From Mbone and Web-based Multimedia to Experiential Multimedia Systems: Time Travel with WABAC Machine

Klara Nahrstedt

klara@Illinois.edu

University of Illinois at Urbana-Champaign

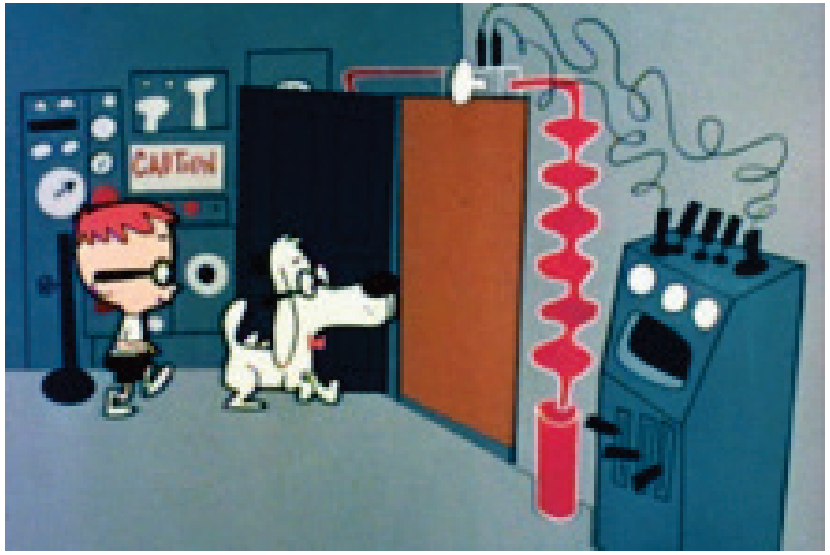
Time Travel with WABAC Machine

Via WABAC Machine or Wayback Machine, 50 Years of "What was, What is and What might be"

- Some of the beginnings
  - From Digital Audio and Etherphone to VoIP and Skype
- Multimedia Protocol Era
  - From Mbone to DASH
- Multimedia Service Era
  - From Vosaic to Human-Centric Multimedia Services
- Multimedia Future and Challenges
  - From 3D TV and Immersion to Experiential Environments

"Sherman and Mr. Peabody enter WABAC machine to witness another time and place in history"

Source: http://en.wikipedia.org/wiki/WABAC_machine



“Sherman, set the WABAC machine to 1980, the beginnings of its’ first digital voice system”

Digital Audio and Voice Systems Era: Then and Now

Digital Audio - 1980

- **Compact Disc Digital Audio** Standard Format – 1980
  - Defined in Red Book
  - Philips and Sony instrumental in developing the CD audio standard
    - Meetings between 1970-1980 between Philips and Sony
- **Concepts:**
  - Signed 16-bit PCM sampled at 44,100 Hz
  - Sample rate according to Nyquist-Shannon sampling theorem
  - Quantization
  - Continuous storage on disk with three parts
    - Audio tracks – program area
    - Lead-in track precedes program area (Table of Content, timecodes)
    - Lead-out track follows program area
  - Bit rate issues
- **Services:**
  - Storage and access to music and audio recordings

- http://en.wikipedia.org/wiki/Compact_Disc_Digital_Audio
- Kees A. Schouhamer Immink (1998). "The CD Story".*Journal of the AES* **46**: 458–465. Retrieved 2007-02-09.

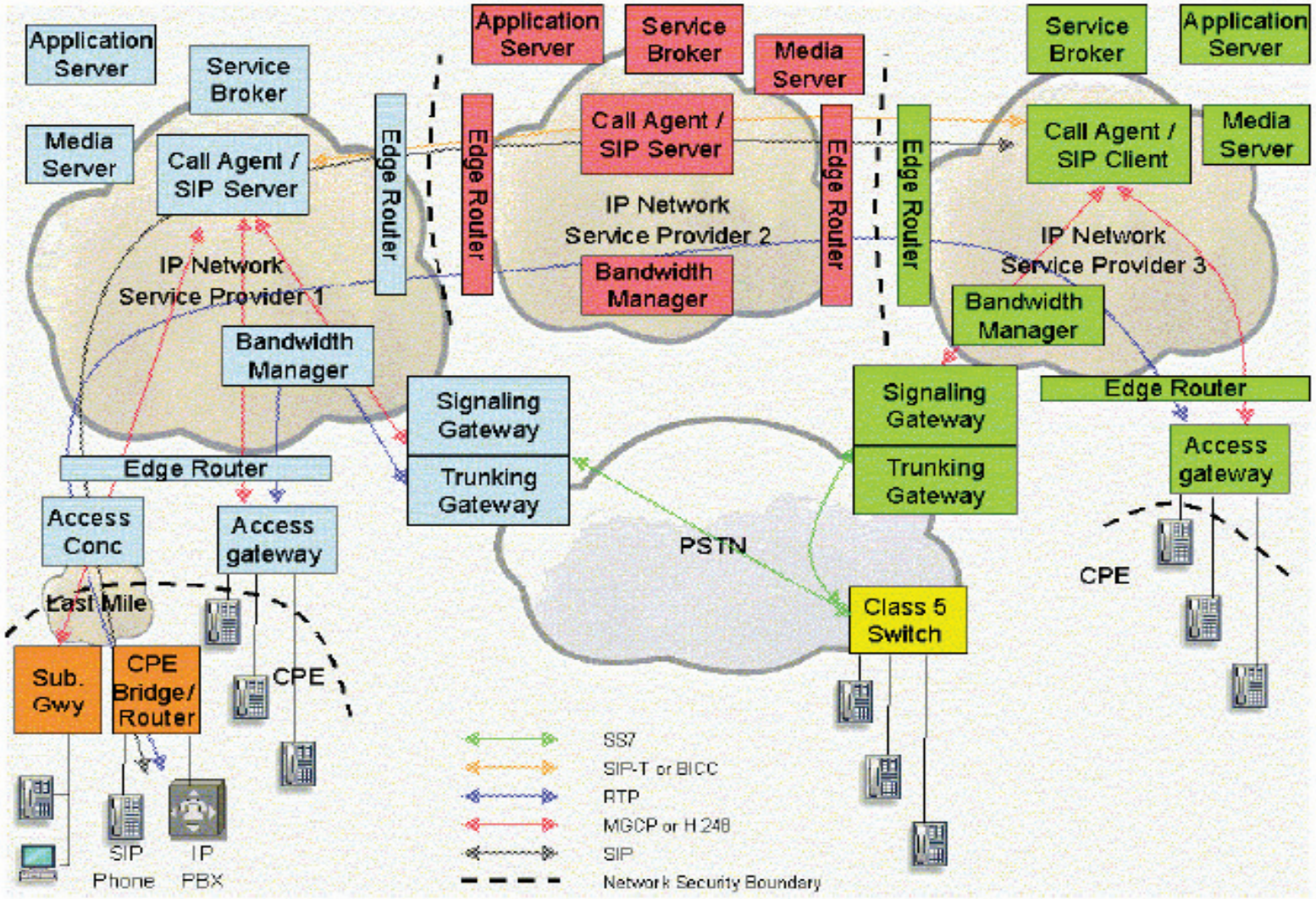
Etherphone 1983-1988

- Xerox PARC Project led by
  - Dan Swinehart, C. Stewart, M. Ornstein
- **Goal:**
  - Adding Digital Voice to Office Computer Network
  - Considering Voice as Data,
  - Introducing Voice Messages Concept
- **Etherphone**^{TM} Concepts:
  - Specially designed processor that
    - Connects to telephone instrument
    - Transmits digitized voice, signaling and control info in discrete packets over Ethernet LAN
- **Services:**
  - Conventional telephone
  - If combined with office workstation, then it provides
    - Voice file service, database service, voice messaging services
  - Integration of voice with other office functions

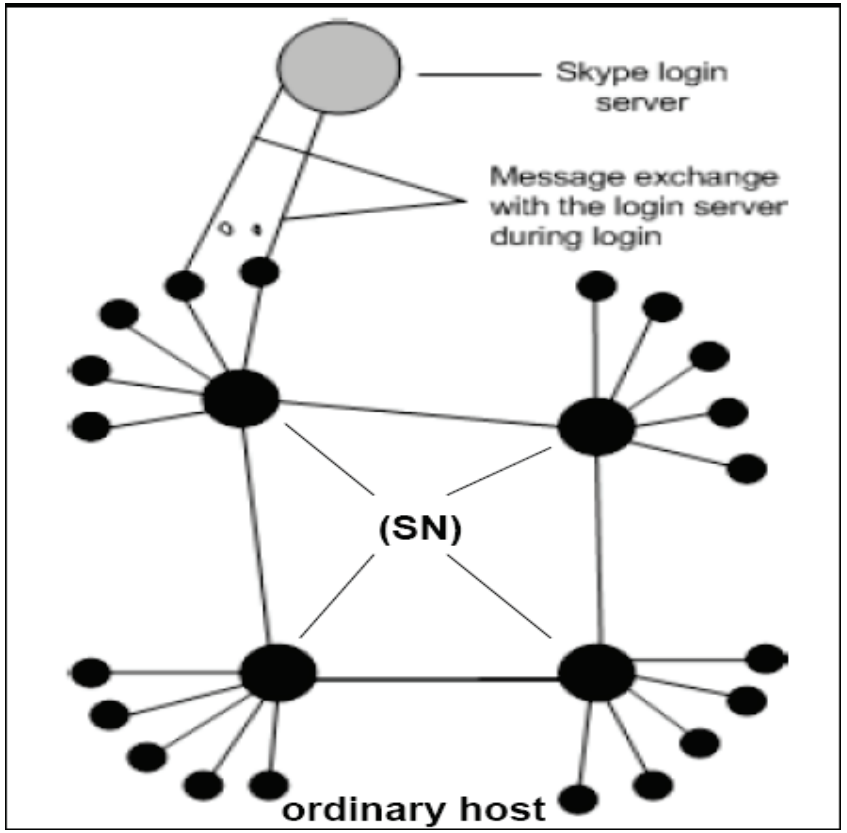
- An Overview of the Etherphone System and its Applications, Polle Zellweger, Douglas B. Terry, Dan Swinehart
- Telephone Management in The Etherphone System, by Dan Swinehart
- Managing Stored Voice in the Etherphone System, by D. Terry, D. Swinehart

http://www.textfiles.com/bitsavers/pdf/xerox/parc/techReports/CSL-83-8_Adding_Voice_to_an_Office_Computer_Network.pdf

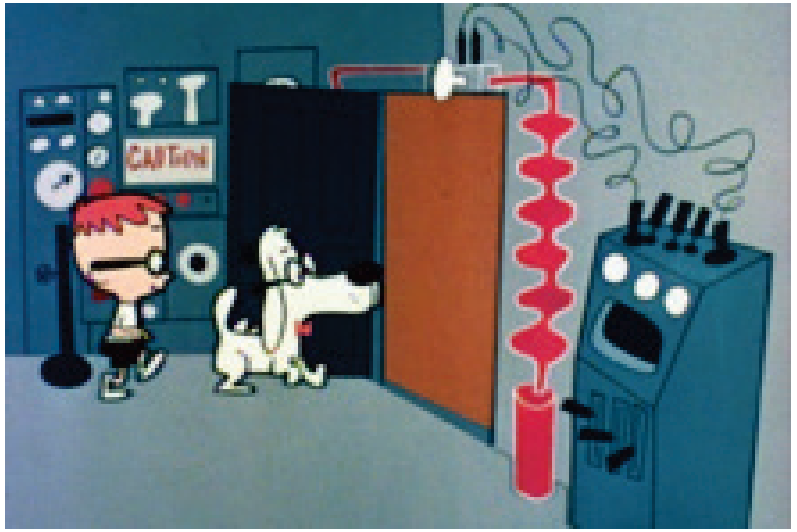
Current Status: Ubiquitous Voice-over-IP and Skype



MSF – Multi-service Switching Forum Example



- An Analysis of the Skype Peer-to-peer Internet Telephony Protocol, S. Baset, H. Schulzrinne, 2004
- Rapid Identification of Skype Traffic Flows, P. Branch et al. , NOSSDAV 2008



“Sherman, set the WABAC machine to 1989, the beginnings of Multimedia Protocols Era”

Digital Multimedia Protocols Era: Then and Now

MBONE - 1989-1997

- **Goal:**
  - Enable multicast audio and video over Internet
- **Concepts:**
  - Mrouter (dedicated multicast router) - segregates MBONE packet delivery, protecting standard packets such as telnet and mail from MBONE experiments
  - Virtual network – shares Internet with multicast support, and some routers are augmented with “tunnels”.
  - Tunneling – enables scheme to forward multicast packets among islands of MBONE subnets through Internet IP routers which do not support IP multicast.
  - MBONE tools – embeds vic/vat and other tools
  - Working Principle - MBONE scheme encapsulates multicast streams into unicast packets which can be passed as regular IP packets along a virtual network of dedicated multicast routers until they reach various destination LAN
- **Services:**
  - Audio and Video conferencing
  - Large-scale Mbone seminars

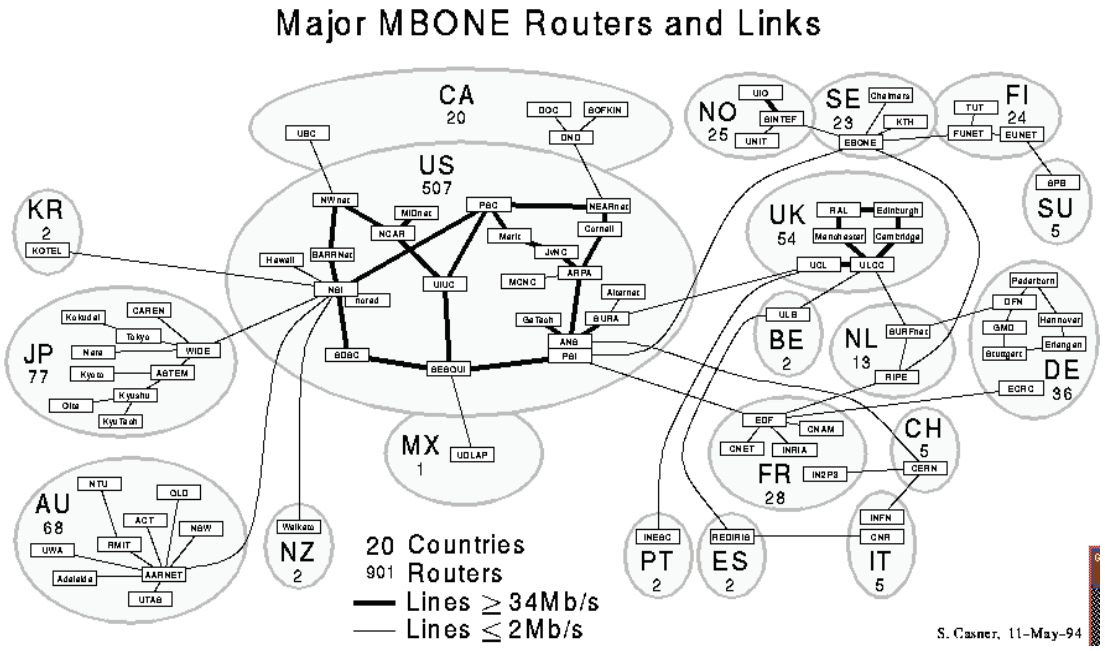
- Casner, Stephen and Deering, Stephen. "First IETF InternEt AudioCast," ACM SIGCOMM Computer Communication Review, San Diego California, July 1992
- Deering, Stephen. "MBONE-The Multicast Backbone," CERFnet Seminar, 3 March 1993.
- MBONE: Multicasting Tomorrow’s Internet, Kevin Savetz, Neil Randall, Yves Lepage, 1996, 1998 http://www.savetz.com/mbone/
- Floor control for large-scale Mbone seminars, Radhika Malpani, Larry Rowe, ACM Multimedia 1997

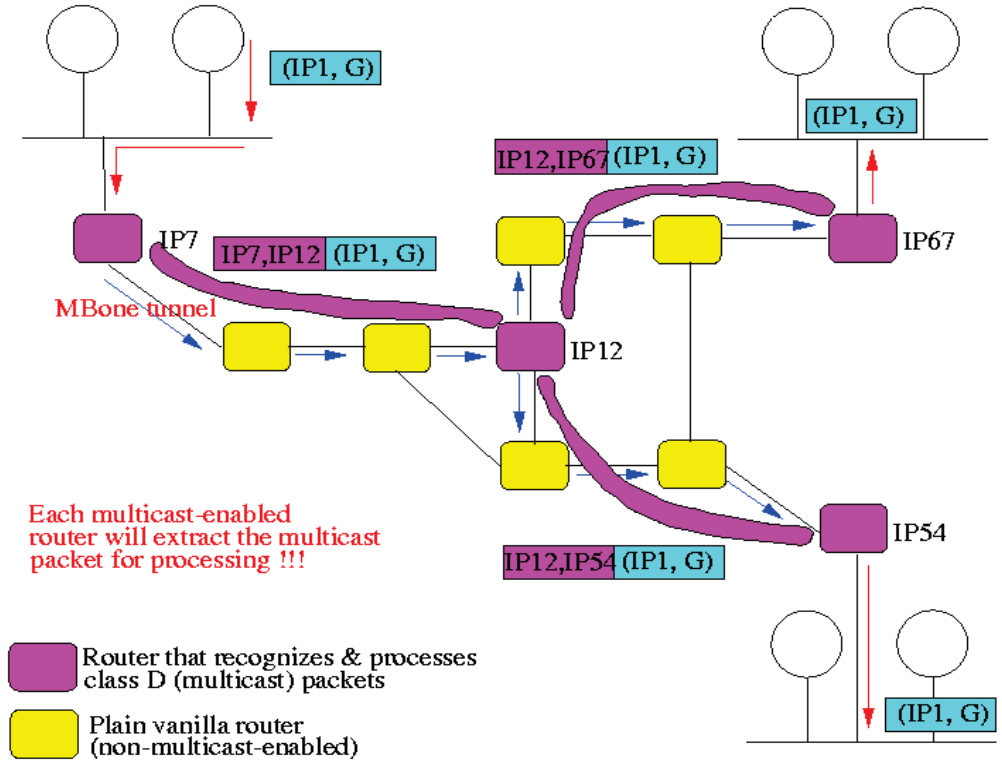
http://www-mice.cs.ucl.ac.uk/multimedia/projects/mice/mbone_review.html

MBONE and IP Multicast

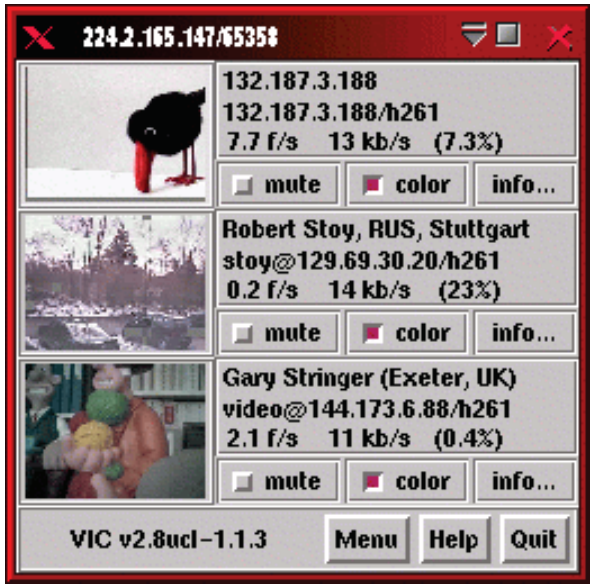
- **Goal:**
  - Enable Internet Protocol Multicast communication after several years of multicast over Ethernet and FDDI
- **Concepts:**
  - IP Multicast addressing
  - Internet standard developed by Steve Deering

Deering, Stephen. "Host Extensions for IP Multicasting", RFC 1112, August 1989





MBONE Tools - vic/vat



- VIC – Videoconferencing Tool and VAT – Audio conferencing Tool
  - Developed by Network Research Group at Lawrence Berkeley National Laboratory in collaboration with UC Berkeley
- **Goal:**
  - Real-time multi-party multimedia applications for audio and video conferencing
- **Concepts:**
  - Intra-H.261 and standard sound cards (halfduplex at first)
  - Based on Real-time Transport protocol RTP (Internet Standard)
- **Services:**
  - Voice-switched viewing windows
    - 'Follow the speaker' feature
  - Interactive "title generation"

http://ee.lbl.gov/vic/
http://ee.lbl.gov/vat/
Van Jacobson
Steve McCanne
1995

MBONE and RTP

- **Goal:**
  - Standardized packet format for delivering audio and video over IP networks
- **Concepts:**
  - RTCP (RTP Control Protocol) – companion protocol to RTP which monitors transmission statistics and QoS, aids synchronization of multiple streams
  - RTP - One of the technical foundations for VoIP and signaling protocol such as SIP (Session Initiation Protocol)
  - RTP - End-to-end, real-time transfer of stream data with jitter compensation and detection of out of sequence arrival of data
  - Profiles and Payload formats
    - Profiles define codecs
    - Payload formats define range of multimedia formats (audio: G.711, MP3, QCELP, video: H.264, MPEG-4, MJPEG, …)
  - RTP – based on the architectural principle known as Application Level Framing (ALF)
- **Services:**
  - Telephony, Video and audio teleconferences

http://www.cs.columbia.edu/~hgs/rtp/
Henning Schultzrinne, November 22, 1995 IETF approved
- RFC 1889, *RTP: A Transport Protocol for Real-Time Applications*
- RFC 1890, *RTP Profile for Audio and Video Conferences with Minimal Control*

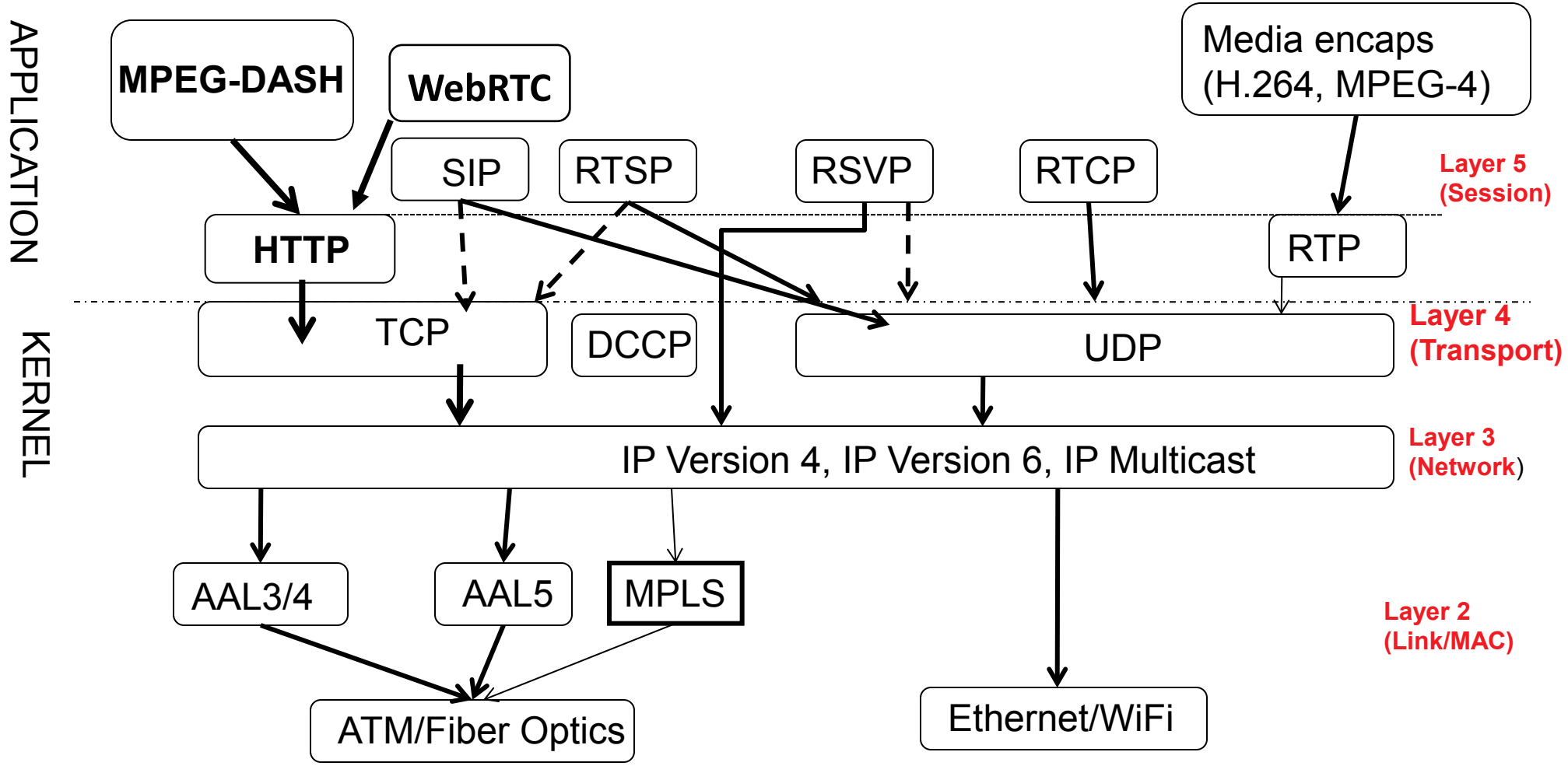
Session Initiation Protocol

- **Goal:**
  - Signaling Communication Protocol widely used for controlling multimedia sessions over Internet Protocol networks
- **Concepts:**
  - Client-Server protocol at the application level
  - Proxy concept
  - Invitation capabilities
  - Support for redirection and mobility
- **Services:**
  - Voice-over-IP

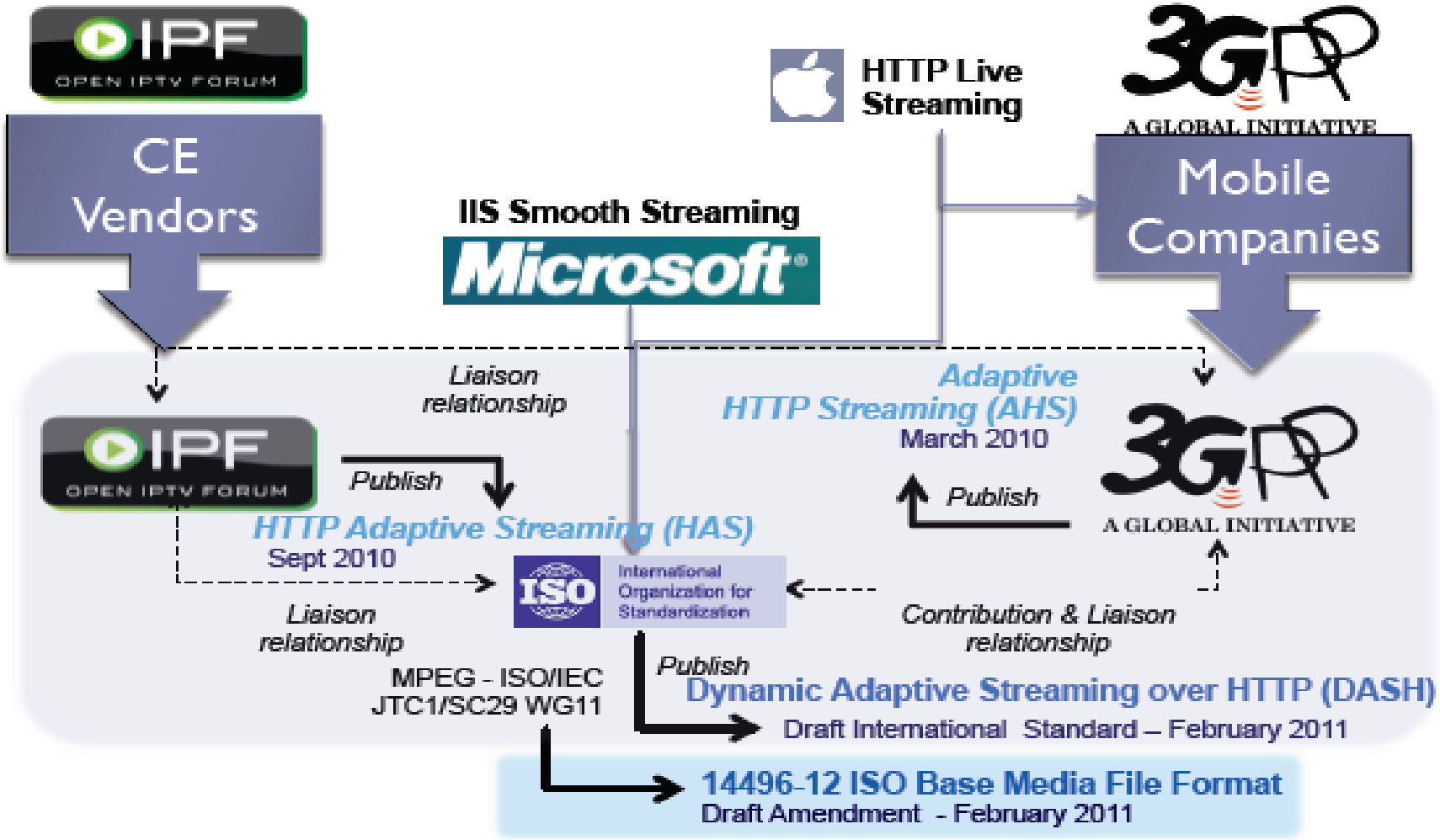
http://www.ietf.org/rfc/rfc3261.txt
SIP: Session Initiation Protocol, J. Rosenberg, H. Schultzrinne, G. Camarillo, A. Johnson, J. Peterson, R. Sparks, M. Handley, E. Schooler, IETF Internet Standard 3261, 2002.

Originally designed by Henning Schultzrinne and Mark Handley in 1996
- http://www.ietf.org/rfc/rfc2326.txt
- http://www.ietf.org/rfc/rfc2974.txt

Current Status: Internet Multimedia Protocol Stack



DASH Standardization History



Thomas Stockhammer, Qualcomm, “DASH – Design Principles and Standards , Presentation at MMSys 2011

Current Status: DASH and WebRTC

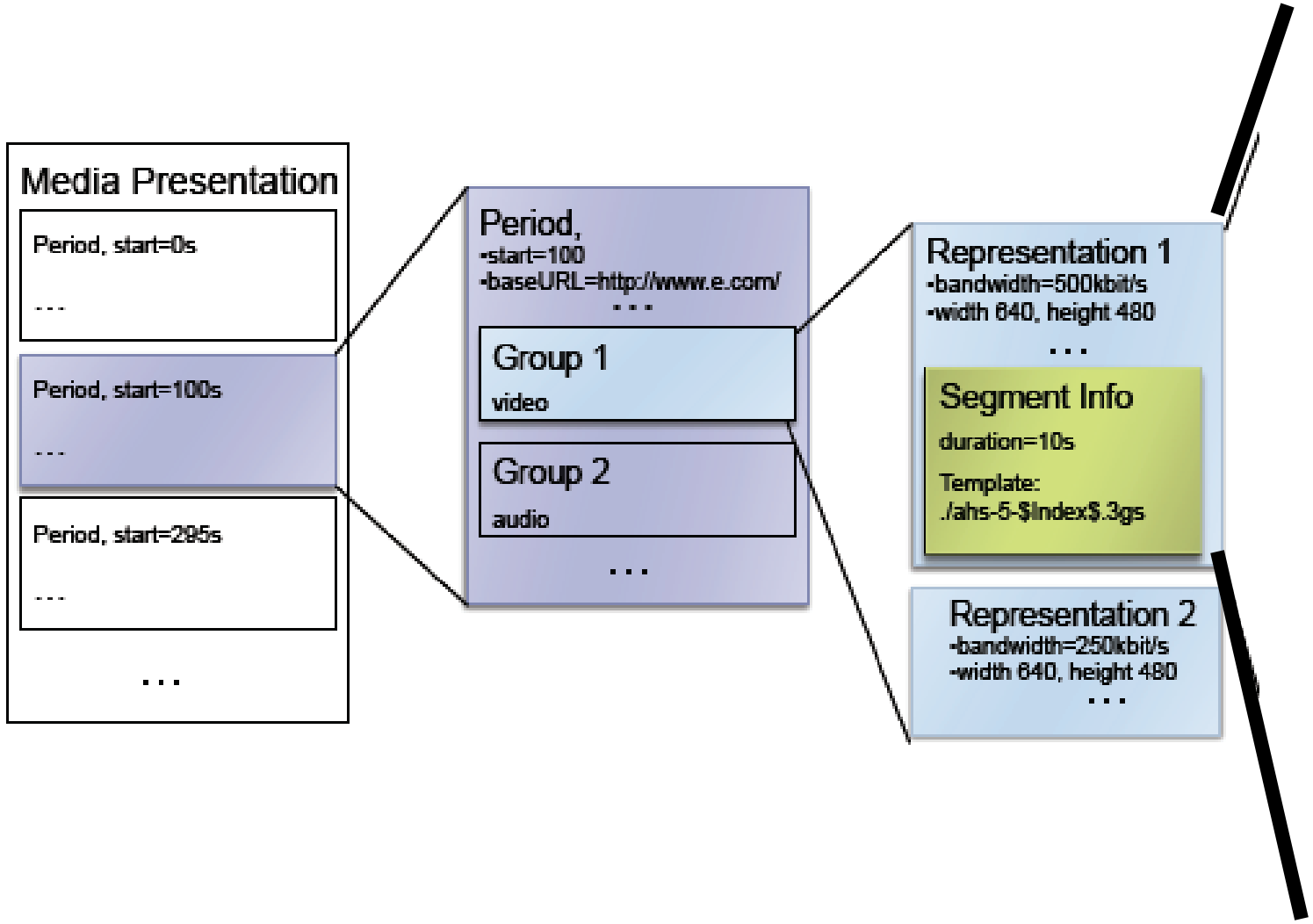
- **Goal:**
  - Provide real-time communication over web interfaces and HTTP/TCP/IP protocol stack
- **WebRTC Concepts** (W3C): 2011
  - Supports real-time communication browser-to-browser applications (Google Chrome, Mozilla Firefox, Opera)
  - Accesses camera and microphone from web browser
  - Sets up audio/video calls
  - Shares data via peer-to-peer
- **Services:**
  - Voice Calling, Voice Chat, P2P file sharing

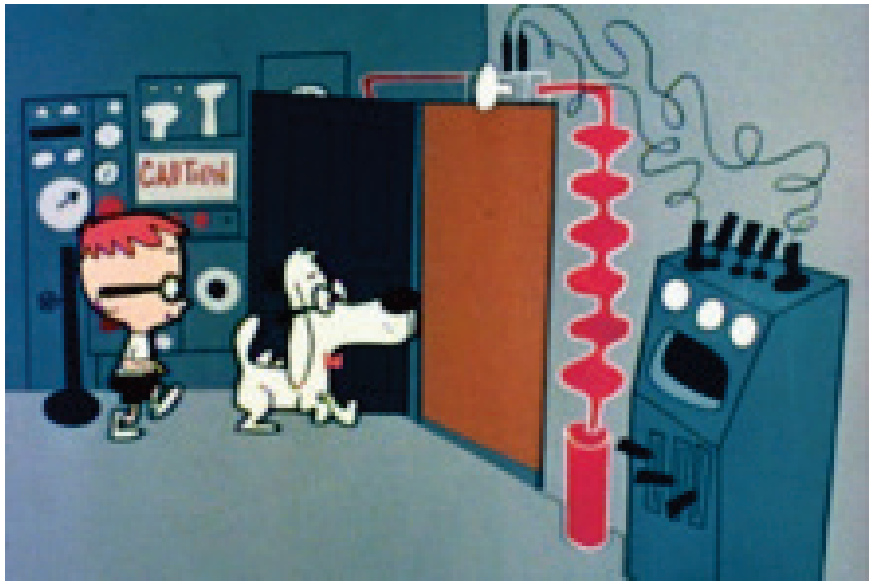
- **Goal:**
  - Enable optimal streaming video viewing experience for diverse range of devices over broad set of connection speeds over web interface
- **Concepts:**
  - Production of multiple files from the same source file to distribute to viewers watching on different devices via different speeds
  - Adaptive distribution of files
  - Transparent operation
  - Adaptive bit rate
  - Media presentation description
  - Concept of segment with media data and metadata – support of large set of formats and profiles
- **Services:**
  - Video-on-demand

Thomas Stockhammer, Qualcomm, "DASH – Design Principles and Standards , Presentation at MMSys 2011
Web Real-time Communications Working Group, http://www.w3.org/2011/04/webrtc/ 2011 Draft
IETF RTCWeb , http://tools.ietf.org/wg/rtcweb/, 2014 Draft

Media Presentation Data Model

MDP - description of accessible segments and corresponding timing



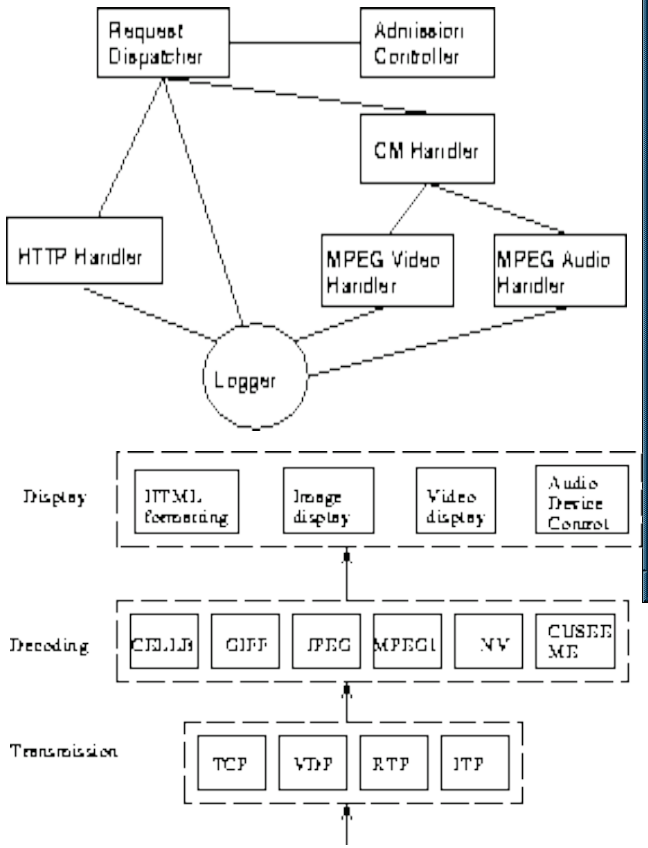


“Sherman, set the WABAC machine to 1996, the beginnings of Multimedia Web and IPTV Services”

Digital Multimedia Web and IPTV Services Era: Then and Now

Video-enhanced Web Service - 1996

- **Vosaic**
  - Video Mosaic from University of Illinois at Urbana-Champaign
  - AV Extension of NCSA Mosaic - 1996
- **Goal:**
  - Design a WWW browser Vosaic (Video Mosaic), that incorporates real-time video and audio into hypertext pages
- **Concepts:**
  - AV transfer in real-time (no file retrieval latency)
  - Real-time streaming protocol DVP (augmented RTP) over MBONE infrastructure
  - Audio and video icons within WWWW hypertext document
  - Vosaic client adapts to received video rate
  - Dual protocols – for data HTTP/TCP and for AV VDP
- **Services:**
  - Video-enhanced Web Service



Zhigang Chen, See-Mong Tan, Roy Campbell, Yongchen Li, "Real-time Video and Audio in the World Wide Web", http://choices.cs.uiuc.edu/Papers/New/vosaic/vosaic.html, 1996.

Interactive Web-based Multimedia – 2001-2010

- **Goal:**
  - Encoding multimedia presentations for delivery and manipulation of multimedia over Web
- **Concepts:**
  - Web-based multimedia for diverse devices
  - SMIL language document
  - Authoring systems for SMIL
  - Encoding peer-level multimedia annotations
- **Services:**
  - Video annotation
  - Editing of Active temporal multimedia documents
  - Personalization and sharing of time-based annotation of videos on web

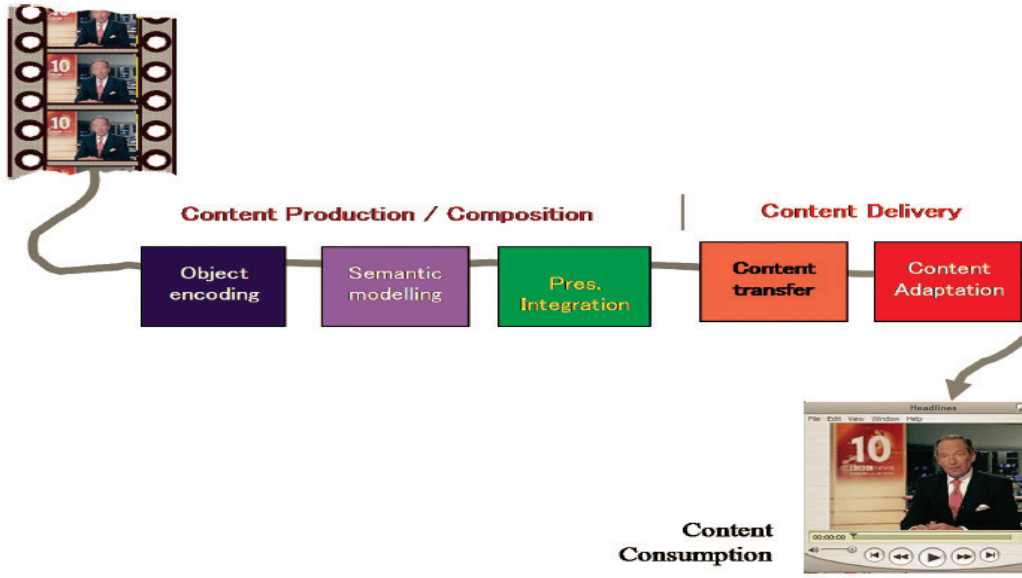
- J. Ayer, Aaron Cohen, Dick Bulterman, "Synchronized Multimedia Integration Language (SMIL) 2.0, W3C, 2001
- Dick Bulterman, SMIL 2.0: Overview, Concepts and Structure, IEEE Multimedia, 8(4), 2001
- Dick Bulterman, SMIL 3.0: Interactive Mltimedia foir the Web, Mobile Devices and Daisy Talking Books, Springer Verlag, December 2008.
- R. L. Guimaraes, P. Caesar, D. Bulterman, Creating and Sharing Personalized Time-based Annotations of Videos on the Web, ACM DocEng, 2010

Current Status: YouTube

https://www.youtube.com/

- **Goal:**
  - Web-based Video sharing website – 2005
- **Concepts:**
  - Adobe Flash Player plug-in on the browser
  - HTML5 standard for YouTube
  - Adobe Dynamic Streaming for Flash (DASH version)
  - Duration of videos up to 15 minutes
  - Diverse number of formats .AVI, .MOV, MP4, .ogg ogv, 3D video formats (Nvidia 3D Vision)
  - Progressive scanning
- **Services:**
  - Promotional platform; individual content creation; training (“How to” videos), education (Khan academy),

Digital Television Service

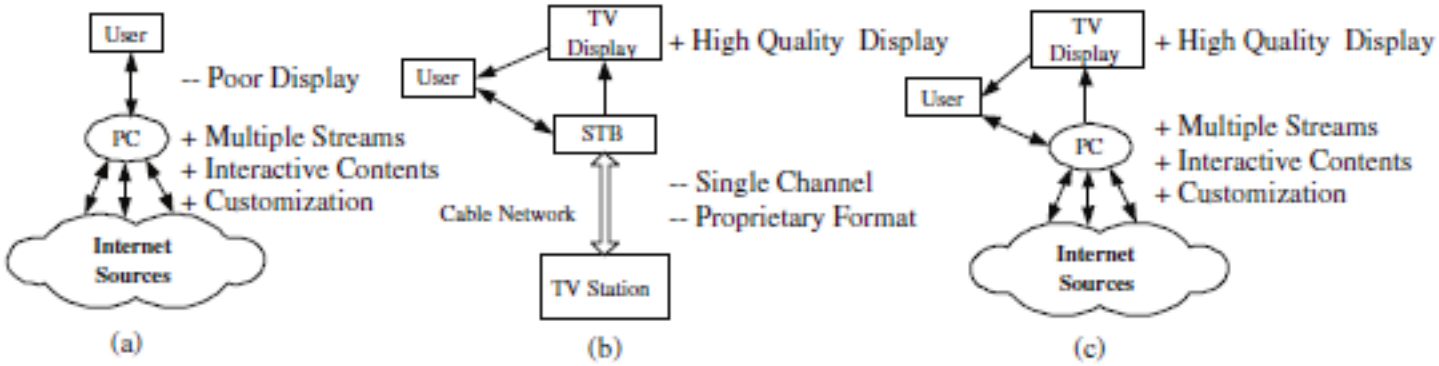


- First generation of interactive television – 2000-2006
  - Digital TV + Set-top Box
- **Goal:**
  - Content sent to end users, primary goal to deliver video con limited interactivity
- **Concepts:**
  - MPEG-2 and MPEG-4 video formats and descriptions
  - Integration and synchronization via SMIL, Flash, Nested Context Language (NCL)
    - Integration languages for multimedia presentations to define flexible spatial and temporal relationships among media
  - Open Cable Platform, Multimedia Home Platform
  - Digital Broadcasting Systems Standards: DVB, ATSC, ISDB
- **Service:**
  - Efficient distribution of massive amount of digital content
  - TiVo ability to alter distribution of schedules
  - Program management - Electronic Program Guide
  - TiVo – PLAY, PAUSE, FORWARD, REWIND

- Pablo Caesar, Dick Bulterman, Luiz Fernando Gomes Soares, “Introduction To Special Issue: Human-Centered Television – Directions in Interactive Digital Television Research” , ACM TOMCCAP 2008.
- L. Ardissonto et al, “Personalized Digital Television Targeting Programs to individual Users, 2004
- L. Smyth, P. Cotter, “A personalized television listings service”, Comm. Of ACM 2000.

Interactive Internet Television (IPTV) – 2003-2008

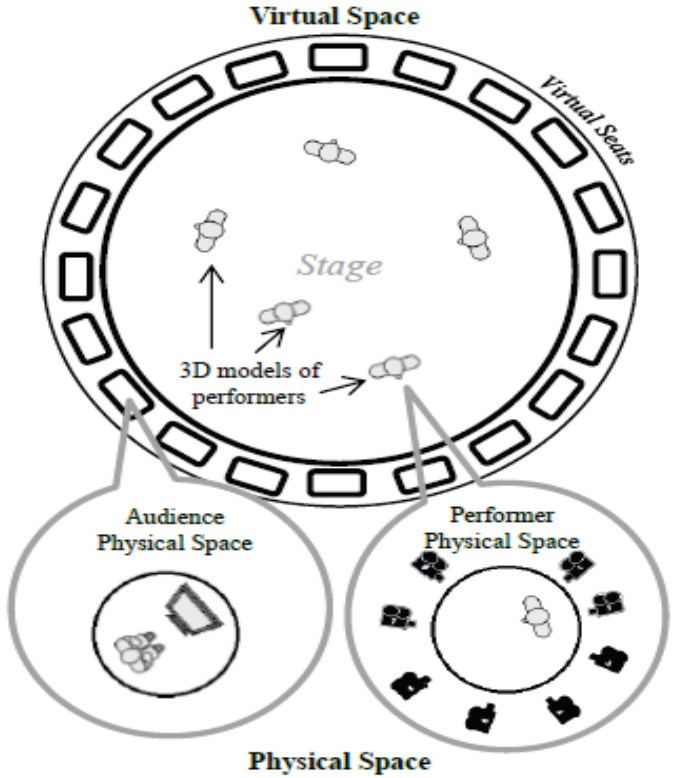
- Enhancements of First Generation Television
- **Goal:**
  - Transmit HDTV content over Internet, but use TV display for high quality; Enrich content and edit content to embed data information (web-related, control information) into MPEG-2 content
- **Concepts:**
  - Active service model
  - Distributed data servers
  - Flexibility and customizability of content
  - HDTV MPEG—2 video transport
  - System-centric
  - Automated directors for content composition
- **Services:**
  - Integrated email and movie watching
  - Editing text and viewing video
  - Multi-channel view via Picture-in-Picture
  - Web-browsing



- E. Amir, S. McCanne, R. Katz, "An active service framework and its application to real-time multimedia transcoding", 1998
- B. Yu, K. Nahrstedt, "Internet-based Interactive HDTV", 2003
- Pablo Caesar, Dick Bulterman, Luiz Fernando Gomes Soares, "Introduction To Special Issue: Human-Centered Television – Directions in Interactive Digital Television Research" , ACM TOMCCAP 2008.

Current Status: Human-centric IPTV

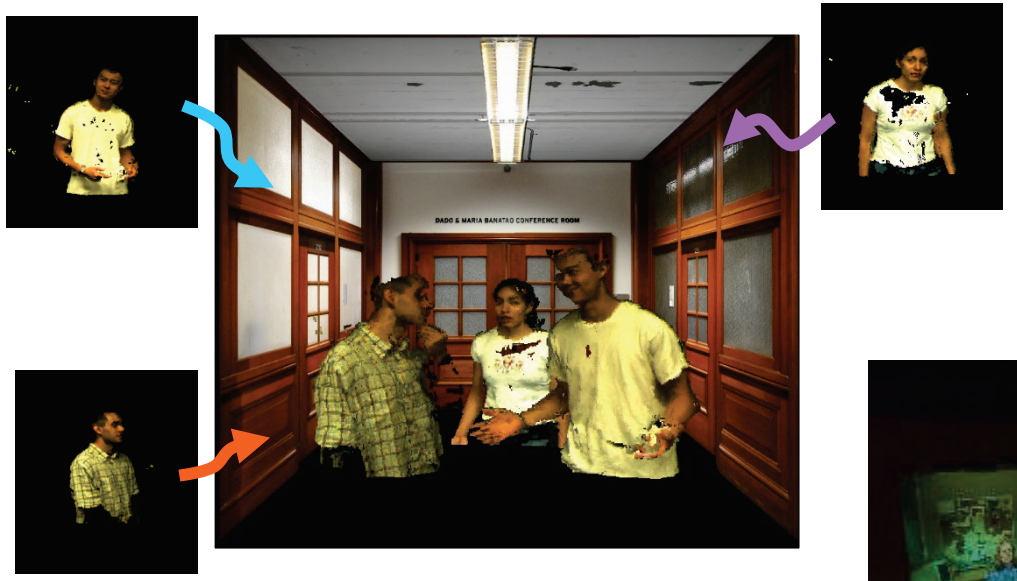
- Second generation of interactive television – 2008-2015
- **Goal:**
  - End user becomes central focus of the interaction infrastructure
- **Concepts:**
  - Shift in content production – lightweight authoring systems
  - Shift in media delivery – media sharing
    - Richer media sharing methods: client-server, peer-to-peer, messaging, chat
      - MessengerTV (Microsoft), Zync! (Yahoo!),
  - Shift in media consumption – social television
    - Merging personal content consumption and social sharing
    - Dual viewing (multiple screens – viewing and/or chatting and/or searching for additional information)
  - Inter-sender and Inter-receiver synchronization
  - Social network integration
- **Services:**
  - Virtual fans (having a virtual seat)



- N. Ducheneaut et al, "SocialTV: Designing for distributed, social television viewing" 2008
- P. Caesar, D. Bulterman et al, "Enhancing social sharing of videos: fragment, annotate, enrich and share", ACM MM 2008
- I. Vaishnavi, P. Caesar, D. Bulterman, "From IPTV to Shared Experiences: Challenges in Architecture Design", 2010.
- S. Chen, K. Nahrstedt, I. Gupta, "3DTI Amphitheater", ACM MMSys 2014

Current Status: Next Generation Telepresence

- 3D Teleimmersion – Holy grail since 1990th
- **Goal:**
  - Instead of having separate video conferencing windows, enable immersion into virtual room
- **Concepts:**
  - Multi-stream streaming platforms
  - Inter-sender and Inter-receiver synchronization
  - Quality of Experiences
  - Session management revisited – IETF Clue
  - View-casting
- **Services:**
  - Tele-dancing, Exergaming, Training, Tele-rehabilitation



- Henry Fuchs: Beyond the Desktop Metaphore: Towards More effective display, interaction and collaboration in the office of the future via a multitude of sensors and displays, 1998
- http://graphics.cs.brown.edu/research/telei/teleImmersionFoundations.html
- Z. Yang, K. Nahrstedt. R. Bajcsy, “TEEVE: The Next Generation Architecture for Teleimmersive Environments”, 2005.

“Sherman, set the WABAC machine to 2025, to see Future Multimedia Systems”

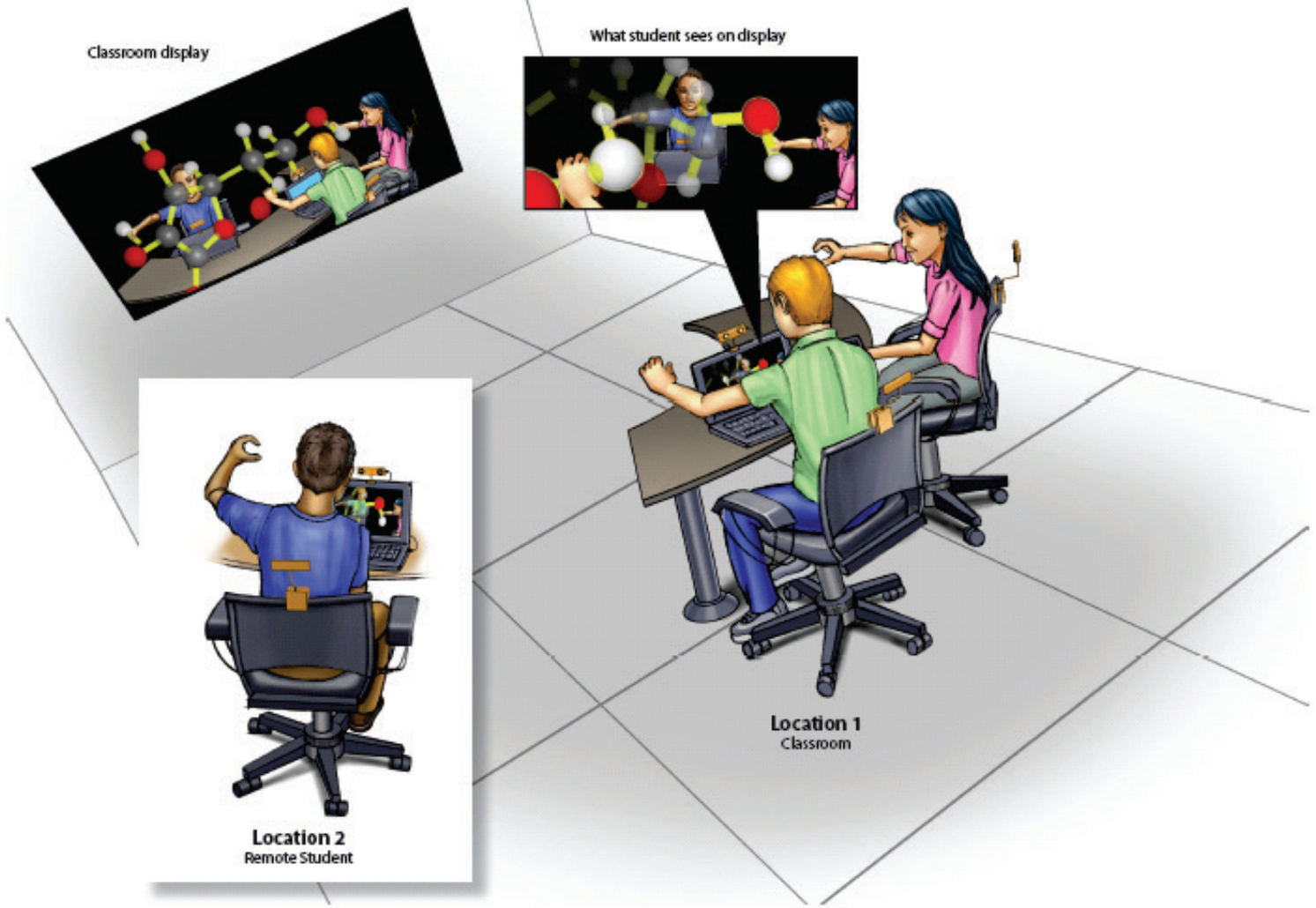
Future Multimedia Systems: Vision and Challenges

Vision

- **Sharing, experiences, habits, activities** are understood by our digital environments

- **Boundaries between conferencing and on-demand environments** are blurred towards experiential multimedia environments enabling seamless socially aware telepresence participation and television interaction

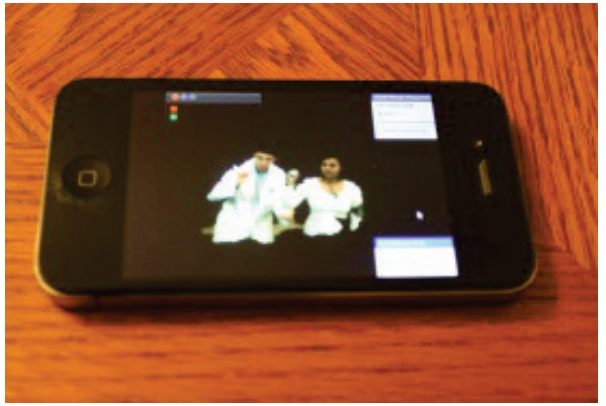
Imagine: Ad Hoc Immersive Spaces for Education



Imagine: Blurring Telepresence with Social Experience for Tele-healthcare

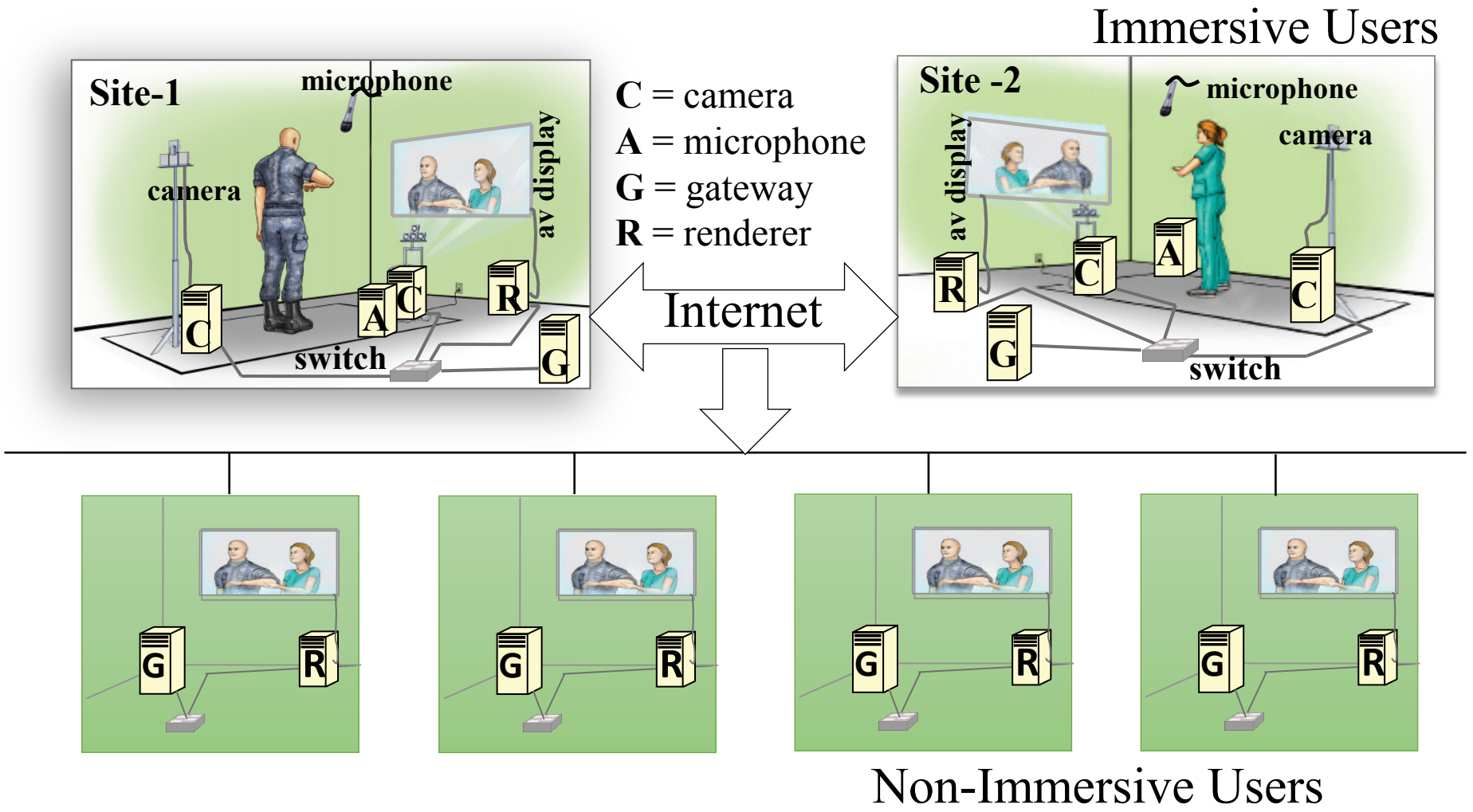
Experiential Applications: Challenges

- **Dynamic 3D multi-resolution cameras**, their calibration and rendering
  - Real-time zoom-in and zoom-out
- **Heterogeneous wireless connectivity** of I/O devices
  - Wifi, Bluetooth, Zigbee
- **Easy Interface and Usage**
  - Fast configuration, setup in clinics, homes
- **High Interactivity** and Perceptual Quality
- **Tactile haptic** – remote touch reproduction
- **Multi-modal real-time communication**
- **Real-time conferencing and storage retrieval**
  - If doctor is present in real-time or as a stored avatar should not matter



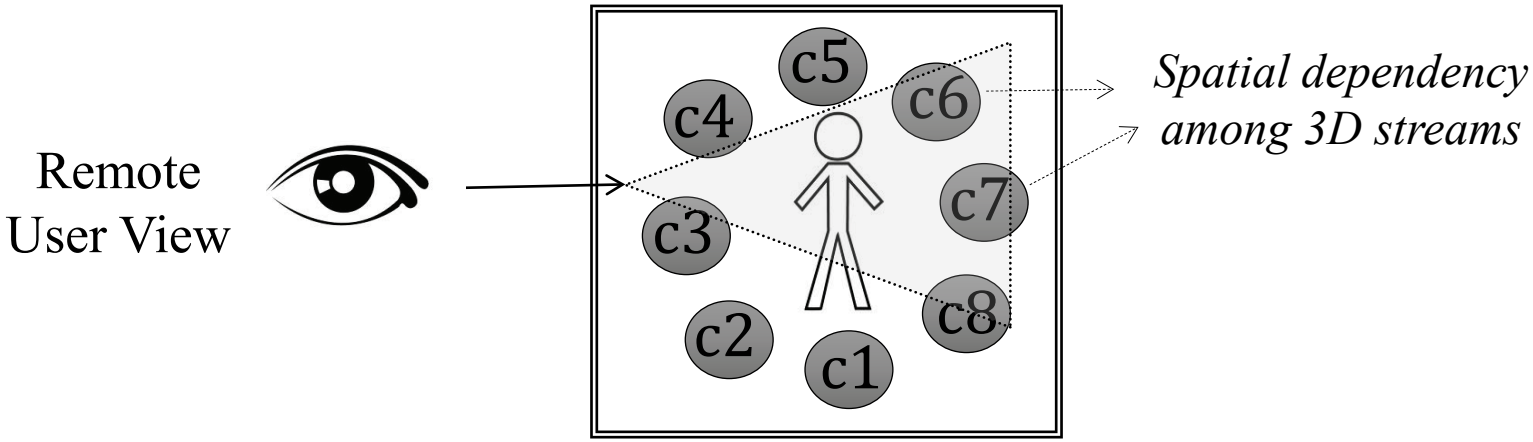
Underlying Experiential Systems: Challenge

- **Heterogeneous users**



Experiential Systems: Challenge

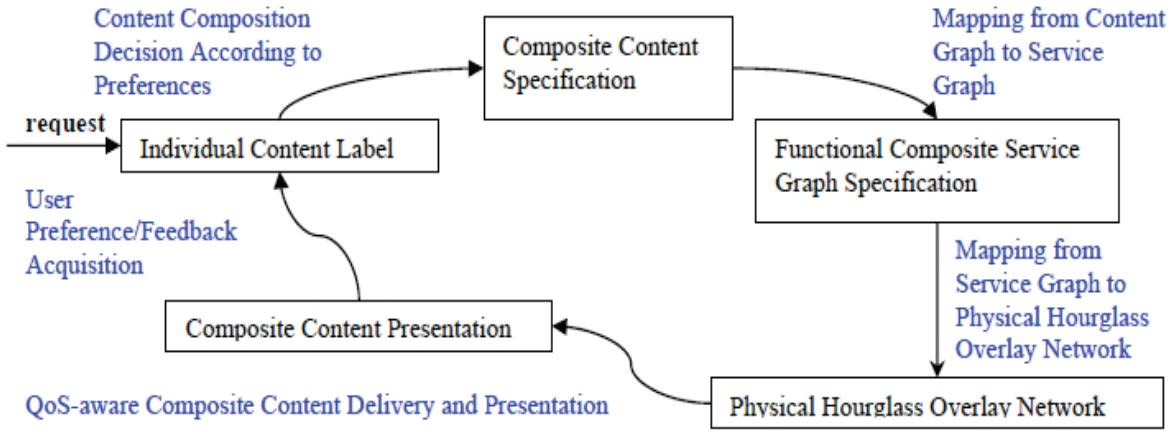
- Heterogeneous users
- **Multi-stream and multi-view environment**

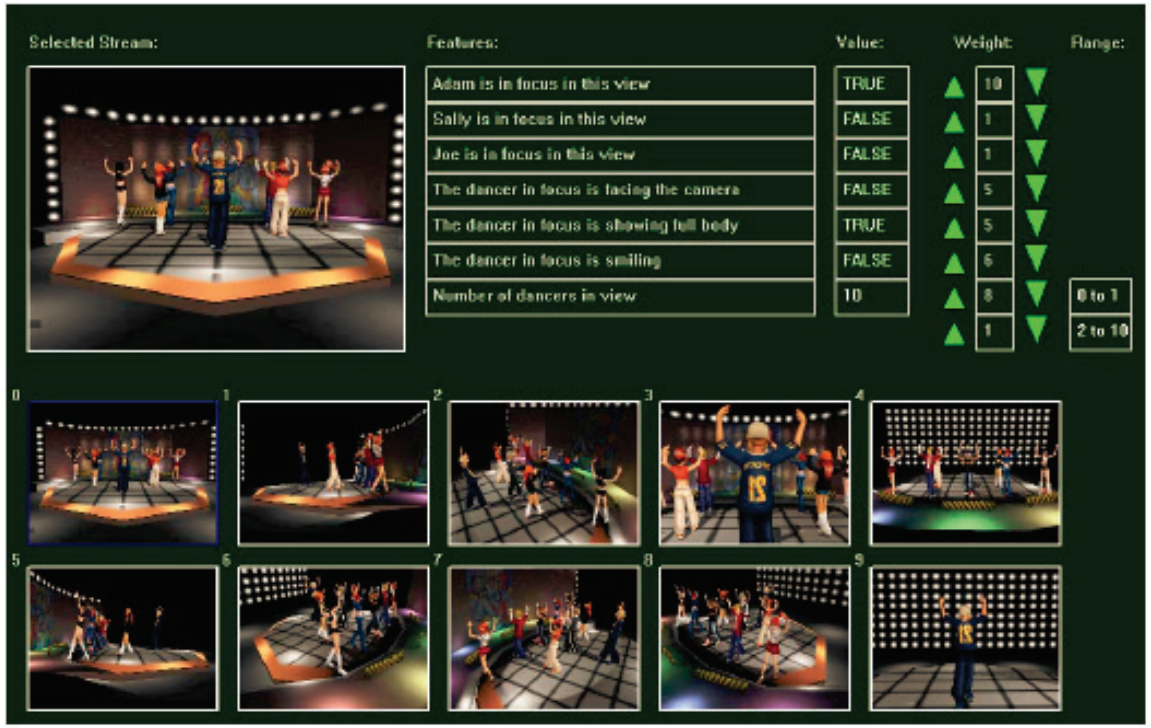


Example of 3D representation captured by different cameras

Experiential Systems: Challenge

- **Multi-view/content composition specification**
  - Content-to-Service-Infrastructure Integration
  - MPEG7, MPEG-21, SMIL 4.0?
  - Differentiation of Content via User-centric Scoring or System-centric automation?
- **Service Composition**
  - Automated or User-in-the-loop?
- Mappings between Content-Service-Infrastructure Compositions?

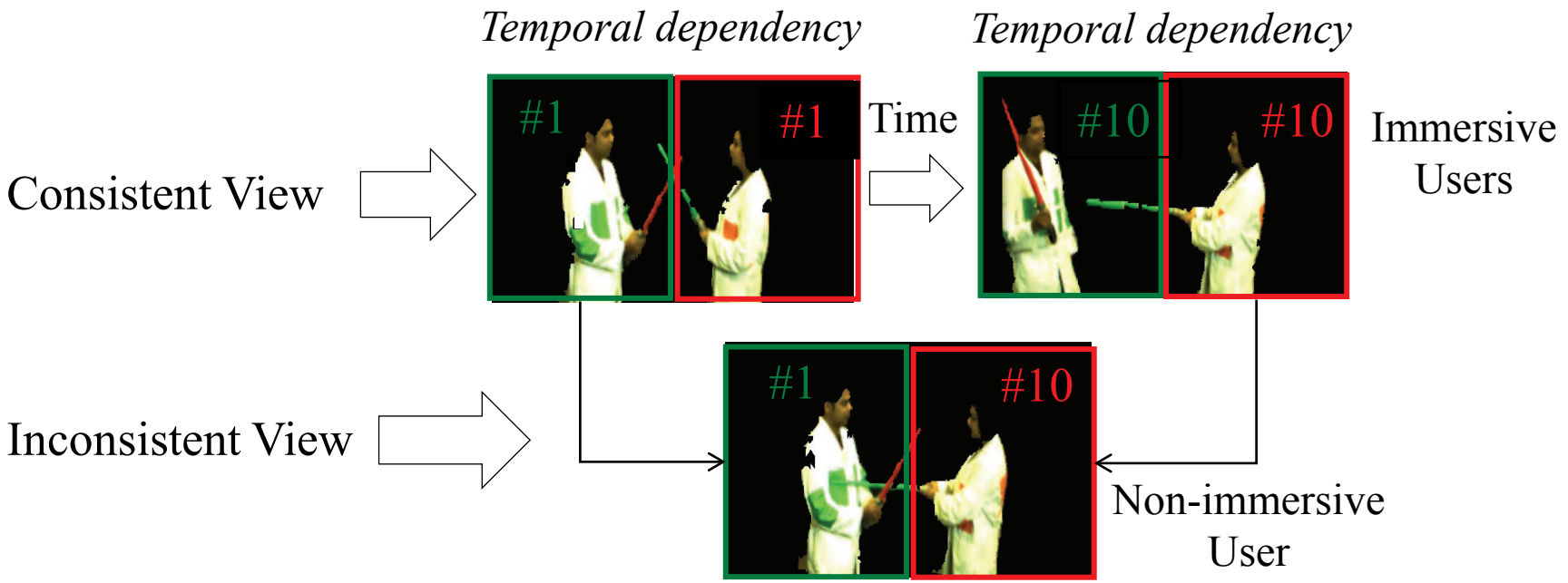




K. Nahrstedt et al. “Hourglass Multimedia Content and Service Framework for Smart Rooms”, 2005

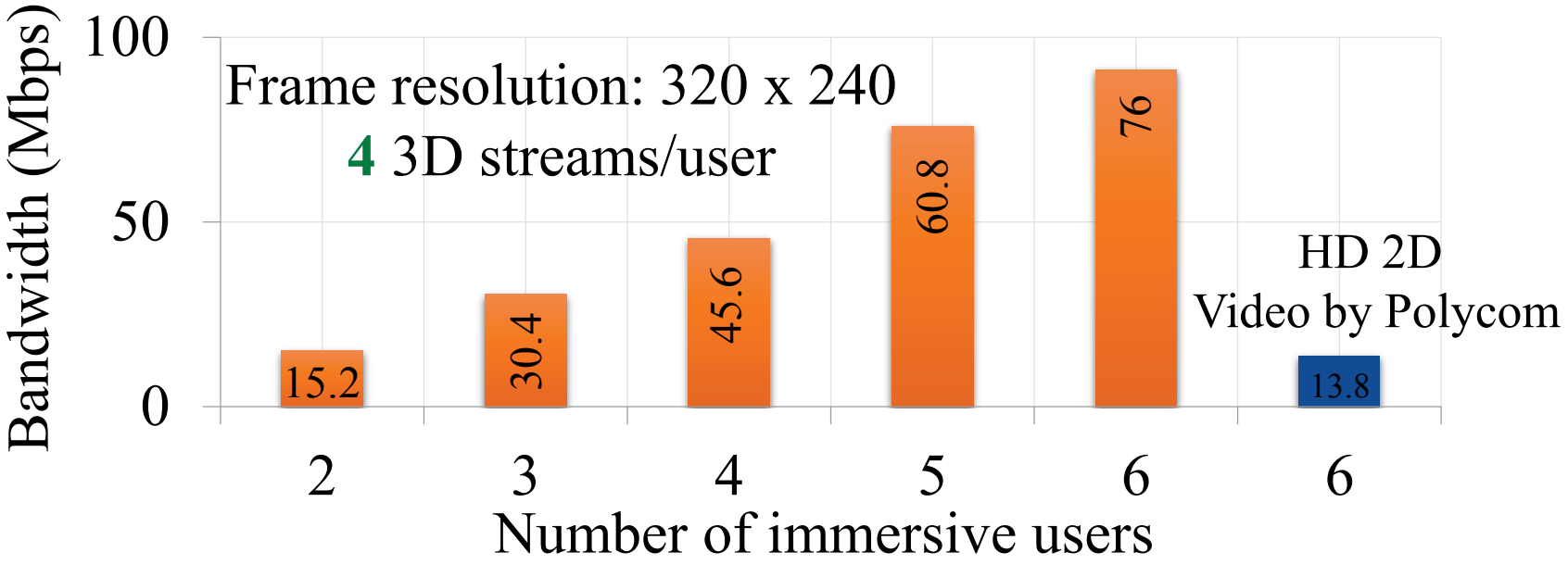
Experiential System: Challenges

- Heterogeneous users
- Multi-stream and multi-view environment
- **Correlated multi-stream dependency**



Experiential Systems: Challenge

- Heterogeneous users
- Multi-stream and multi-view environment
- Correlated multi-stream dependency
- **High demand on network resource**



Experiential System: Challenge

- Heterogeneous users
- Multi-stream and multi-view environment
- Correlated multi-stream dependency
- High demand on network resource
- **Activity dependent user expectations**

Immersive Users

- Stream selection
- Multi-stream skew
- End-to-end delay
- Quality of streams

Final Notes for 2025

- **2D and 3D Cameras** are getting cheaper
  - Extensive usage in homes, offices, other public spaces
    - Enlarge viewing areas – smart projections
    - Help with viewing multi-modal information (gestures – Kinect; Google glasses)
- **Microphone arrays** will serve in multiple tasks (not only voice application)
- **Mobile and diverse multimedia sensors** (e.g., body sensors) will dominate
- **Neuroscience** will assist in multimedia experience
- **Olfactory and haptic senses** will be ubiquitous for movie watching
- **Petabit networks** will be deployed, releasing the pressure on networks