern California (USC ISI) but also available from many other places

http://www.isi.edu/

ARIE VAN PRAAG — CERN IT PDP — E-Mail: a.van.praag@cern.ch



High Performance Networking

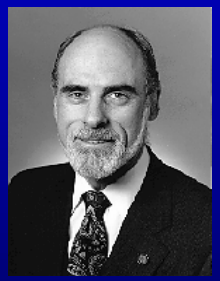
What's Brings ARPAnet 1973

Protocols:

NCP — Network Control Protocol

IP — Interface Protocol

Bob Kahn



Vinton Cerf

NCP >> TCP

TCP

TCP/IP

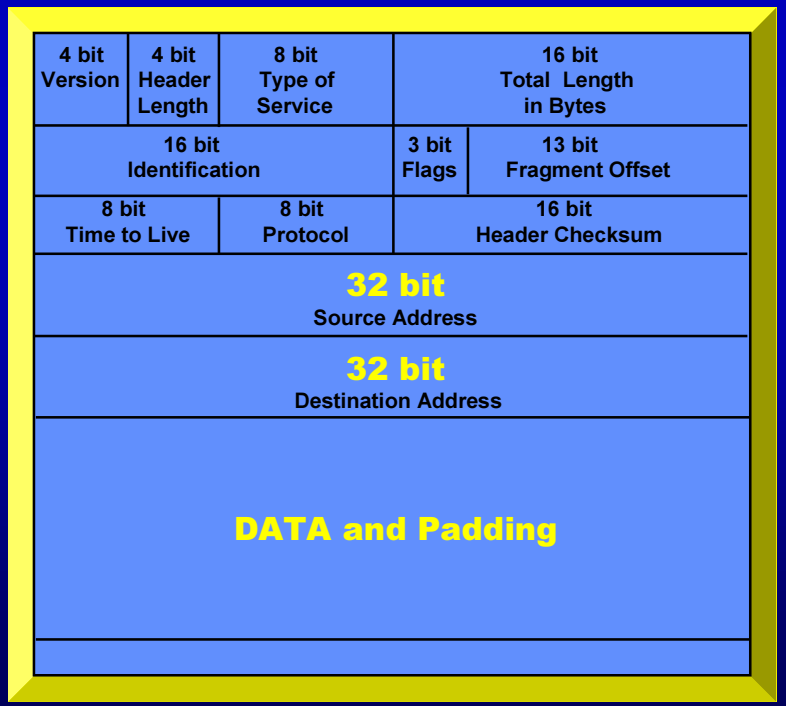
ARIE VAN PRAAG — CERN IT PDP — E-Mail: a.van.praag@cern.ch

High Performance Networking

IPv4 = Internet Protocol

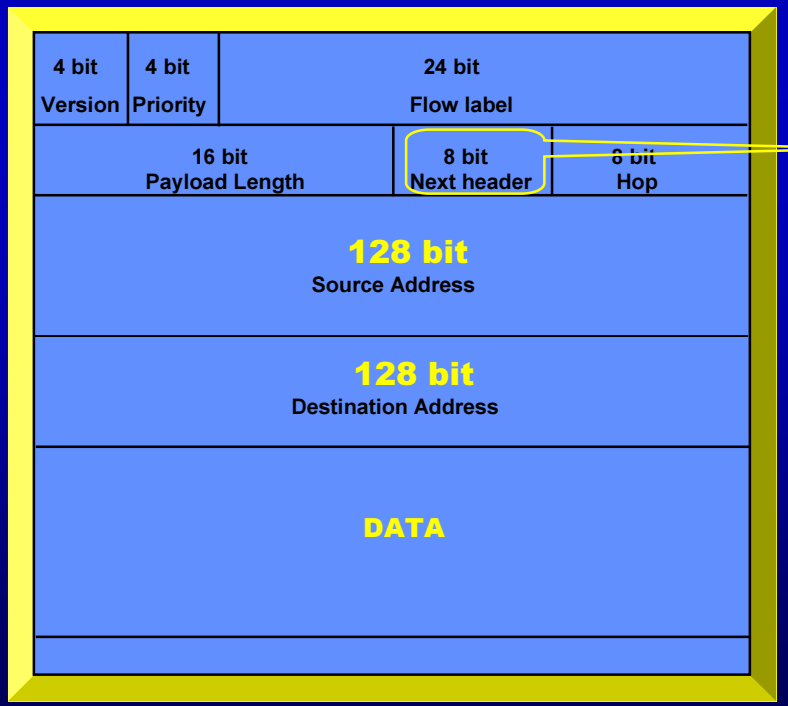
Remember RFC 0001: Destination **5 bits** = 64 nodes
Link **8 bits** = 128 nodes
Total of **4048** address Space

IPv4 Header



Total of **4 200 000 000** Address space

IPv6 According RFC 1883



Total of **3.4 e 38** Address space

340 000 000 000 000 000 000 000 000 000 000 000

ARIE VAN PRAAG — CERN IT PDP — E-Mail: a.van.praag@cern.ch

High Performance Networking

IPv6 Next Headers

IP Next Header Field definitions:

0	Hop by Hop Options	May accelerate caged operating systems
4	IP	IP Traffic
6	TCP	TCP Frame
17	UDP	UDP Frame
43	Routing	Specific Routing Instructions
44	Fragment	Partial Transfer
45	Interdomaine Routine	
46	Resource Reservation	
50	Encapsulation Security	
51	Authentication	Security and Encryption keys
58	ICMP	For Network Test Purposes
59	No next Header	
60	Destination Options	

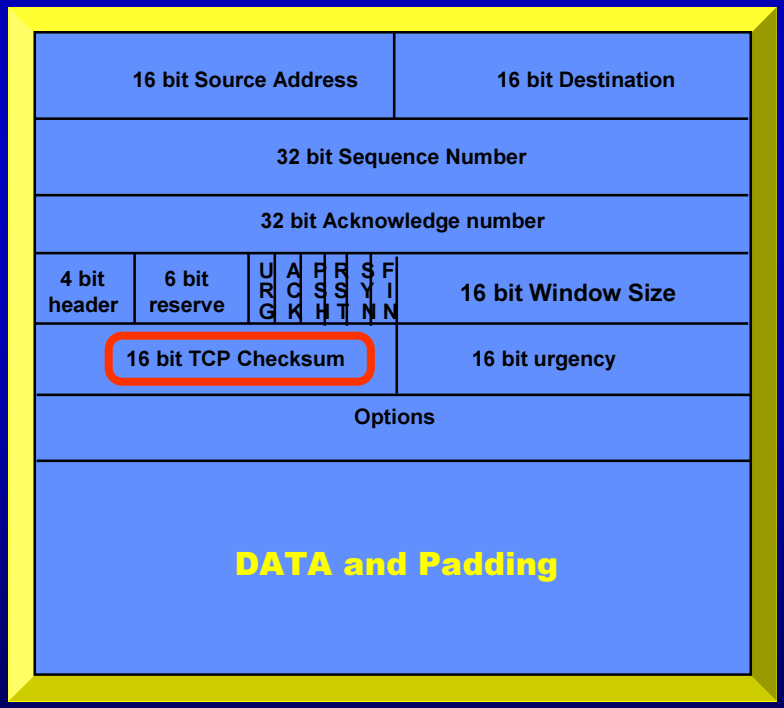
Example of concatenated headers.

ARIE VAN PRAAG — CERN IT PDP — E-Mail: a.van.praag@cern.ch

High Performance Networking

TCP Protocol

TCP Header



A problem for High Performance Networking in the TCP header is the Cheksum.

It has to be generated top to bottom on the data and be inserted upfront in the header.

This was normally done as a separate algorithm inside the TCP Stack before moving the data to the network.

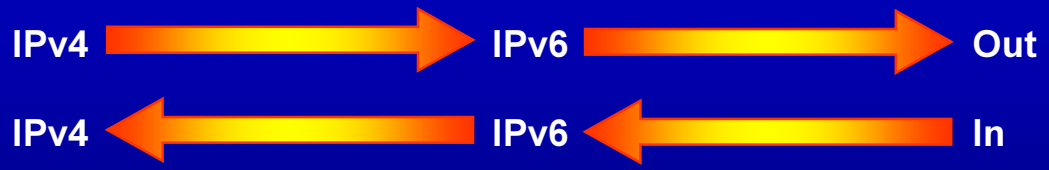
Solutions will be given with Gigabit Ethernet and with the Gigabyte System Network, GSN.

ARIE VAN PRAAG — CERN IT PDP — E-Mail: a.van.praag@cern.ch

High Performance Networking

IPv4 versus IPv6

- As seen IPv4 and IPv6 are different on many points and not compatible.
- Very difficult to mix this two protocols on one network, which means that networks should be separated ?
- What is needed during the conversion period are routers placed on the border line able to recognize the IP type and if necessary complete header to what is used in the local network.



- **This promises a difficult time until the conversion is complete.**

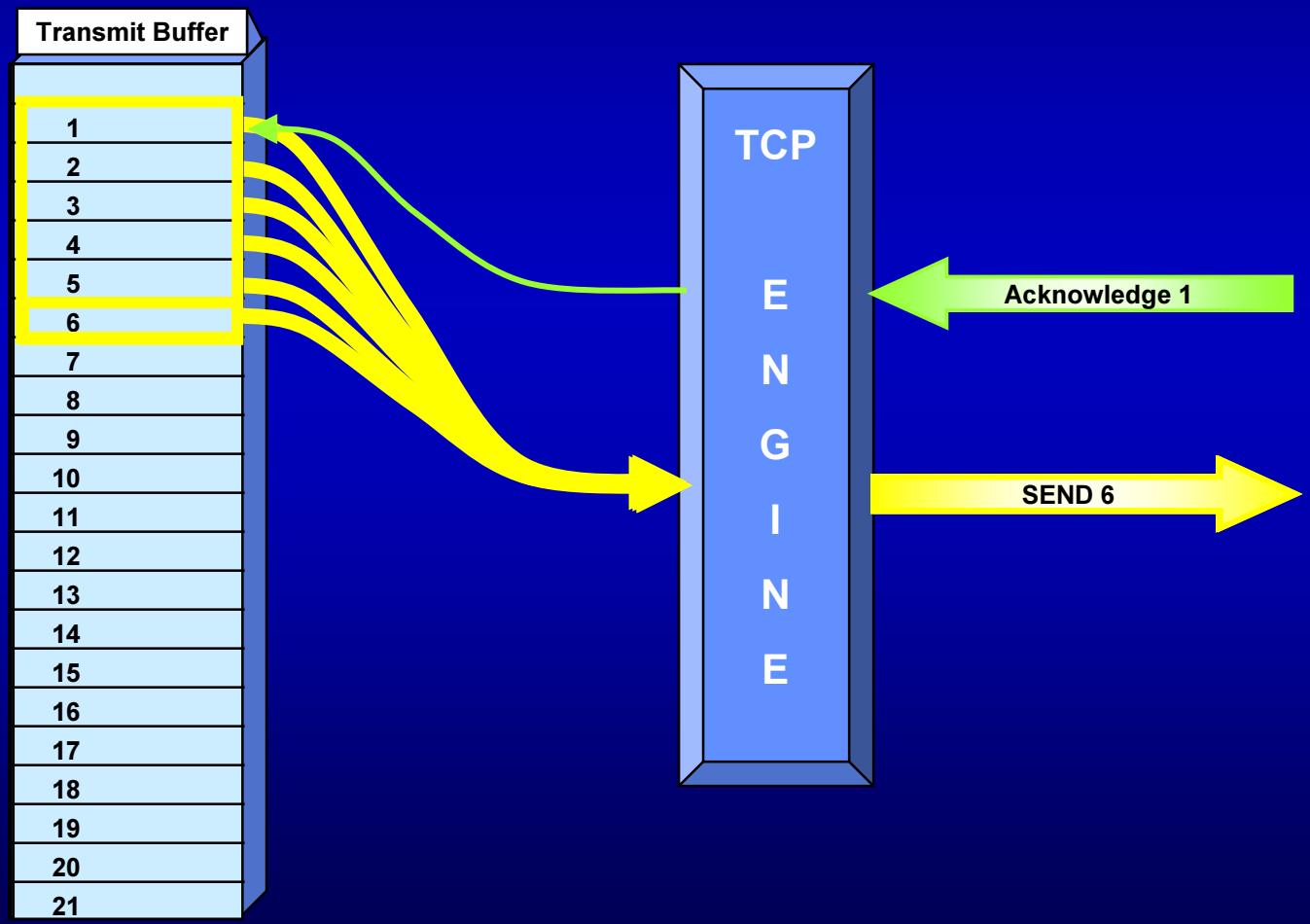
Solutions:

- Encapsulate IPv6 frames in IPv4 on IPv4 networks. (tunneling)
- Concatenate the IPv4 32 bit addresses with a known standardized prefix to run on IPv6 networks (2002::/16)

ARIE VAN PRAAG — CERN IT PDP — E-Mail: a.van.praag@cern.ch

High Performance Networking

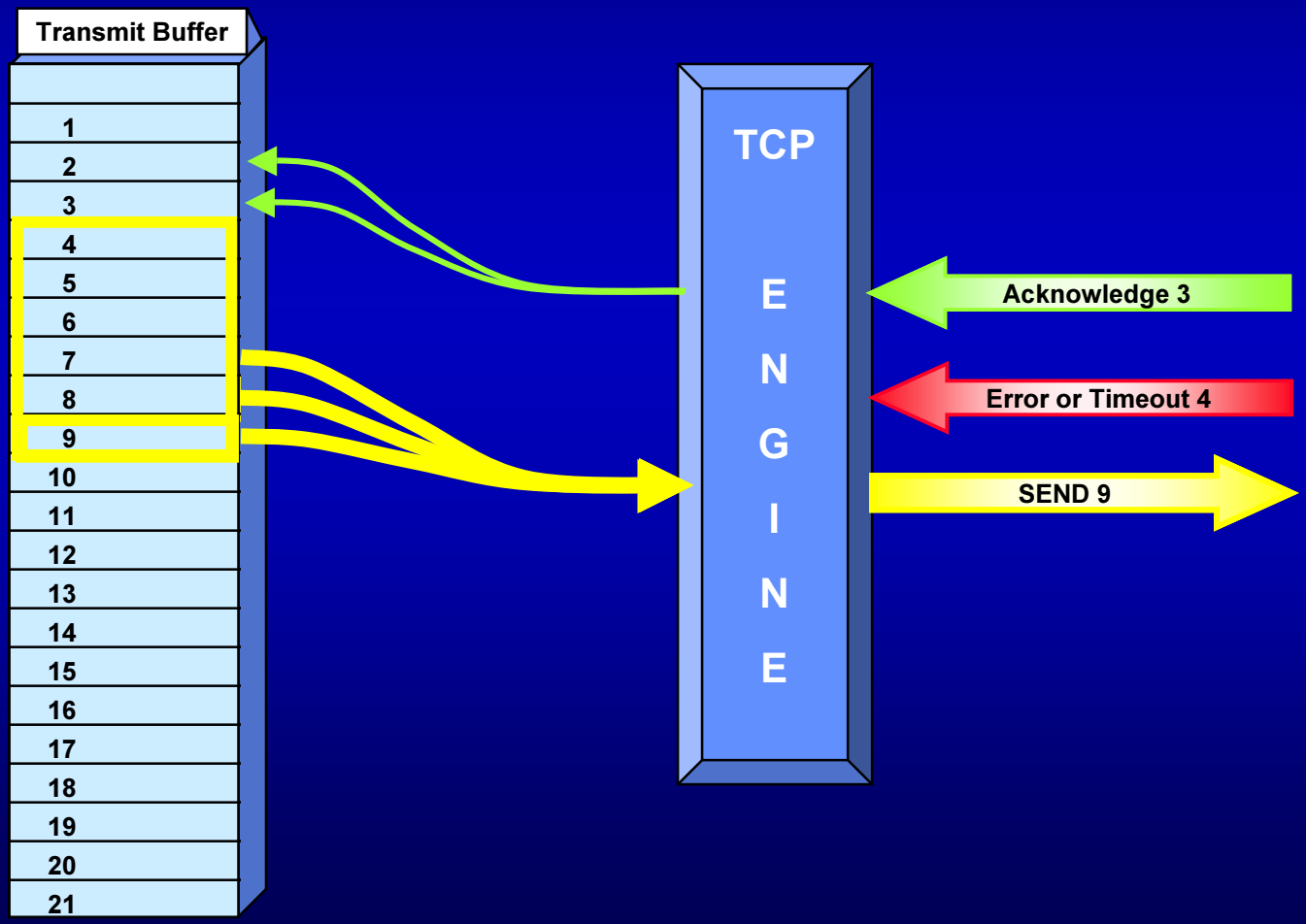
TCP = Transfer Control Protocol



ARIE VAN PRAAG — CERN IT PDP — E-Mail: a.van.praag@cern.ch

High Performance Networking

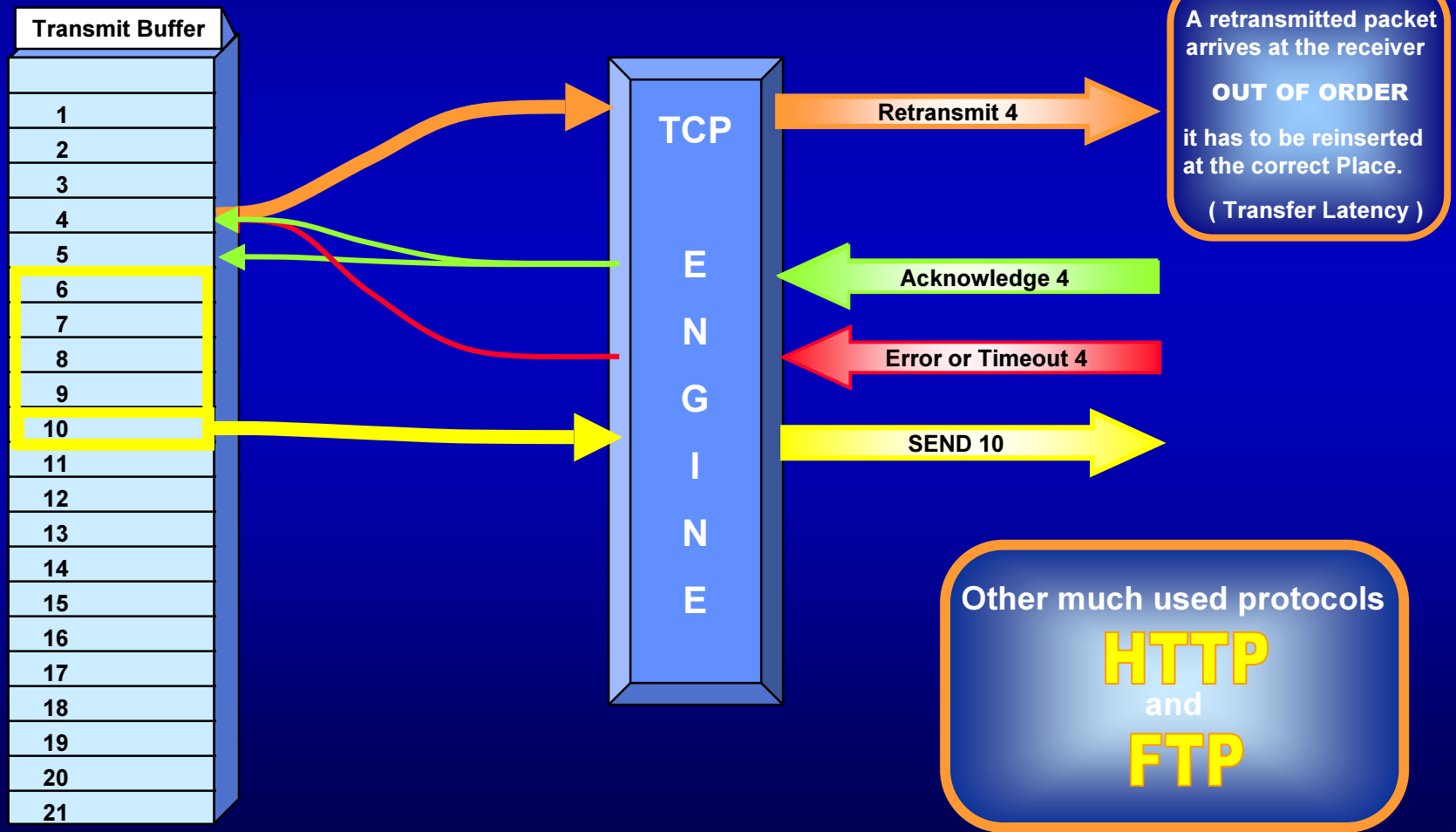
TCP = Transfer Control Protocol



ARIE VAN PRAAG — CERN IT PDP — E-Mail: a.van.praag@cern.ch

High Performance Networking

TCP = Transfer Control Protocol

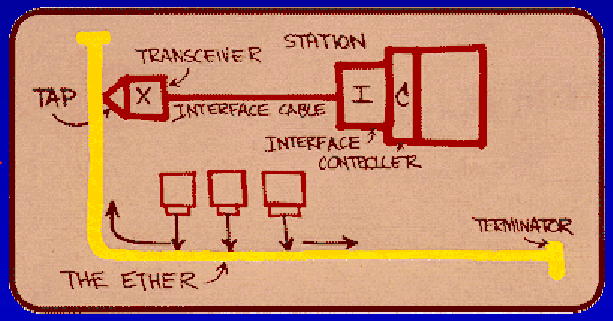


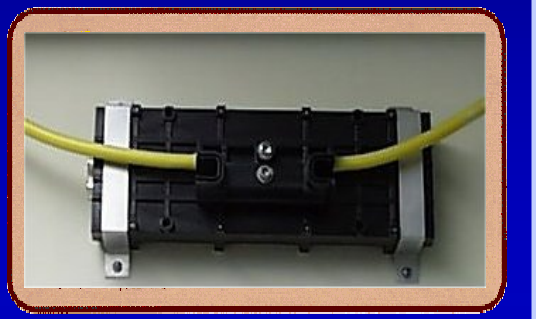
ARIE VAN PRAAG — CERN IT PDP — E-Mail: a.van.praag@cern.ch

High Performance Networking
Ethernet

Bob Metcalfe

His Ethernet Idea

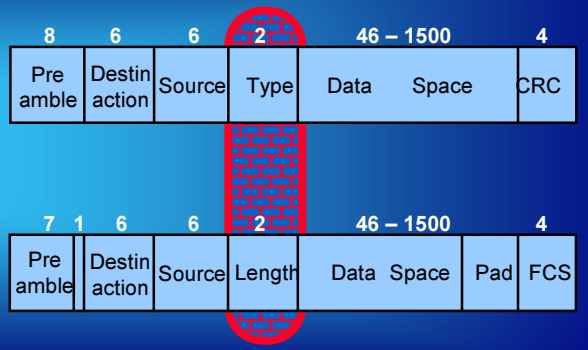




This developments leads finally to:

Industry — 1972 - Digital (DEC) & XEROX → DIX

IEEE — 1980 – IEEE standard 802.3 → 802.3



ARIE VAN PRAAG — CERN IT PDP — E-Mail: a.van.praag@cern.ch

High Performance Networking

Network Development In Europe (at CERN)

1971 A PDP 11 in the Central Library is coupled to the CDC6600 in the Central Computer center using "Gandalf" the terminal distributing system. 9600 Bit/s Distance 2 Km.

1973 Start of CERNnet with a 1 Mbit/s Link between the computer center and experiments 2 Km away.
Protocols: CERN changed progressively during 1980's to TCP/IP

1985 HEPnet in Europe
Developed to connect CERN computers to a number of Physics Institutes.

1986 The CHI project develops an interface for Fastbus and VAX machines to G703.3 at 155.5 Mbit/s

1987 Inside CERN 100 machines
Outside CERN 6 Institutes *(5 in Europe, 1 in USA)*

1989 CERN connects to the Internet.

1990 CERN becomes the Largest Internet site in Europe.

ARIE VAN PRAAG — CERN IT PDP — E-Mail: a.van.praag@cern.ch

High Performance Networking

SONET

1972 The first Electro optical specifications and synchronization protocol are published by an industry consortium called :
Synchronous Optical NETwork or SONET. Bandwidth 3.125 Mbit/s

1985 SONET was adapted by the ANSI standards body T1 X1 as Synchronous Fibre Optics Network for Digital communications.

1986 CCITT (now ITU) joined the movement.

Implemented	Optical Level	Europe ITU	Electrical Level	Line Rate (Mbps)	Payload (Mbps)	Overhead (Mbps)	H Equivalent
1989	OC - 1	---	STS - 1	51.840	50.112	1.728	---
1992	OC - 3	SDH1	STS - 3	155.520	150.336	5.184	STM- 1
1995	OC - 12	SDH4	STS - 12	622.080	601.344	20.736	STM- 4
1999	OC - 48	SDH16	STS - 48	2488.320	2405.376	82.944	STM-16
2002	OC-192	SDH48	STS-192	9953.280	9621.504	331.776	STM-64

ARIE VAN PRAAG — *CERN IT PDP* — *E-Mail: a.van.praag@cern.ch*

CERN

High Performance Networking

About bandwidth

Bandwidth:
Load a Lorry with 10 000 Tapes
100 G Byte each.

Move it over 500 Km

Drive time is 10 Hours

Bandwidth = 10^{15} / 10X3600 = 270 GByte/s

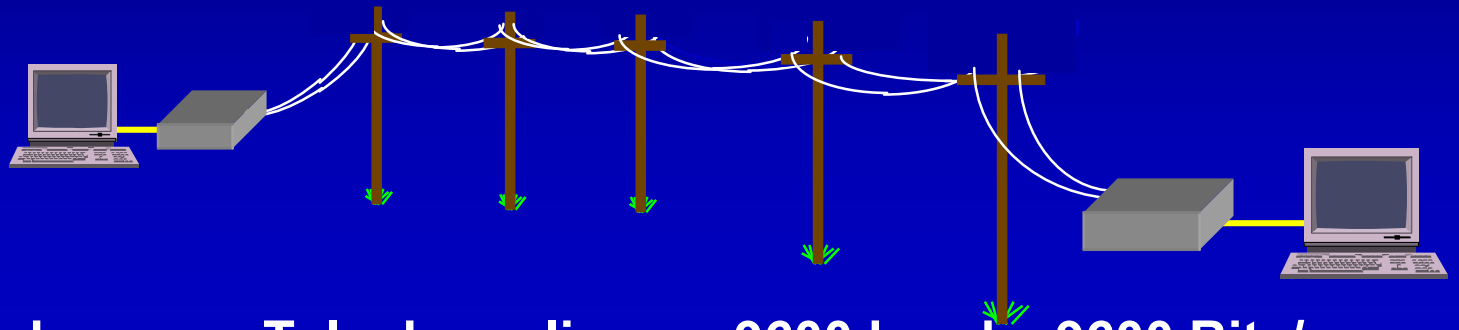
Corresponds to SONET OC 51 152

Latency 10 Hours
Latency Distance Dependent

ARIE VAN PRAAG — *CERN IT PDP* — *E-Mail: a.van.praag@cern.ch*

High Performance Networking

About Latency



Modem over Telephone lines 9600 baud = 9600 Bits/s

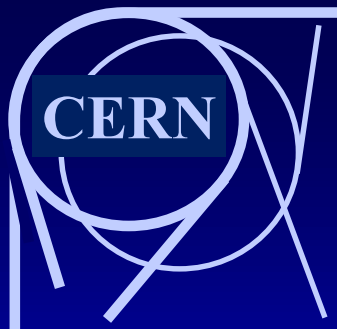
1 Byte = 8 bits >> 8 X 100 usec >> 800 u sec

A 1 MHz Clock Processor does 800 instruction in this time.

1 Peta Byte of data needs 1 10^{8} sec or 3 Years to transfer

Latency is only important as it gets large in relation to the transfer time

ARIE VAN PRAAG — CERN IT PDP — E-Mail: a.van.praag@cern.ch



High Performance Networking
Some Statements

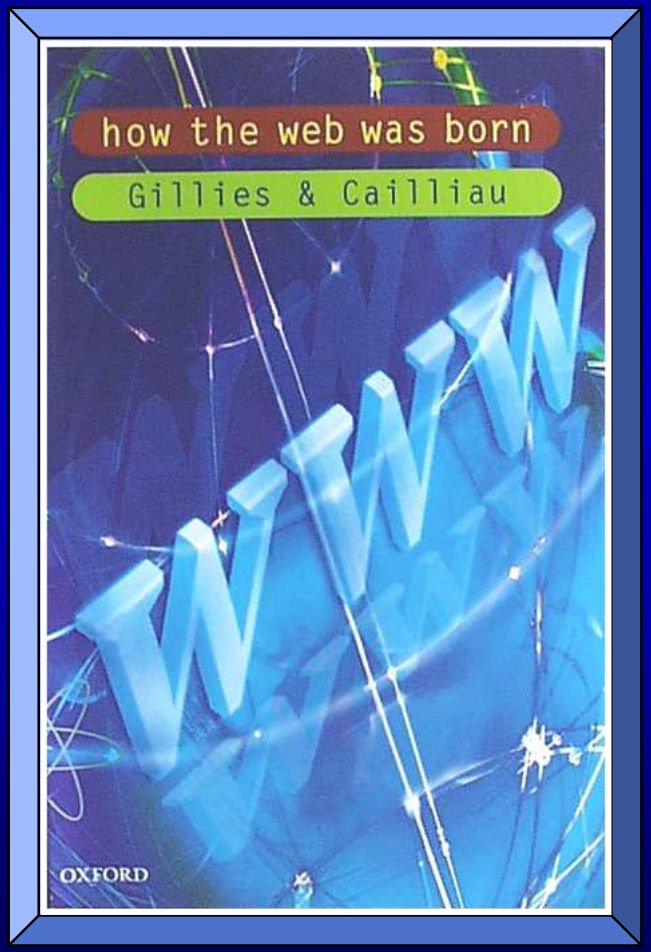
- The higher the Bandwidth the more important gets Latency.
- Flow control brings Security but also Latency.
  - Flow control is good for the LAN.
  - No flow control is better for the WAN.
- Latency is always Distance Dependent. (except satellite connections)
  - Without Flow control the Pipe has to be filled.
  - With Flow Control the distance has to be done multiple times.
- Small Frame Sizes kill Processor Efficiency.
- High Performance Networks need Operating System Bypass.
- A Network technology transparent for protocols is the better.
- A Network technology transparent for Frame Size is the better.

ARIE VAN PRAAG — *CERN IT PDP* — *E-Mail: a.van.praag@cern.ch*

CERN

High Performance Networking

HOW THE WEB WAS BORN



HOW THE WEB WAS BORN

James Gillies
Robert Cailliau

Oxford University Press
Great Clarendon street
Oxford OX2 6DP

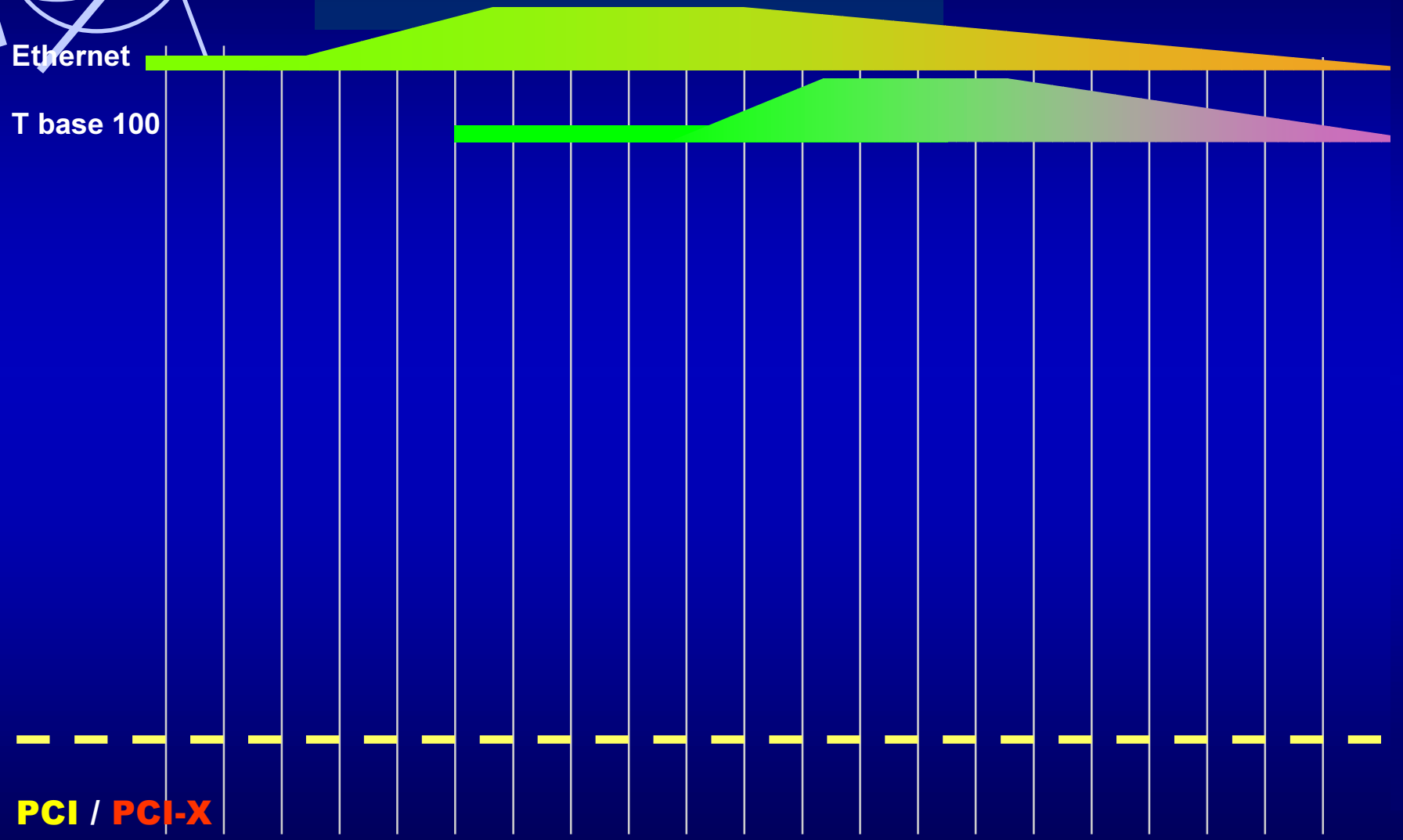
ISBN0-19-286207-3

SFr. 20.- **(at CERN)**

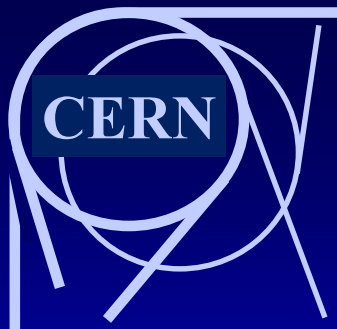
ARIE VAN PRAAG — *CERN IT PDP* — *E-Mail: a.van.praag@cern.ch*

Standards & Popularity

(made in 1995 and extended 2000)



ARIE VAN PRAAG CERN IT PDP E-Mail: a.van.praag@cern.ch



High Performance Networking References:

- How the web was born, James Gillies, Robert Cailliau, Oxford University Press, ISBN0-19-286207-3
- High Performance Networking, at CERN and Elsewhere, Arie van Praag, CERN, IT/PDP, 15 June 2001, CERN-IT-2002-002
- TCP/IP Illustrated, Vol1, W. Richard Stevens, Addison Wesley Publishing Corp, 1994, ISBN 0-201-63346-9
- TCP/IP Unleashed, K. S. Siyan, T. Parker,Sams Publisher, March 2002, ISBN 0-672-32351-6
- For the RFC Library at the University of Southern California, http://info.internet.isi.edu/in-notes/rfc/files/