A Brief Prehistory of VoIP

Stephen Casner

Adapted from a talk given jointly with
Danny Cohen
presented at Google in 2010

Both of us were at the
University of Southern California
Information Sciences Institute
when this work was performed

June 13, 2021

Danny Cohen, 1937-2019

Timeline (1/2)

1962 - Packet switching invented

1969 - The ARPAnet was born

1973 - ARPA's crazy idea: Packet Speech
- ARPA initiated the NSC program
- NVP implemented for ARPAnet

1974 - CVSD over the ARPAnet
- LPC over the ARPAnet
- TCP defined (Cerf+Kahn paper)

1975 - Voice Message System demonstrated

Timeline (2/2)

1976 - CVSD 4-site teleconf (ISI,LL,CHI,SRI)

1978 - TCP/IP split, UDP defined, PV movie

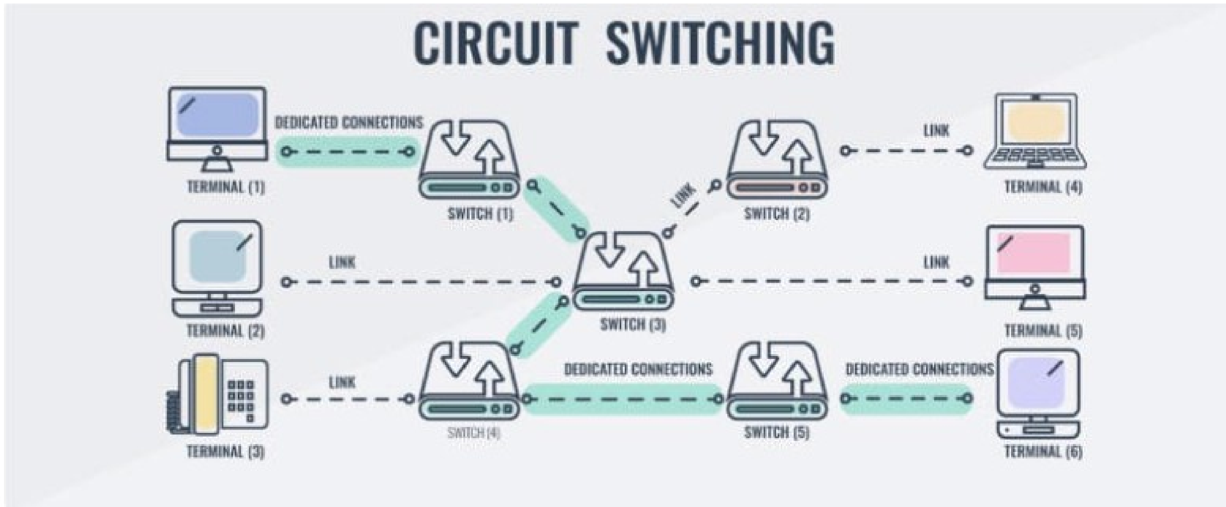
1981 - NVP-II defined for use over IP

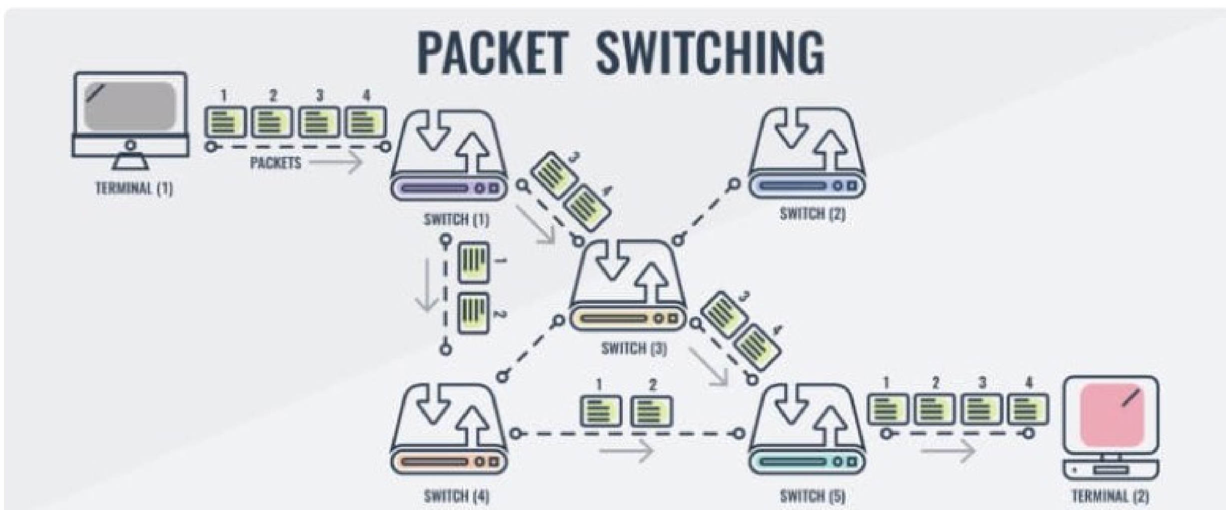
1992 - IETF AVT WG formed, first audiocast

1994 - MBone carried Hubble repair, Stones

1995 - The term “VoIP” coined
- ITU-T adopted RTP for H.323

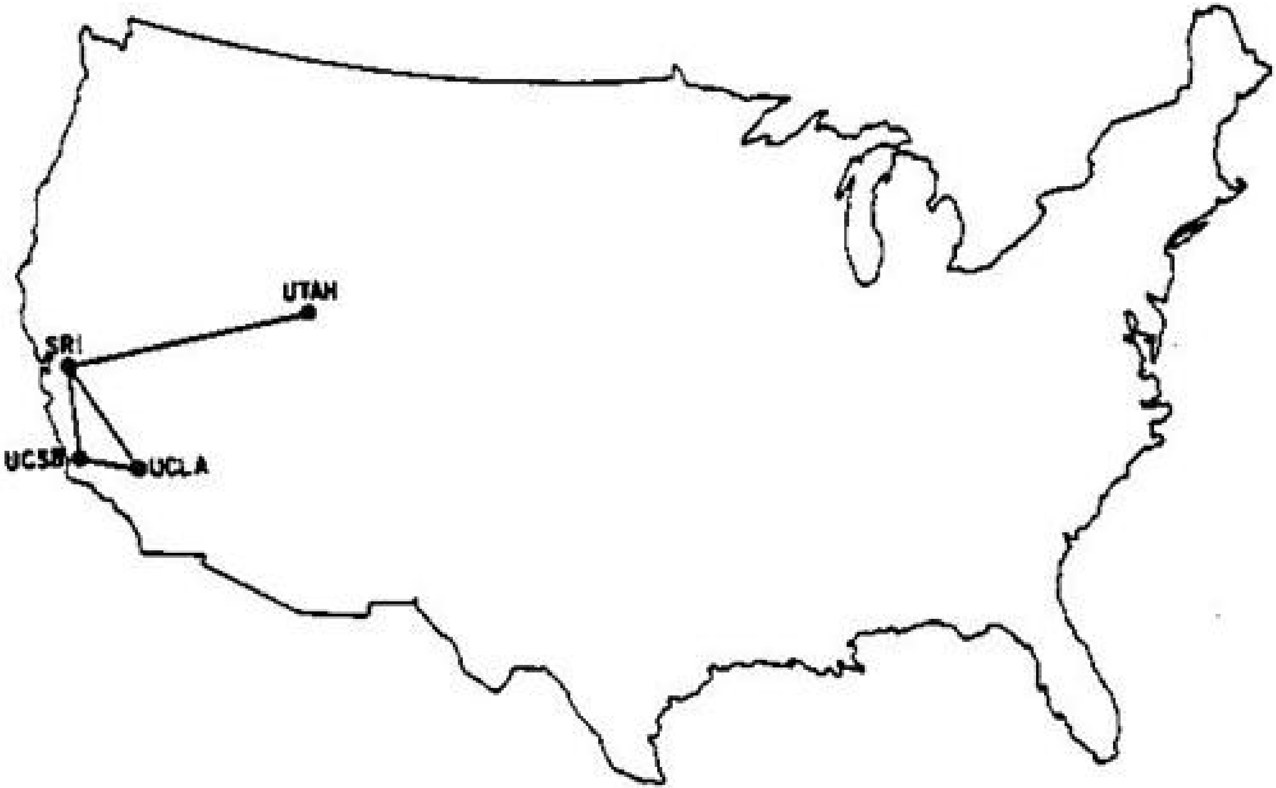
1996 - RTP specification RFC 1889 published
- IETF began working on SIP





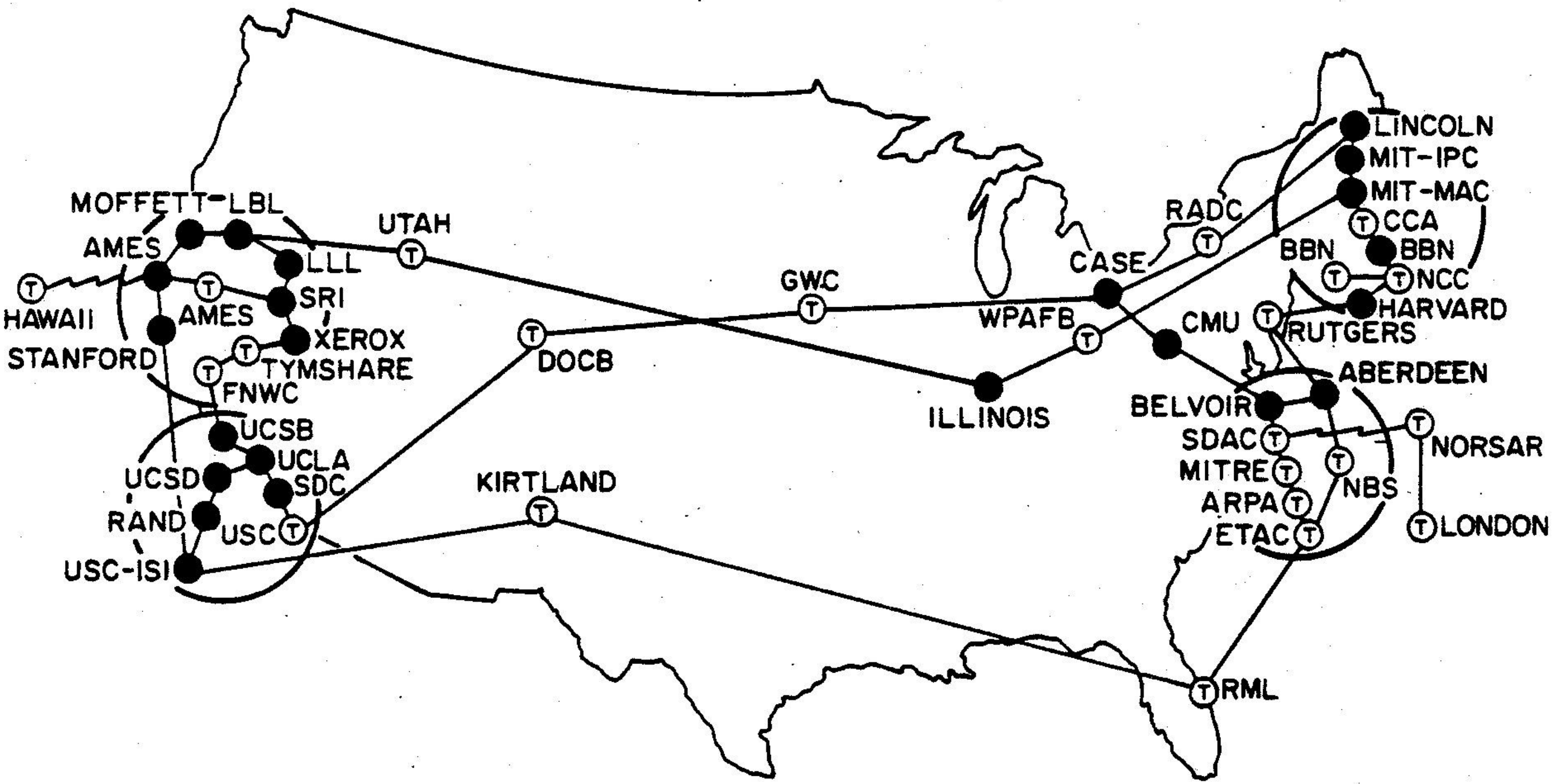
From https://www.comparitech.com/net-admin/circuit-switching-vs-packet-switching/

The ARPAnet in December'69



4 sites
4 lines of 50Kbps each

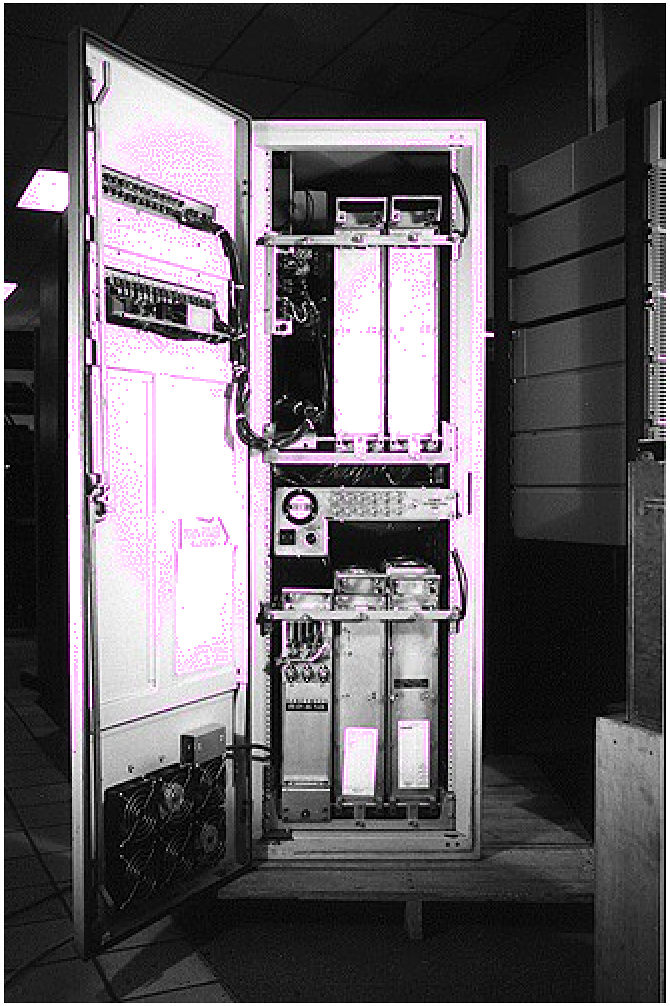
The ARPAnet in June 1974



About 45 sites

3 cross country lines of 50Kbps each

MIL-SPEC IMP of the ARPAnet



The ARPAnet

An IMP per site, up to 64 IMPs each with 4 hosts

The IMPs interconnected by Bell-303 modems bridging between a 50Kbps digital interface on one side and analog circuits on the other side.

Addresses were 8=6+2 bits (IMP + host)

IPv4 uses 32 bits, IPv6 uses 128 bits

ARPA's Crazy Idea - 1973

ARPA suggested to use packet switching networks for real-time interactive speech, for military and civilian applications.

The Network Secure Communication (NSC) program was tasked to develop packet voice

Carriers didn't recognize that this would be "disruptive technology"

The Explicit Objective of NSC

Provide Proof-of-Concept for the use of packet-switching networks for interactive telecommunication among people.

The Implicit Objectives

High voice quality (intelligibility, recognition)

Real-time (low delay, high bandwidth, ...)

Tele-conferencing

Multi-media

Voice-mail

Interoperability with the telephone network

ARPA's NSC Program

Bob Kahn was the PM in charge

Contracts: BBN,CHI,ISI,MIT,SCRL,SRI,Utah

Based on cooperation, not competition:
All succeed together or all fail together

Two-pronged program: speech compression and real-time packet communication

This presentation is mostly about the real-time packet communication part of the program.

Speech Compression ...

... necessary to fit in 50Kbps lines

- CVSD, 8Kbps, domain-independent waveform encoding, light computing
- LPC-10, 2.4Kbps, domain-dependent vocal tract model, heavy computing

LPC required about half of a 19” rack for an array processor (SPS-41 or FPS AP-120B)

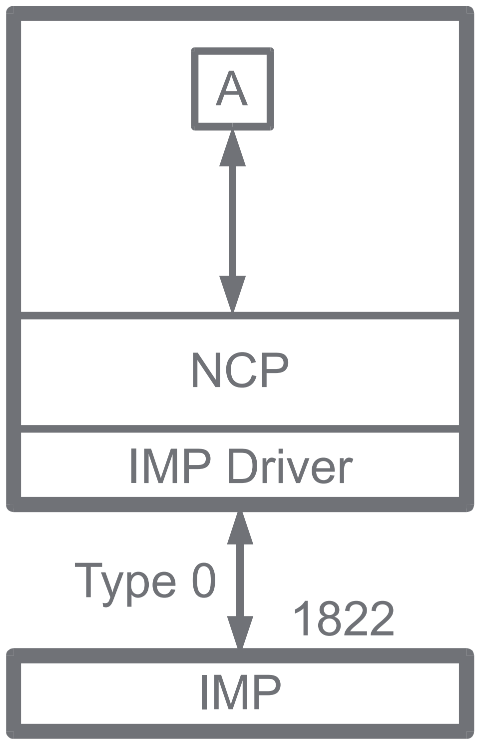
The FPS AP-120B (12MFLOPS)

FIGURE 2.36 An overall view of an FPS AP-120B installation. The AP-120B occupies only 29 inches of rack space, and is attached to a PDP 11/34 with two disc units, a tape reader and output printer. The control teletype or VDU is not shown. (Photograph courtesy of D Head and Floating Point Systems, S A Ltd.)

NCP: Network Control Protocol

- Host-to-Host protocol was NCP;
  no IP, no TCP, no UDP
- NCP: data integrity, flow control, and error recovery (timeouts+retransmissions), like today's TCP

Initial ARPAnet Protocols



NCP Type of Service

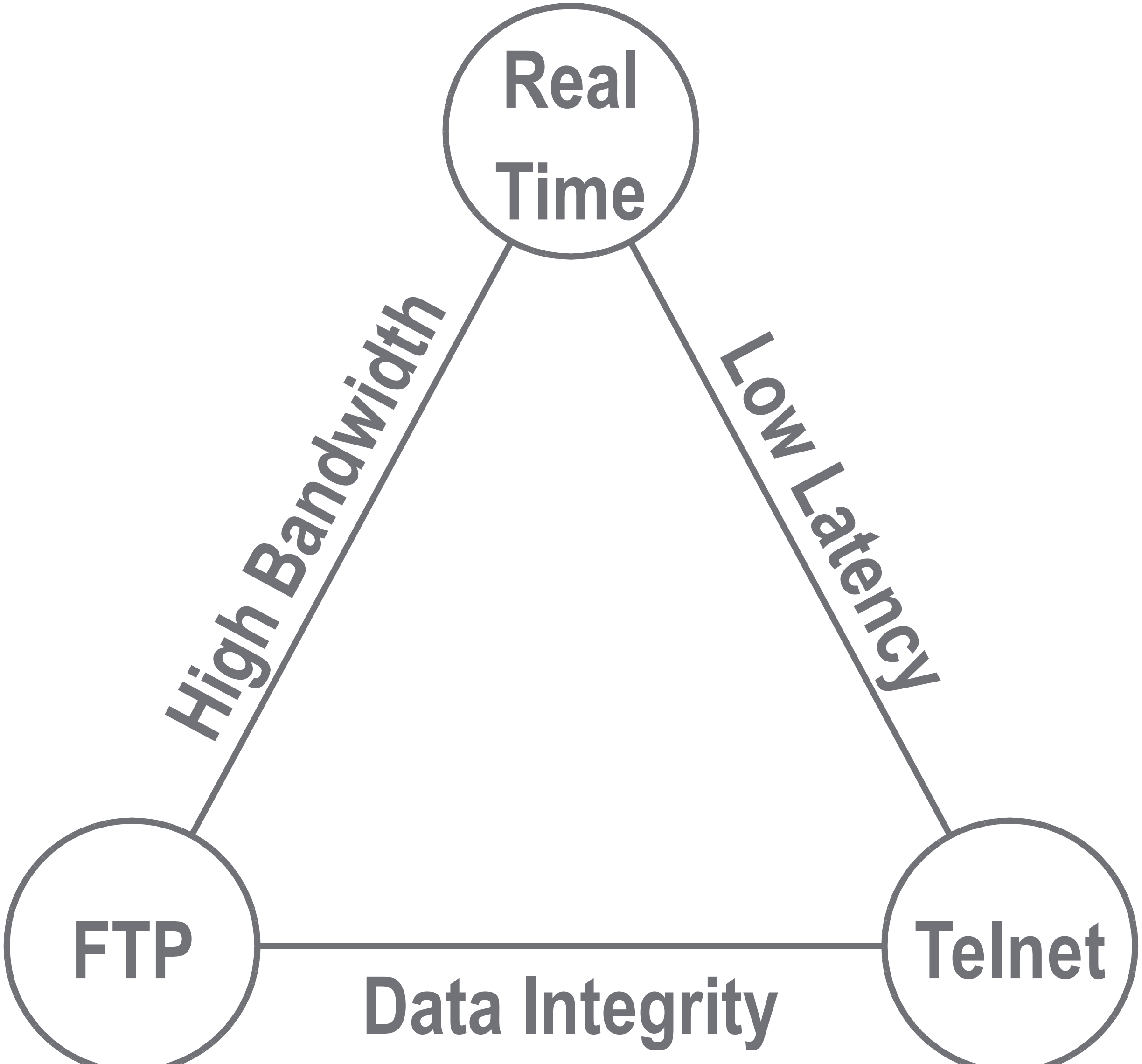
- The NCP provided what everyone wants: reliable error-free in-order delivery.
- No one wants erroneous data
- No one wants to lose data
- This was the only Type-of-Service offered

NCP Type of Service

- ~~The NCP provided what everyone wants: reliable error-free in-order delivery.~~
- ~~No one wants erroneous data~~
- ~~No one wants to lose data~~
- ~~This was the only Type-of-Service offered~~

But it's not good enough for realtime traffic

Choose 2 out of 3, No More



Realtime Communication

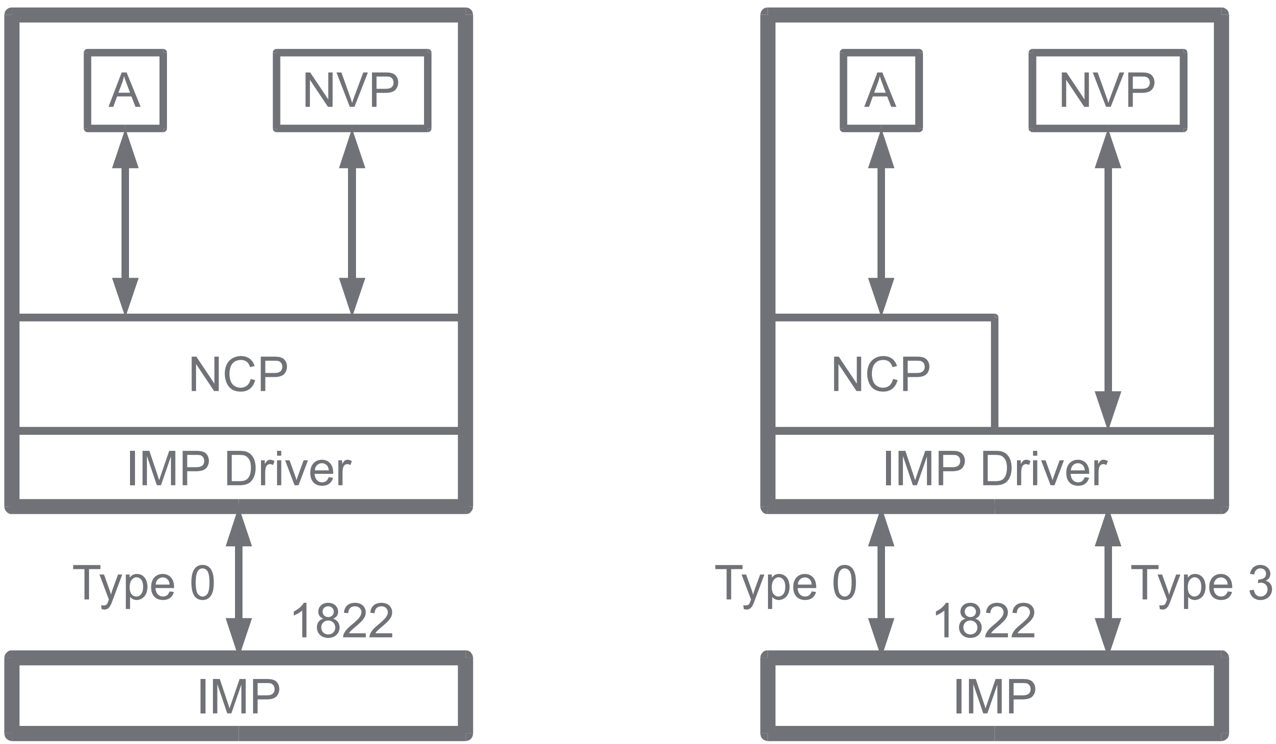
In many cases of realtime communication, new data obsoletes previous data (e.g., weather reading and stock markets).

Therefore, it does not pay to retrieve lost or damaged data if it causes large delay.

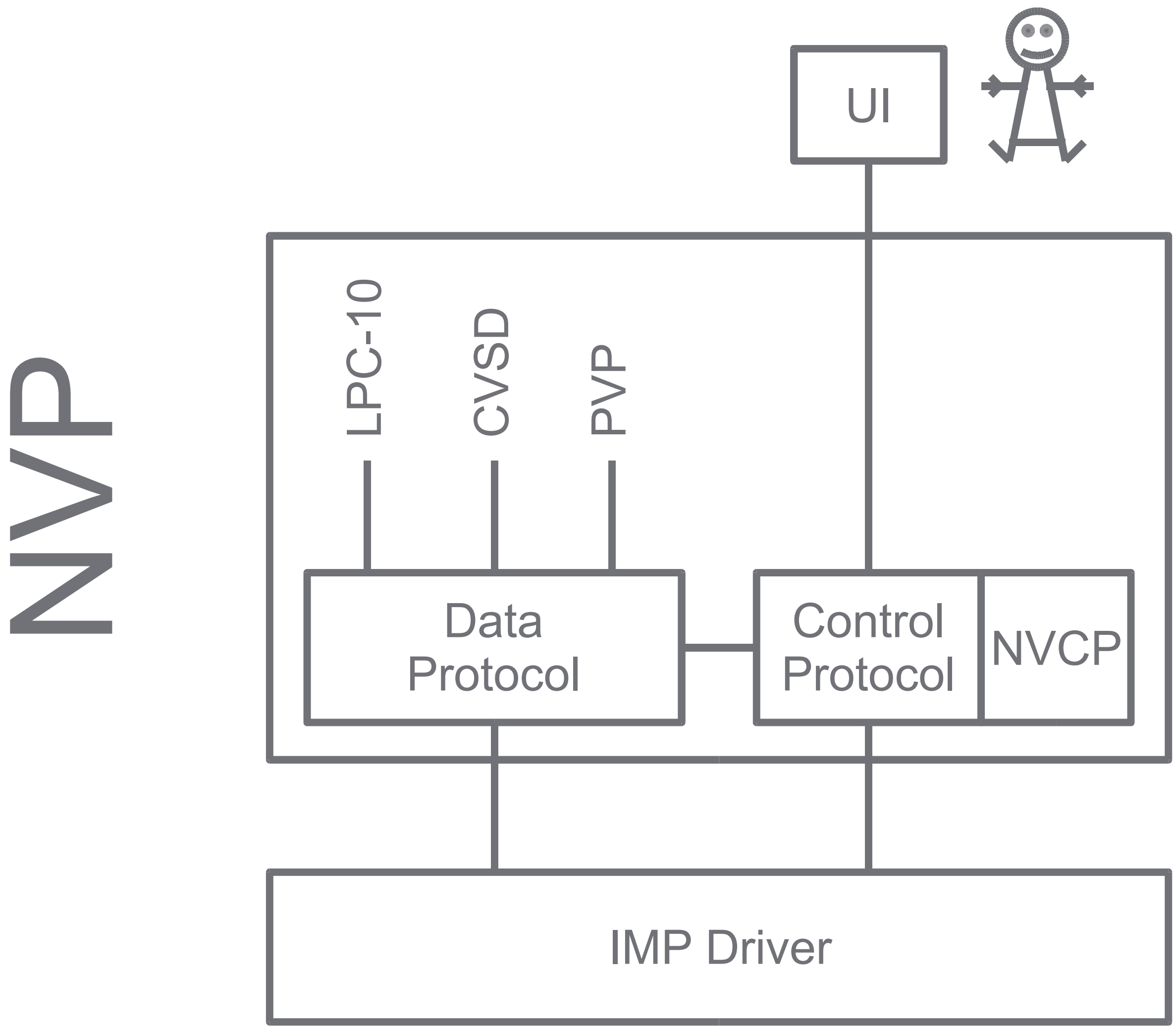
Realtime is like milk: keep the newest
Non-realtime is like wine: keep the oldest

NCP Bypass with Raw Packets

In order to have reasonable delay the NCP was bypassed using raw-packets (subtype-3)



NVP: Network Voice Protocol



NVP Data Protocol

- Simple header for sequencing, timing

NVP Control Protocol

- Connection establishment, negotiation

NVCP = NVP + Conferencing

- Provides floor-control and UI extensions
- UI extensions support voting and invitations
- Also added extensions for voice messaging

What packet size/rate is best?

SOME INITIAL MEASUREMENTS OF ARPANET PACKET VOICE TRANSMISSION

Stephen L. Casner, Eric R. Mader, E. Randolph Cole

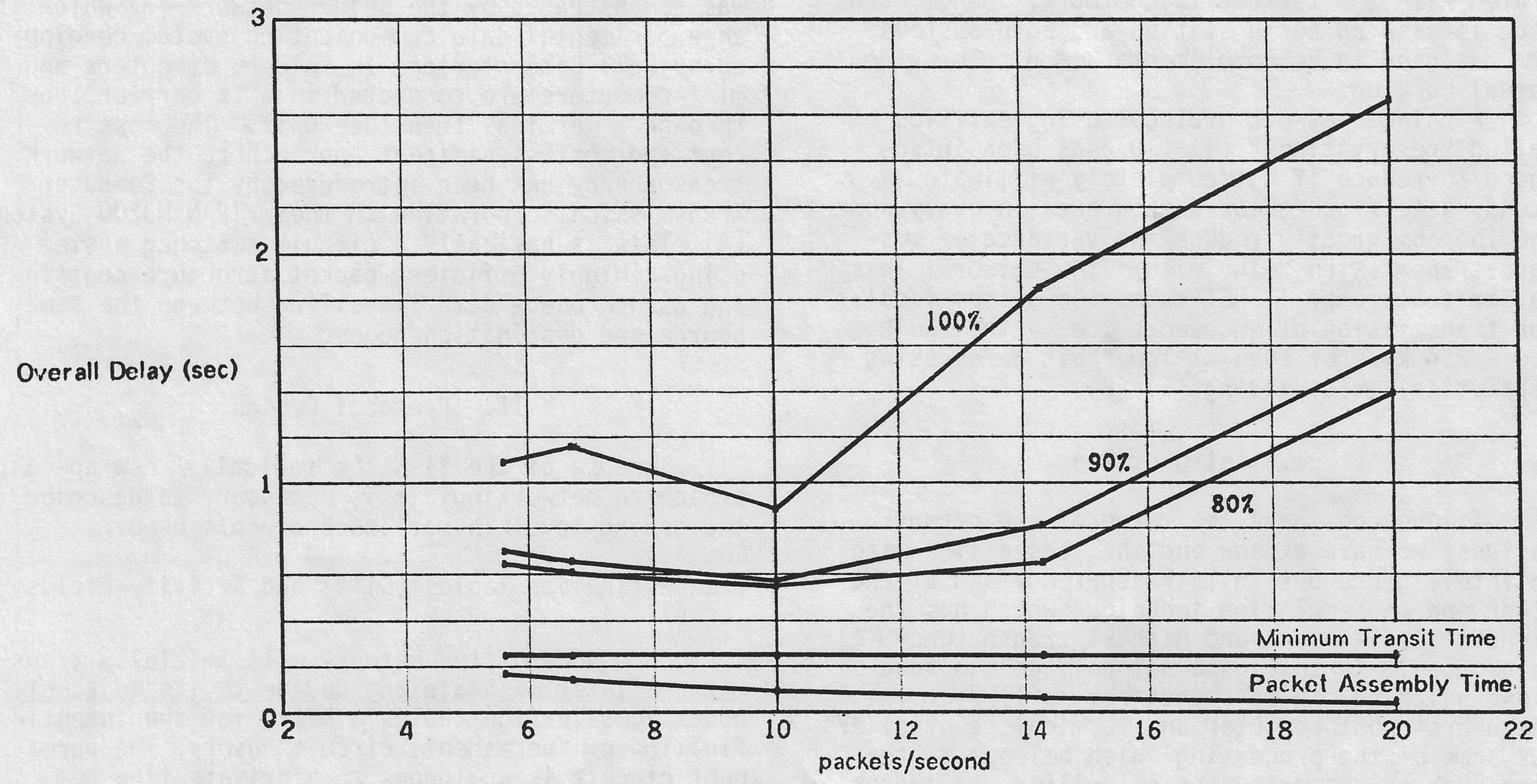
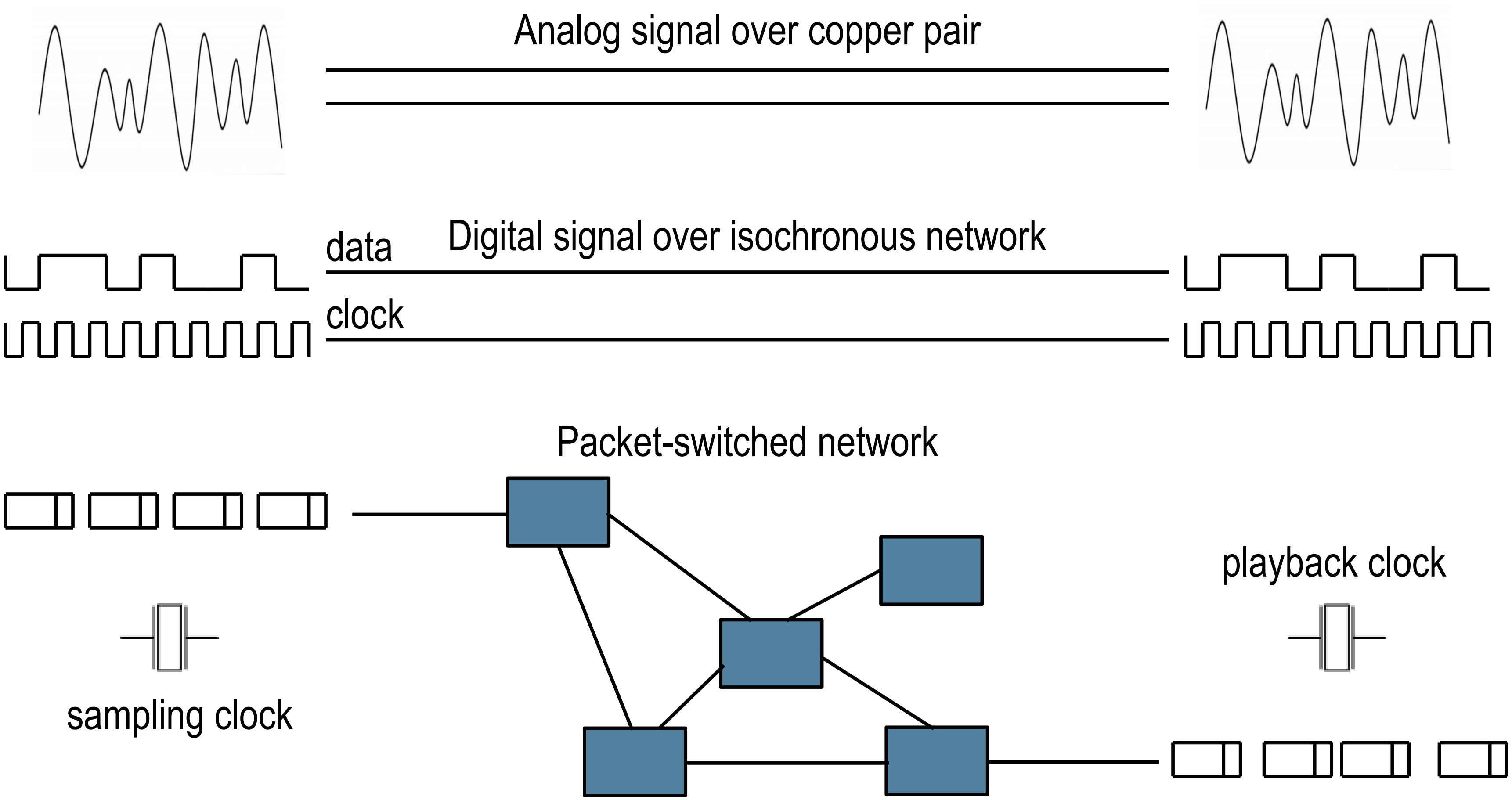


Figure 6.
Percentage of Packets Received "In Time"

From IEEE NTC 1978

Transmitter-Receiver Sync



Two Strategies For Timing

MIT Lincoln Laboratory:

- Establish playout clock based on first packet arrival plus reconstitution delay
- Accurately reproduces silence durations, but late packets must be discarded
- Adjust reconstitution delay for next talkspurt

USC Information Sciences Institute:

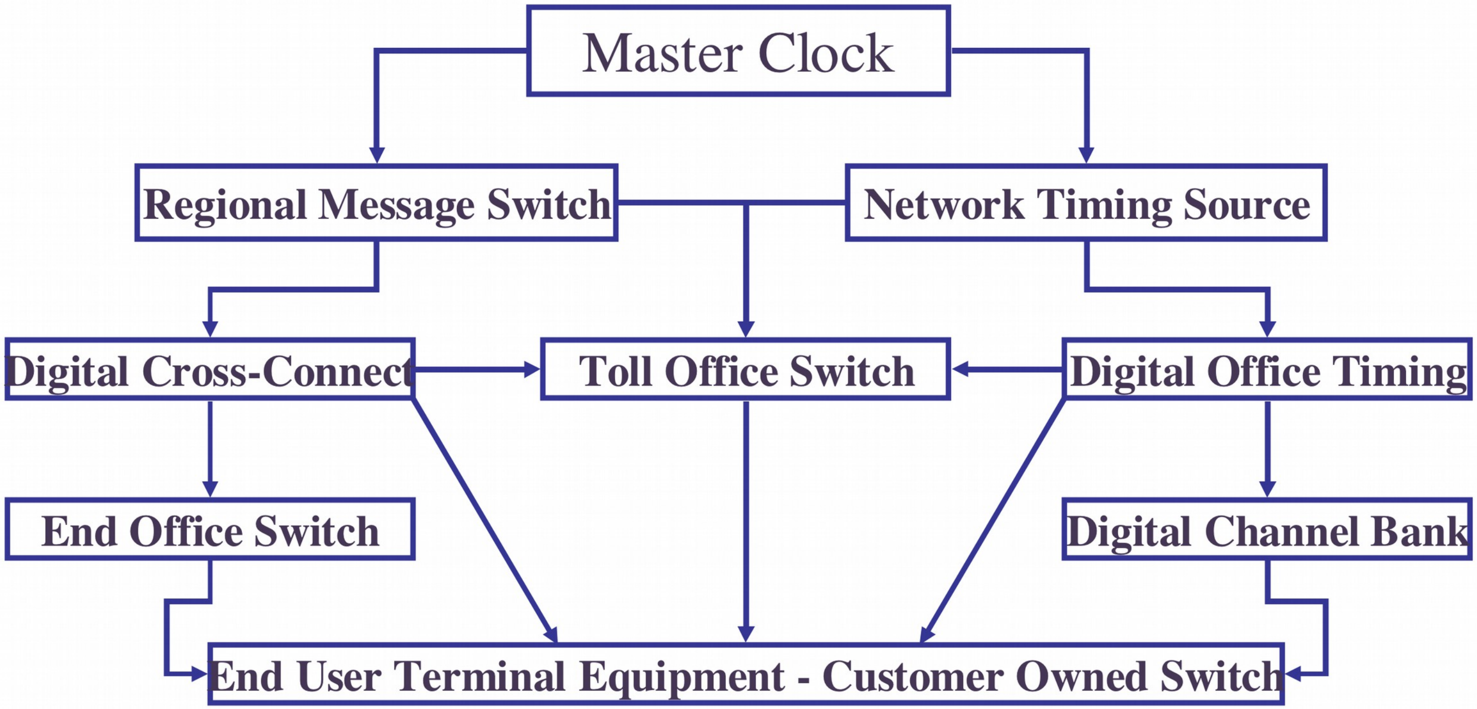
- Let playout clock slip if packets arrive late, automatically increasing reconstitution delay to play all received packets

Continuous, live broadcast?

- Playout clock too fast $\Rightarrow$ "Buffering…"
- Playout clock too slow loses live timing, needs unlimited buffer

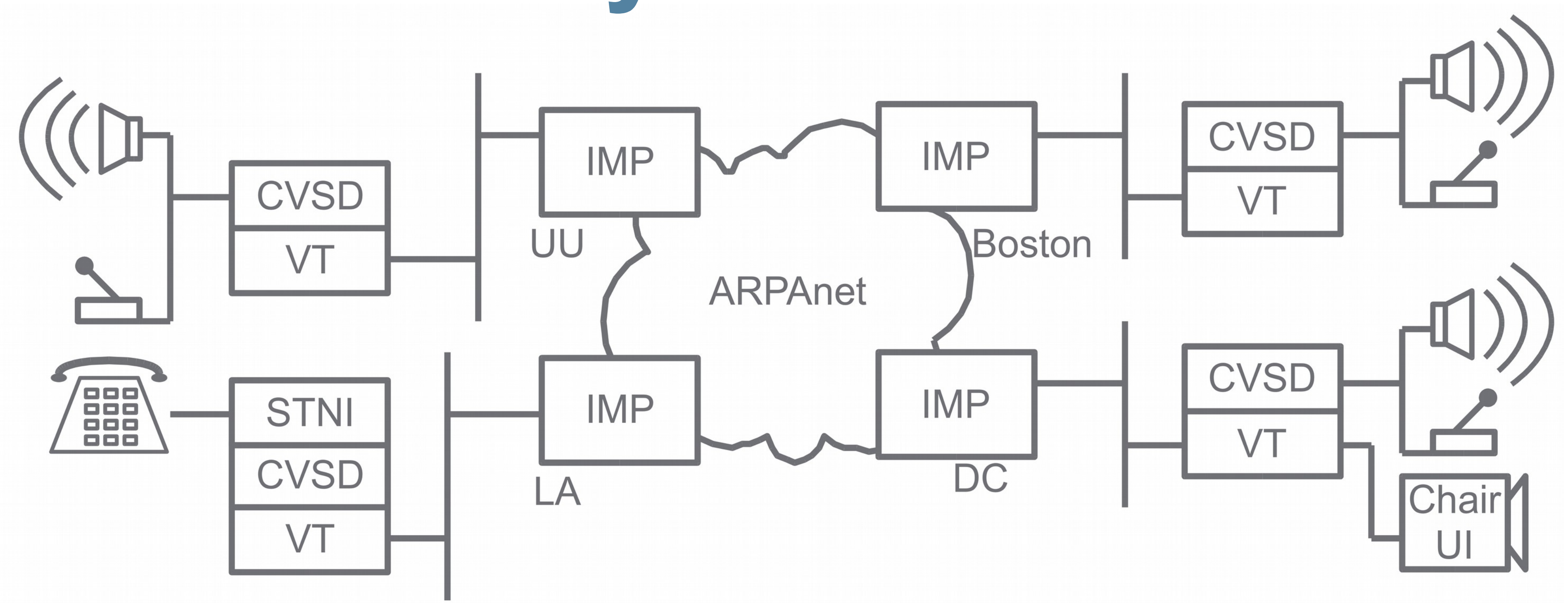
Ideally, the video frame rate and audio sample clock should be adjusted dynamically in a manner similar to the Network Time Protocol using the source timestamps and arrival times as the equivalent of the round-trip transmit and receive time pairs.

Clocking Distribution



From https://ece.utdallas.edu/~torlak/courses/ee4367/lectures/Clock.pdf

The January 1978 Movie

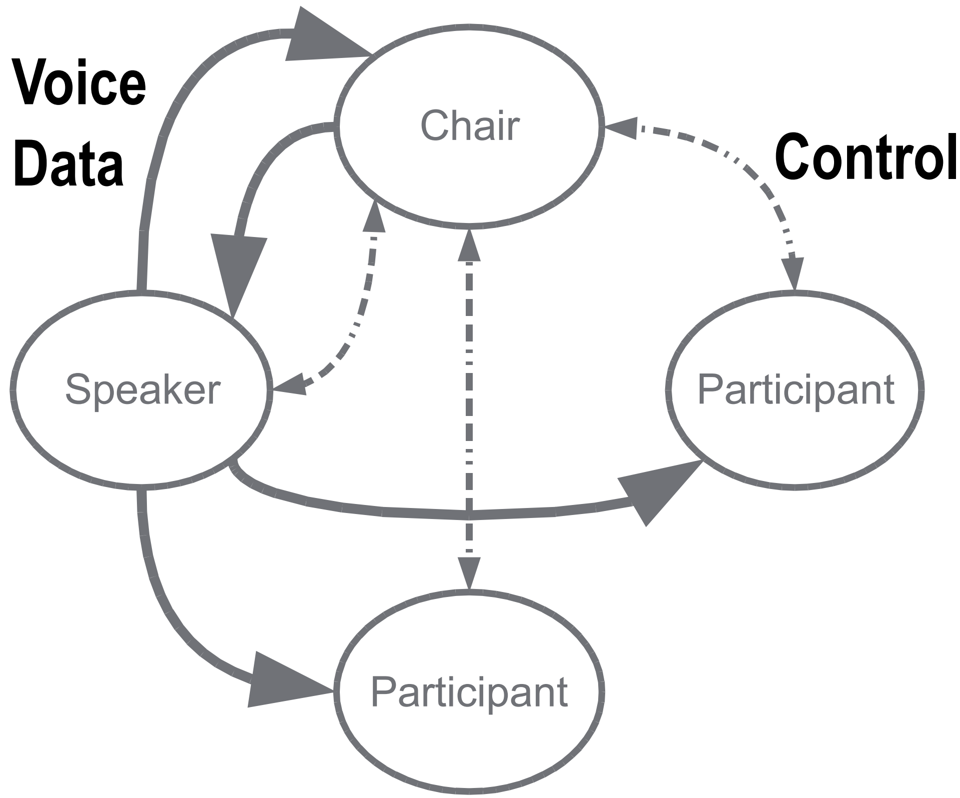


The movie shows a simulated four-node teleconference with CVSD-vocoded speech.

STNI = PSTN Interface

- Interoperability with the public switched telephone network
- Connects any external phone with a local voice terminal (VT)
- Supports DTMF signaling, both ways
- Can be called from any phone, and can call out to any phone, allowing toll bypass
- Note similarity to *SKYPE*

Floor-Controlled Data Paths



- Chairman talks only to the speaker
  - Only one decoder, so receive one stream
  - LPC can't encode multiple-speaker audio

The movie is available at:
http://www.youtube.com/watch?v=MGat1jRQ_SM

or 50MB MPEG download from
http://casners.us/dv.mpg
http://ee.stanford.edu/~gray/dv.mpg

You may have noticed...

- Occasional screeches: vocoding errors
- Danny's voice was not his own
- Movie production avoided need for lip-sync
- Various user interface devices
- Voting without switching to each participant

The Internet was Born 1976-83

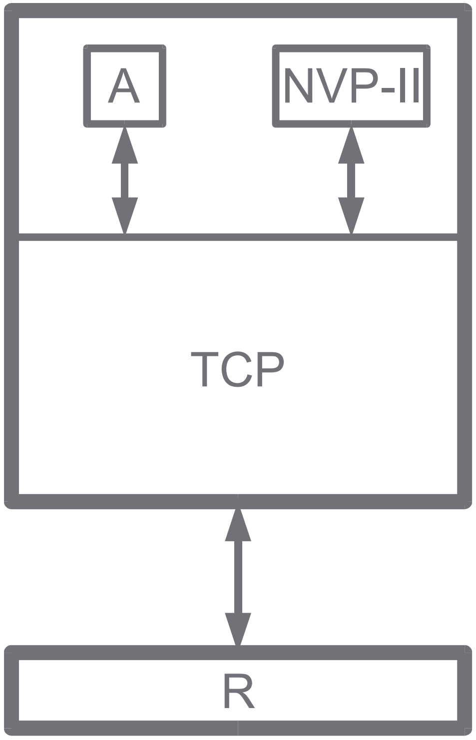
By replacing the ARPAnet's NCP with TCP, it became possible to create the Internet.

Initial Internet was ARPAnet plus many networks connected to it.

8-27-1976: TCP gateway PRnet – ARPAnet.

1-1-1983: NCP disabled on ARPAnet.

Transmission Control Protocol (TCP)



TCP Type of Service

TCP provided what everyone wants:
reliable error-free in-order delivery.

No one wants erroneous data

No one wants to lose data

This was the only Type-of-Service offered

TCP Type of Service

~~TCP provided what everyone wants: reliable error-free in-order delivery.~~

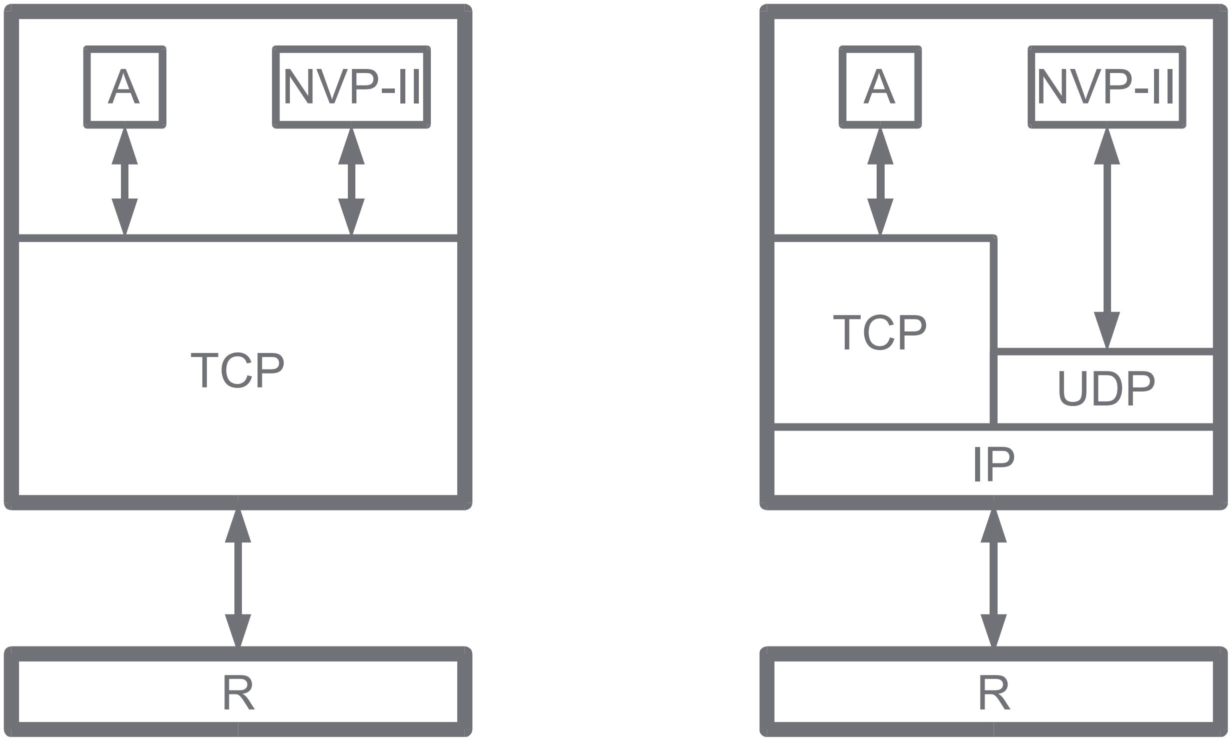
~~No one wants erroneous data~~

~~No one wants to lose data~~

~~This was the only Type-of-Service offered~~

UDP was added later as its price was right

The Split of the Original TCP



- IP = Envelope, TCP = Letter inside
- UDP was defined for IP-without-TCP
- The split is v4 of TCP (not of IP!)

IPv6? What about IPv5?

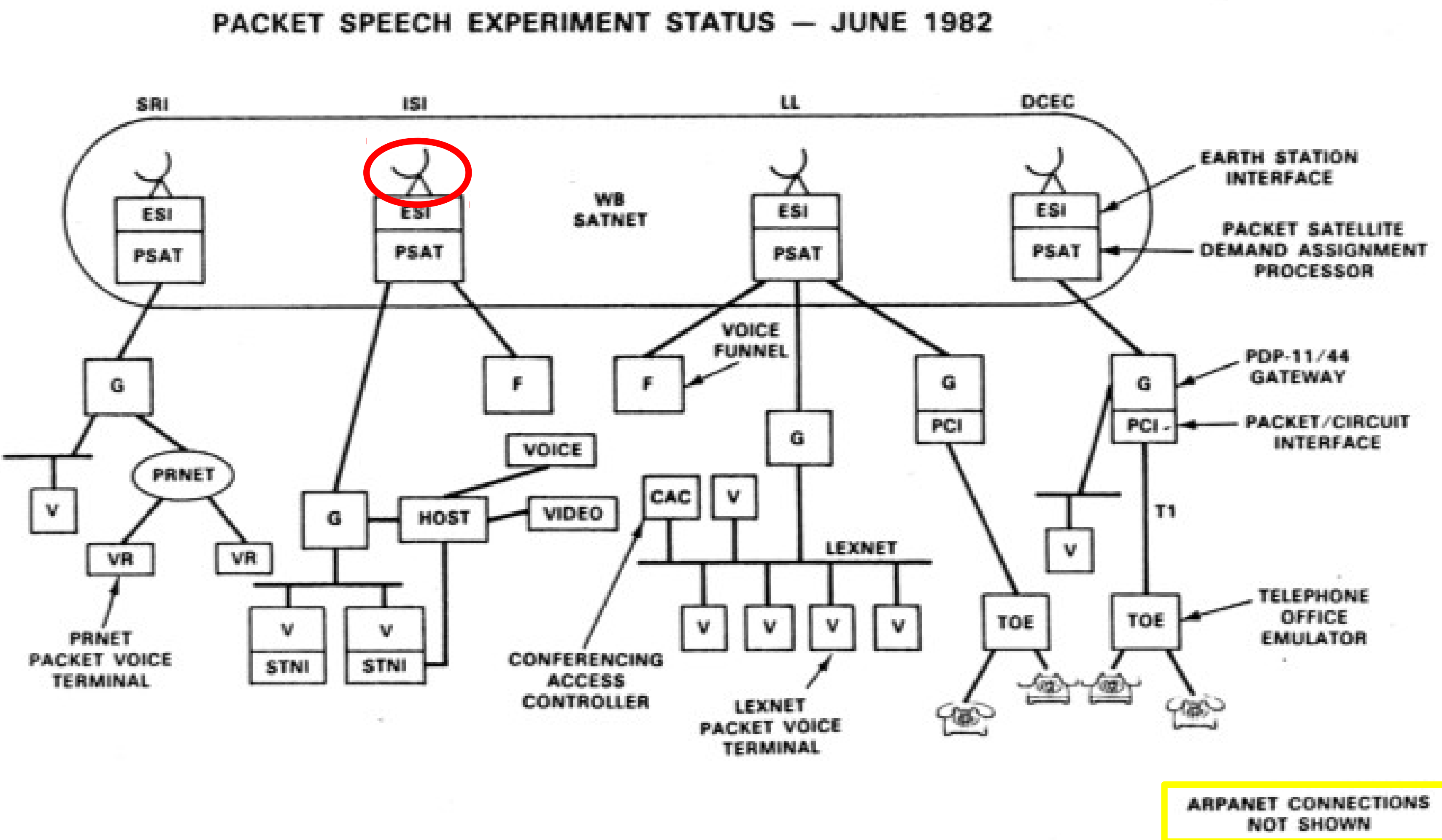
IEN 119, ST - A Proposed Internet Stream Protocol

RFC 1190, Experimental Internet Stream Protocol, Version 2 (ST-II)

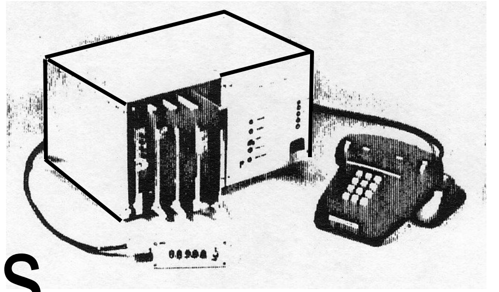
- Connection-oriented, not datagram like IP
- Smaller header using per-hop identifiers
- Point-to-point or point-to-multipoint
- Requests network bandwidth reservation

Many detractors said best-effort IP was sufficient.

Packet Voice Internet, 1982



1982 NSC Program Review



Significant demo events:

- Lincoln Lab packet voice terminals (multi-microprocessor-based, 1.5 ft^3)
- Embedded variable-quality voice coding
- DSP-based LPC encoder/decoder
- 5-party conference, SRI, ISI, 2 @ LL, DCEC (WBnet, PRnet, LEXnet, and STNI to PSTN)

1982-1992 VoIP “Hibernation”

Packet-speech activities waited for the Internet and computers to become ubiquitous, high-performance, low-cost, and audio-capable.

- Web accelerated Internet popularity & growth
- Workstations gained audio devices
- Commercial parties began to take notice

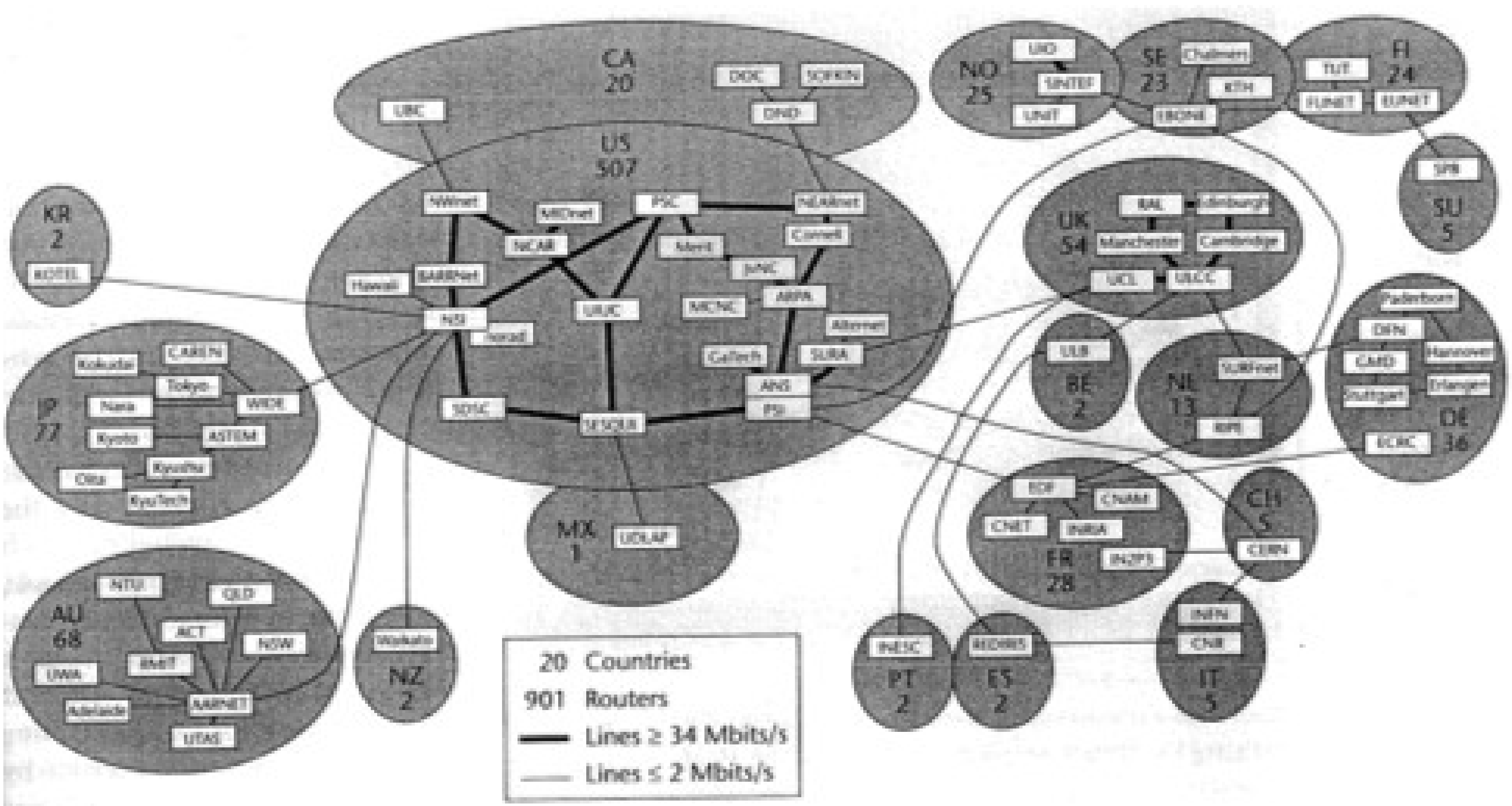
We weren't all sleeping...

- Networks evolved from ARPAnet to WBnet to TWBnet and DARTnet and ...
- Packet video developed at ISI
- Packet voice & video conferencing systems simmered in “experimental production”
- Steve Deering developed IP Multicast
- IETF RSVP, IntServ and DiffServ projects prepared the Internet for RT services & QoS

3/1992: First IETF Audiocast

- NVP-II over UDP over IP multicast
- DARTnet core of 34-node multicast network
- 20 sites spanning 16 time zones
- Audio outbound from and inbound to meeting
- First meeting of the IETF Audio/Video Transport working group (home of RTP)

MBone: Multicast Backbone

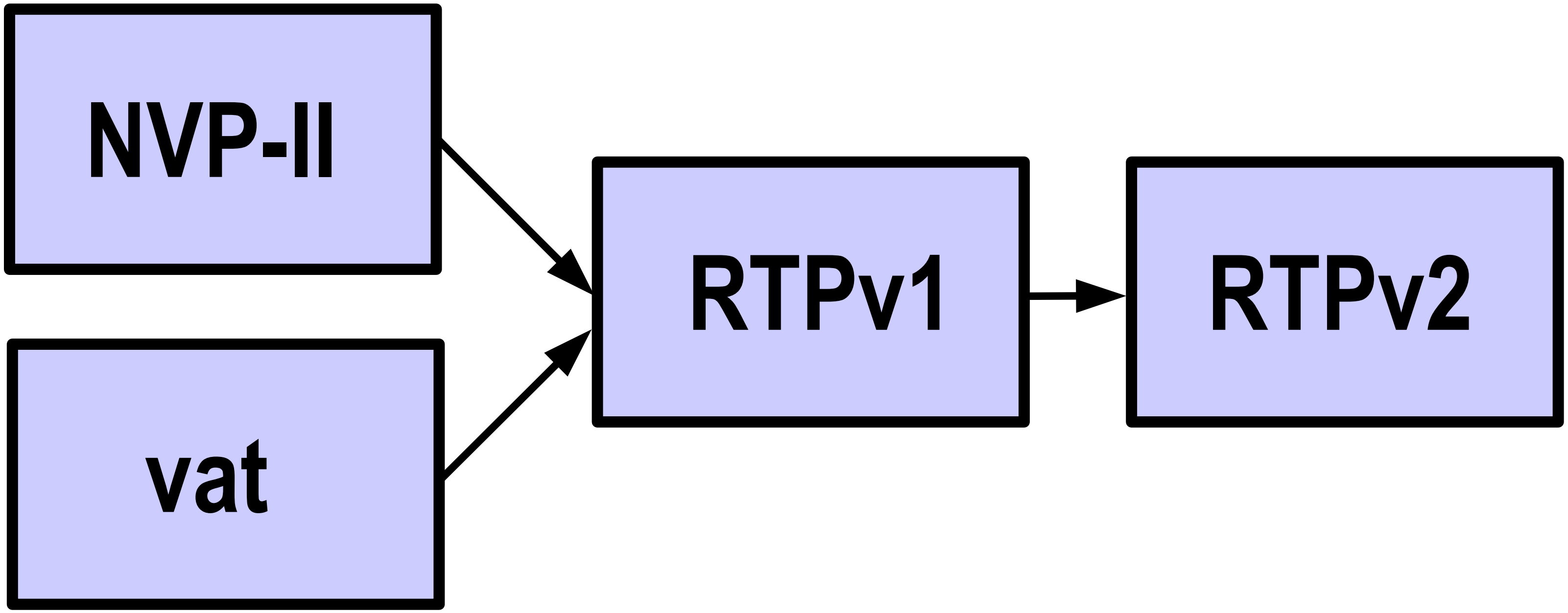


MBone 1994: 20 countries, 901 routers

MBone Highlights

- Provided audio-chatroom for network geeks
- Streamed NASA TV of first Hubble repair
- Broadcasted Rolling Stones concert
  - December 1994 Newsweek article
- Distributed music performance - ACM MM'94
- Session directory sent via IP multicast using Session Announcement Protocol (SAP)

Real-time Transport Protocol



- Concepts from NVP: sequence + timestamp
- Concepts from vat: on-the-fly vocoding switch, source identifiers, start-of-talkspurt bit

NVP-II Data Protocol Header

```
  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|     SEQUENCE NUMBER       |       TIMESTAMP, in vocoder frames   |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|          Header Checksum          |       Token Length (Words)   |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
```

- Sequence number <u>and</u> timestamp needed:
  - Sequence number detects loss
  - Timestamp indicates intentional gaps
- Header checksum for unreliable transports
- Token length (control tokens are optional)
- Data length comes from lower layer

“vat” Data Protocol Header

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|V=0|    NSID   |T| 0 | Format  |          Conference ID        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  Timestamp (in audio samples)                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
:                     0-63 Site Identifiers                     :
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- Audio format/encoding to switch on-the-fly
- Timestamp in samples facilitates format switch
- Start-of-Talkspurt flag indicates gaps
- 0-63 source identifers for pre-mixed audio
- Conference ID for validation

RTP Version 1, November '93

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|V=1| ChannelID |P|S|   format    |       sequence number         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     timestamp (seconds)         |     timestamp (fraction)      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| options ...                                                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- Timestamp always in [seconds:fraction] for simple inter-media synchronization
- Sequence number (in packets) detects loss
- End-of-sync-unit flag: detection without delay
- Audio format/encoding to switch on-the-fly
- Allowed control tokens as options in data pkts
- Channel ID for multiplexing RTP packets

RTP Version 2, 1996 & 2003

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|V=2|P|X|  CC   |M|     PT      |       sequence number         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           timestamp                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           synchronization source (SSRC) identifier            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            content source (CSRC) identifiers                  |
|                             ....                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- Timestamp in samples facilitates format switch
- Sequence number (in packets)
- Start-of-Talkspurt flag for audio; end for video
- Audio format/encoding to switch on-the-fly
- Source identifer for multiplexing and validation
- 0-15 source identifers for pre-mixed audio

RTP Version 2 Philosophy

- Separate control and data ports to allow a 3rd-party multicast monitor to get only control
- Fixed 12-byte header, minimal options: fast processing, simple header compression
- Highly scalable for group size, data rate, and any media with inherent notion of real time
- Absolute vs. incremental values for robustness against packet loss

RTP Version 2 Devolution

In 2009, some backtracking:

- Allow multiplexing RTP and RTCP on one port due to NAT and limited port space
- Header extensions made easier to use, so now multiple options can be chained (but audio & video codecs have options galore anyway)

Always Another Vocoder

DVI4	12-bit DAT, 20- and 24-bit Linear Sampled Audio
G.722	Adaptive Multi-Rate (AMR)
G.723	Adaptive Multi-Rate Wideband (AMR-WB)
G.726	BroadVoice Speech Codecs
G.728	Enhanced AC-3 (E-AC-3)
G.729	Enhanced Variable Rate Codecs (EVRC)
G.729D	Enhanced Variable Rate Wideband Codec (EVRC-WB)
G.729E	Extended Adaptive Multi-Rate Wideband (AMR-WB+)
GSM	G.711.1
GSM-EFR	G.719
L8	G.722.1
L16	G.729.1
LPC	Internet Low Bit Rate Codec
MPA	MP3 Audio
PCMA	MPEG-4 Audio/Visual streams
PCMU	Selectable Mode Vocoders (SMV)
QCELP	Speex Codec
VDVI	Variable-Rate Multimode Wideband (VMR-WB)
	Vorbis Encoded Audio

Control Protocol Evolution

- NVP only scratched the surface of the control side of VoIP – rudimentary sessions
- RTP Control Protocol (RTCP) supports only “loosely controlled” sessions without membership control or negotiation
- For a complete VoIP solution, need session signaling for locating participants, negotiating capabilities, billing, ...

NVP-II Control Protocol

- Series of control tokens with parameters:

  CONNECTION-NAME(address,port,id)
  VOCODING
  I-AM-RINGING
  BYE
  (and several others)

- Provides simple capability negotiation

RTCP: RTP Control Protocol

- Provides feedback on quality of transmission to sender and other multicast receivers
- Carries persistent identifier for source
- Allows counting participants to scale timers and limit control bandwidth consumption
- Optionally conveys session control info for “loosely controlled” sessions
- Carried on a separate transport-level port

Sessions: Beginnings of SIP

- Multicast sessions use multicast directory
- P2P calls need explicit session initiation
  - Addressed in IETF Rem-Conf, ConfCtrl, and MMUSIC working groups
  - Inputs included Etherphone (PARC), Touring Machine (BellCore), MMConf (BBN), CCP (ISI)
  - Standardization of multipoint conferencing came later

History *vs.* Prehistory

Lots of work has been done since early 1990's, sorry we can't cover it all:

- ITU-T adopted RTP for H.323 and defined control protocols to work with it
- SIP development required multiple IETF WGs, is now deployed
- RTSP was developed for streaming

VoIP History a la Google

“The history of VoIP shows that this technology started as far back as 1995 when a small company called Vocaltec released, what was believed to be the first internet phone software”...,

“This was still a major milestone as it represented the first ever IP Phone”

(Downloaded from Google “VoIP History” on 7-20-2010)

VoIP or IPuV?

The name **VoIP** suggests that IP came first and that voice was inserted over IP.

However, the voice application came first, then IP was created under the voice to support it.

IPuV may be a more appropriate name, but is more difficult to pronounce.

Conclusion

- VoIP and Packet Video are major components of the Internet (the “information revolution”)
- Their roots were developed and demonstrated publicly by ARPA projects, starting in the 70's
- Advances in computing, communication, and storage made them practical and ubiquitous
- The carriers no longer think we are crazy...

Recommended Reading

Linear Predictive Coding and the Internet Protocol

Robert M. Gray (2010),
Foundations and Trends® in Signal Processing,
now Publishers, Hanover, MA
ISBN:978-1-60198-348-0

This book is more about audio signal processing and less about packet comm.

End

Movie: http://www.youtube.com/watch?v=MGat1jRQ_SM
Or http://ee.stanford.edu/~gray/dv.mpg
Or http://casners.us/dv.mpg