Networked Multimedia and Internet Video

Colin Perkins
“IP video will represent 80% of all traffic by 2019, up from 67% in 2014”

Source: Cisco Visual Networking Index, 2015
Phase 1: Experimental

- Initial experimentation – understanding the architecture

- Network voice protocol (NVP), 1977
- Internet Streaming Protocol (ST, ST-II)
- Real-time Transport Protocol (RTP), 1996

- Is packet-based networked multimedia feasible?
  - Scepticism from traditional telephony, TV, and video communities
  - Media packetisation, fragmentation, and reassembly – use of application level framing for robustness
  - Re-sequencing and timing recovery
  - Is sufficient quality-of-service possible without network support?
  - Self-identifying vs signalled media formats; signalling needed to negotiate media formats and parameters
  - Scalable media transport protocols – Mbone
  - Custom protocols → application-layer transport over UDP as IP matured
Phase 2a: Telephony and Conferencing

- Telephony convergence with IP
- SIP: the protocol that ate the IETF

- Standardisation on RTP-based media over UDP
  - Peer-to-peer media channel
  - Server-assisted NAT traversal as IP addressing crunch hit

- Innovation in signalling vs legacy systems
  - Separate signalling channel and media path
  - Session and media description formats
  - Multi-stage negotiation vs offer-answer model
  - Signalling protocols vs APIs for defining signalling protocols

- Complexity – is a standard signalling protocol needed?
Phase 2b: Video on Demand

- Proprietary streaming protocols and applications
  - RealAudio, QuickTime, Flash
  - Standardisation: RealAudio → RTSP

- Resource constraints suggested media flows needed special treatment
  - Dedicated media server – packetisation complexity, custom protocols
  - Dedicated client applications – complex codec, error resilience
  - Signalling over TCP, RTP-over-UDP for media
  - Media and signalling multiplexed in single flow
  - Custom, but increasingly HTTP-influenced, protocols

- Business model disrupted by growth of web CDNs
Phase 3: IPTV

- IP multicast video vs broadcast TV industry
- Allowed ISPs to replicate a cable TV service …just as it became obsolete?

- IP multicast provides scalable distribution channel for traditional TV content
- Combine protocol mechanisms prototyped in Mbone tools with broadcast TV
- A multicast group per TV channel

Phase 4: Web Video

- HTTP-based adaptive streaming – web infrastructure now performant enough
  - Apple HTTP live streaming (HLS)
  - Microsoft Smooth Streaming
  - Adobe HTTP dynamic streaming

- Converging on MPEG DASH standard
  - Web CDNs, cloud computing, and cheap storage
  - Multiple representations of the content
  - Adaptation decisions at the client; real-time HTTP GET requests
Reflection

- Benefit of common media transport
- Impact of legacy systems
- Impact of Moore’s law on protocols and codecs
Benefits of Common Media Transport

- RTP media transport widely favoured:
  - Conferencing and telephony
  - Streaming using RTSP
  - Many IPTV deployments
- In principle, can gateway between protocols without touching media packets

- Will MPEG DASH play the same role for streaming?

- Common signalling standards elusive
  - Complexity of the problem space, non-invented-here, or…?
Impact of Legacy

- SIP failed because it cared about the PSTN
  - IP telephony standards aimed for parity with legacy standards
  - Feature and bug compatibility with the PSTN → overwhelming complexity
    - Early media (e.g., ring tones, 1-800-go-FedEx), DTMF and fax-over-VoIP, …
  - Security, identity, and PSTN gateways
    - Wiretapping regulations
    - Interworking with other systems
    - SDP security descriptions → no end-to-end security, no authentication, trusted middleboxes

- Legacy issues recurring with WebRTC
Impact of Moore's Law (1)

- Custom protocols → web protocols
  - Early systems required highly optimised protocols running over UDP
  - Increasingly superseded by generic web protocols
  - Hard problems become easy over time

- TCP is not suitable for real-time?
  - TCP congestion response fills in-network queues; induces latency but ensures consistent throughput to receiver
  - Increasing network capacity allows sustained rate sufficient for TV streaming

- Latency vs sustained bandwidth vs expected media rate variation: increasing range where trade-off allows TCP
Impact of Moore's Law (2)

- Why are we still arguing about codecs?
  - H.264 vs VP8 → H.265 vs VP9
  - Patents and licensing fees

- Performance differences negligible between codecs of same generation
  - 2x performance changes across generations → 10 years
  - Dwarfed by increase in network, storage, and processing capacity

- Excessive codec support is harmful
  - Implementations are complex: signalling, parsing, and decoding
  - Low-level programming for performance → security issues
  - Surely we can build a type safe byte code for downloadable codecs?
Current Challenges

• Political and engineering issues:
  • Congestion control
  • Network neutrality
  • Media quality and codecs

• Hard problems, but in the long term will likely be resolved by Moore’s law
Security

• Current systems not end-to-end secure
  • Signalling metadata, including media keys, exposed to servers, gateways, and media processing middleboxes
  • Required by wiretapping regulations

• Open issues for enabling strong security
  • Legal issues
  • Media path security via DTLS-SRTP, but relies on certificate authority
  • Identity provision – who are you calling?
  • Gateways to legacy systems with weaker security models
  • Use of in-network processing for conference servers
Balkanisation

- Fragmentation of conferencing and telephony application space
  - Differences in identity provision – is federated identity management feasible?
  - Differences in signalling protocols and/or use of WebRTC signalling building blocks – how to translate protocols?
  - Can we maintain media path compatibility?

- How to maintain uniform TV viewing experience across content providers with smart TVs?
  - Infrastructure quality and support issues – security and longevity

- Fragmentation of the network – NAT hurts interactive video
  - ICE helps ensure happy eyeballs
  - But, latency penalty for call setup
Ossification

- Lower-layer innovation increasingly difficult
  - NATs, firewalls, proxies, caches, and other middleboxes galore
  - Each understands (a subset of) current protocols

- Network assumes HTTPS over TCP/IP

- Such ossification is essential – implies backwards compatibility – but hinders the deployment of new features

- Promising future directions for video – content centric networking – require changes to these lower layers for full effect

Futures

- Increasing video quality and ubiquity, telepresence, virtual reality, immersive environments
  - Bigger, better, faster, more... fits within the existing architecture
  - Expect interesting congestion control/network neutrality/bandwidth crunch for the next 5-10 years, especially in mobile

- Architecture evolves to support a content-centric model with pervasive caching but zero visibility
  - Hints with MPEG DASH and CDNs today, but doesn’t go far enough for a world of ubiquitous multimedia content – have router vendors internalised Moore’s law for storage?

- Privacy concerns will push processing to the edges
  - Cannot trust servers to mediate