Audio and Video Over Packet Networks — Issues, Architecture and Protocols

Henning Schulzrinne
GMD Fokus, Berlin
hgs@fokus.gmd.de

Interop'94 Paris

October 27, 1994

Overview

- audio and video characteristics
- problems of integrating audio & video with data
- network support for integration
- RTP: real-time transport protocol

Why bother integrating?

- use existing workstations as audio/video terminals
- use existing LAN/WAN infrastructure
- efficiency:
  - true (LAN/WAN) multicast instead of MCUs and bridges – same application scales from two to hundreds of receivers
  - variable-bit rate (VBR) video, but interoperation with H.261 standards
  - audio silence suppression – important for large-scale conferences

Why bother integrating (cont'd)?

- added control functionality:
  - directory services
  - visual speaker indication
  - source selection at receiver
- integration of application-sharing and data applications, WWW

Disadvantages of packet audio/video

- no resource reservation → quality may suffer (but: RSVP)
- may push overloaded networks over the edge
- low frame rates for workstation video codecs
- hands-free speaking in infancy (echo)
- packetization overhead, delay
- operating systems ill-suited for real-time applications → single-user, no background load
- office environment: acoustics, lighting, …

Characteristics of digital audio and video

	audio	video
rate (kb/s)	13…64…1500	200…1500…6000
loss tolerance	$\leq 5\%$	10^{-5}…1%
packet size	small	large
traffic	interrupted CBR	VBR

One-way delay tolerance:

conference audio without echo cancellation: 40 ms

conference audio with echo cancellation: 150 ms

playback audio/video: $\geq$ 500 ms (“VCR response”)

Integration of Real-Time and Data Traffic

- TCP (and TP4, …) not suited to carry real-time traffic:
  - flow control → window backoff and slow start
  - retransmission delay
- TCP backoff nice for making room for real-time traffic
- options:
  - low load, rely on data traffic backoff (but NFS doesn't)
  - priorities at MAC/link level (starvation!)
  - bandwidth-allocating scheduling at link level
- transport protocols can't help (except with set-up)!

Integration of Real-Time and Data Traffic

- single-priority token rings cannot guarantee bounds (token loss), but may be sufficient
- MAC priorities (token ring, FDDI) won't help (much)
- priorities often not even implemented
- ATM: promises, but (most) switches don't implement traffic control

Ethernet Audio/Video Capacity

802.3 10Base-T Ethernet by Tobagi *et al.* (Infocom'94):

without data

delay	loss	64 kb/s	384 kb/s	1536 kb/s
20 ms	0.001	55 (35%)	14	4
20 ms	0.01	64 (41%)	17	5
100 ms	0.001	89 (57%)	18	5
100 ms	0.01	104 (67%)	20	5
bandw. limit		156	26	6

- effect of TCP backoff not considered
- 100Base-T: about 10 times capacity, as expected
- multicast "defeats" Ethernet switches

Isochronous Ethernet (Iso-Ether) as alternative?

Offers fixed-rate (ISDN basic-rate [BRI]) channels:

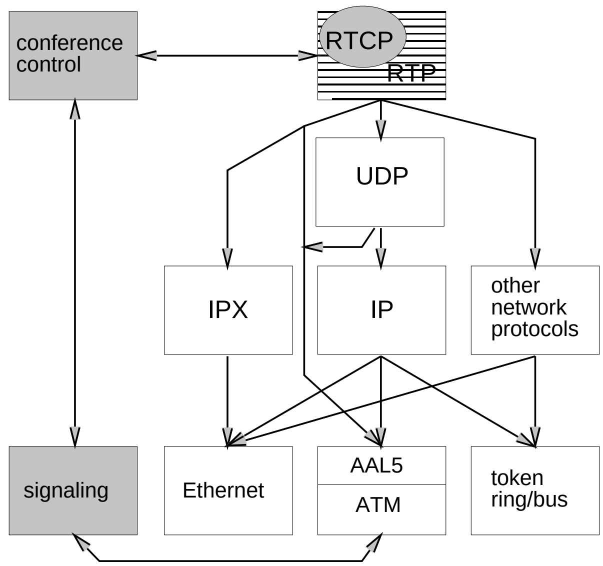
+ predictable quality

? aggregation of 64 kb/s BRI channels?

– not optimal for VBR video, silence-suppressed audio

? multicast other than through MCU (multipoint control unit)?

RTP: The Big Picture



RTP: Real-Time Transport Protocol

- only part of puzzle
- product of Internet Engineering Task Force, AVT WG
- final draft stages
- under consideration by Interactive Multimedia Association (IMA)
- support for functions, but does not restrict implementation

RTP: Goals

lightweight: specification and implementation

flexible: provide mechanism, don't dictate algorithms

protocol-neutral: UDP/IP, ST-II, IPX, ATM-AALx, …

scalable: unicast, multicast from 2 to $O(1000)$

separate control/data: some functions may be taken over by conference control protocol

secure: support for encryption, possibly authentication

RTP: Functions

- segmentation/reassembly done by UDP (or similar)
- resequencing (if needed)
- loss detection for quality estimation, recovery
- intra-media synchronization: remove delay jitter through playout buffer
- intra-media synchronization: drifting sampling clocks
- inter-media synchronization (lip sync)
- quality-of-service feedback and rate adaptation
- source identification

RTP: Bridges, Translators, …

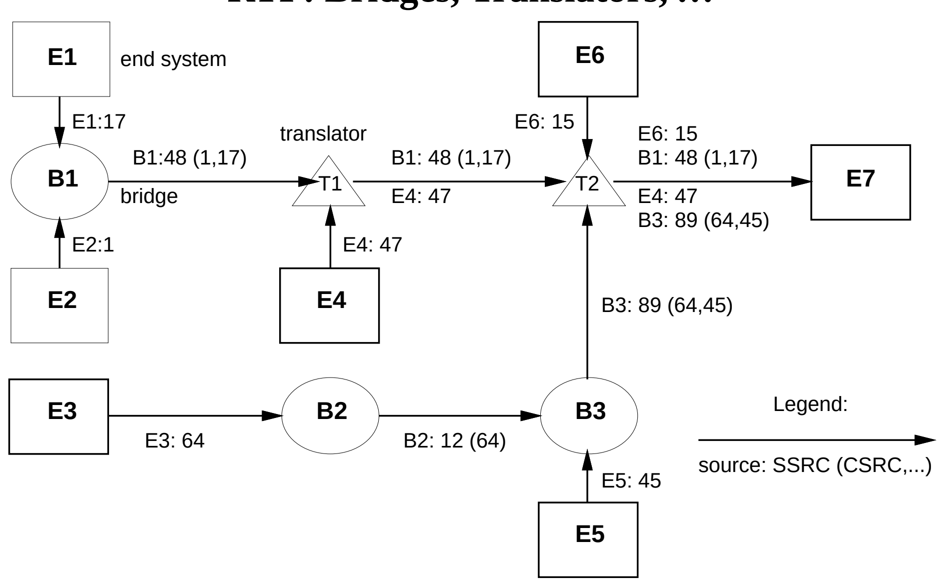
bridge:

- several media stream → one new stream (new encoding)
- mixer: reduced bandwidth networks (dial-up)
- appears as new source

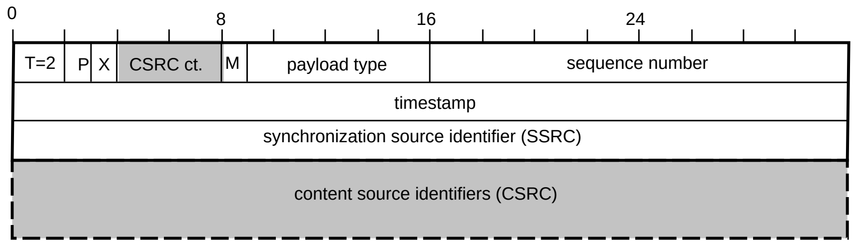
translator:

- single media stream
- *may* convert encoding
- protocol translation, firewall

RTP: Bridges, Translators, …



RTP: Packet format



P: padding (for encryption)

M: marker bit; indicates frame, talkspurt

CC: content source count

Payload type: audio, video encoding method

RTP: Control Protocol – Algorithm

Goals:

- estimate current number of participants – dynamic
- participant information → talker indication
- quality-of-service feedback → adjust sender rate
- scale to $O(1000)$ participants, small fraction of data bandwidth

→ randomized response with rate ↓ as members ↑

- limited by tolerable age of status
- give active senders more bandwidth?

alternative: probabilistic probes (see Turletti's paper in this session)

RTP: Control Protocol – Types

stackable packets, similar to data packets

sender report: bytes send → estimate rate;
timestamp → synchronization

reception reports: number of packets sent and expected → loss,
interarrival jitter, round-trip delay

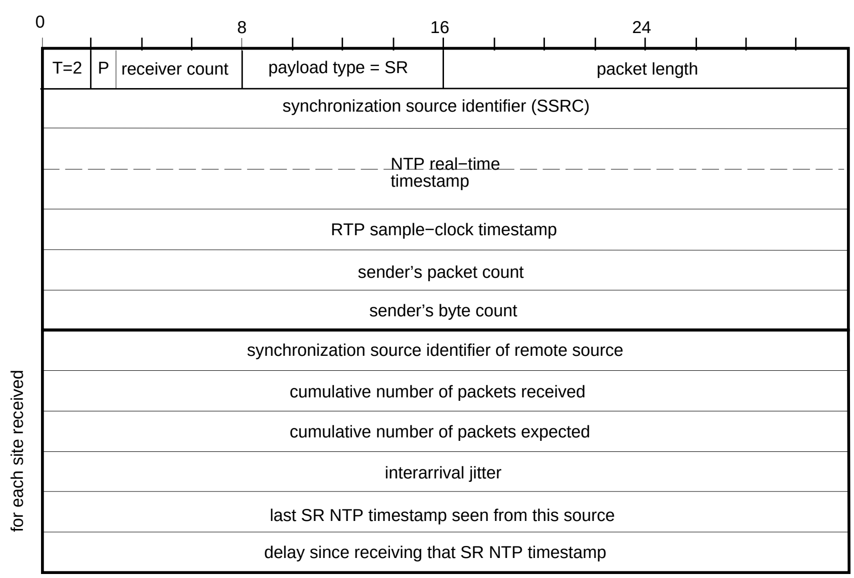
source description: name, email, location, …

explicit leave: in addition to time-out

format mapping: dynamic definition of media formats

extensions: application-specific

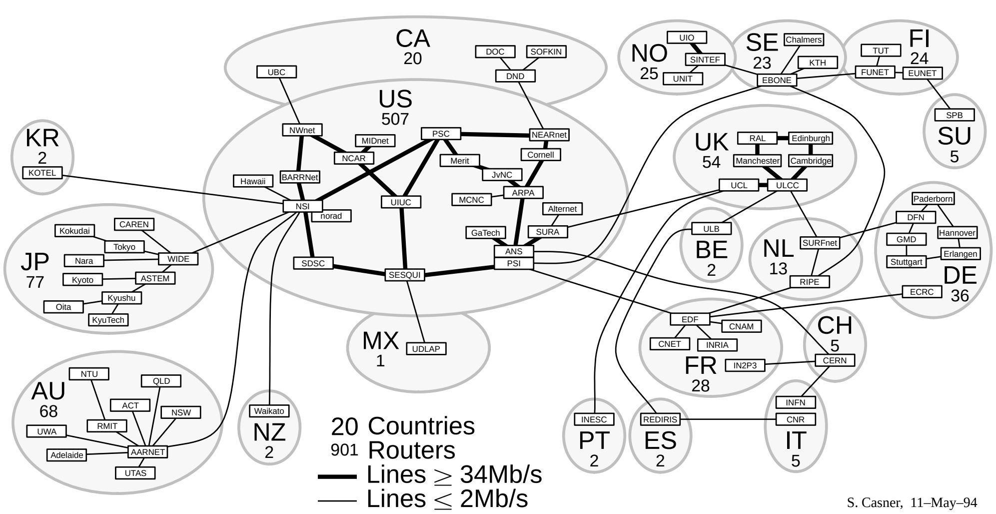
RTP: Control Protocol — SR



Audio/Video in the Internet: The MBONE

- MBONE ≡ multicast backbone
- overlay network over Internet
- needed until deployment of multicast-capable backbone routers
- IP-in-IP encapsulation, “tunnels”
- used mainly for conferences and seminars, weather maps, …
- limited capacity, resilience

Major MBONE Routers and Links



RTP Implementations

"Adaptive applications" → adjust to current network delay

tool	where	application
nv	Xerox PARC	Unix video (proprietary video encoding)
NeVoT	GMD Fokus	Unix audio
ivs	INRIA	Unix audio/video (H.261)

more in progress (PC, Mac, …)

`http://www.fokus.gmd.de/minos/employees/hgs/rtp/faq.html`

Conclusion

- current LANs/WANs can support small number of audio/video connections
- need range of approaches:
  - adaptive applications
  - signaling
  - resource reservation
  - $\boxed{\text{transport protocols}}$
  - switch and router support for QoS