Ollscoil na Gaillimhe
University of Galway

CT 420
Real-Time Systems

Soft RTS

Dr. Jawad Manzoor
Assistant Professor
School of Computer Science

Contents

Ollscoil na Gaillimhe
University of Galway

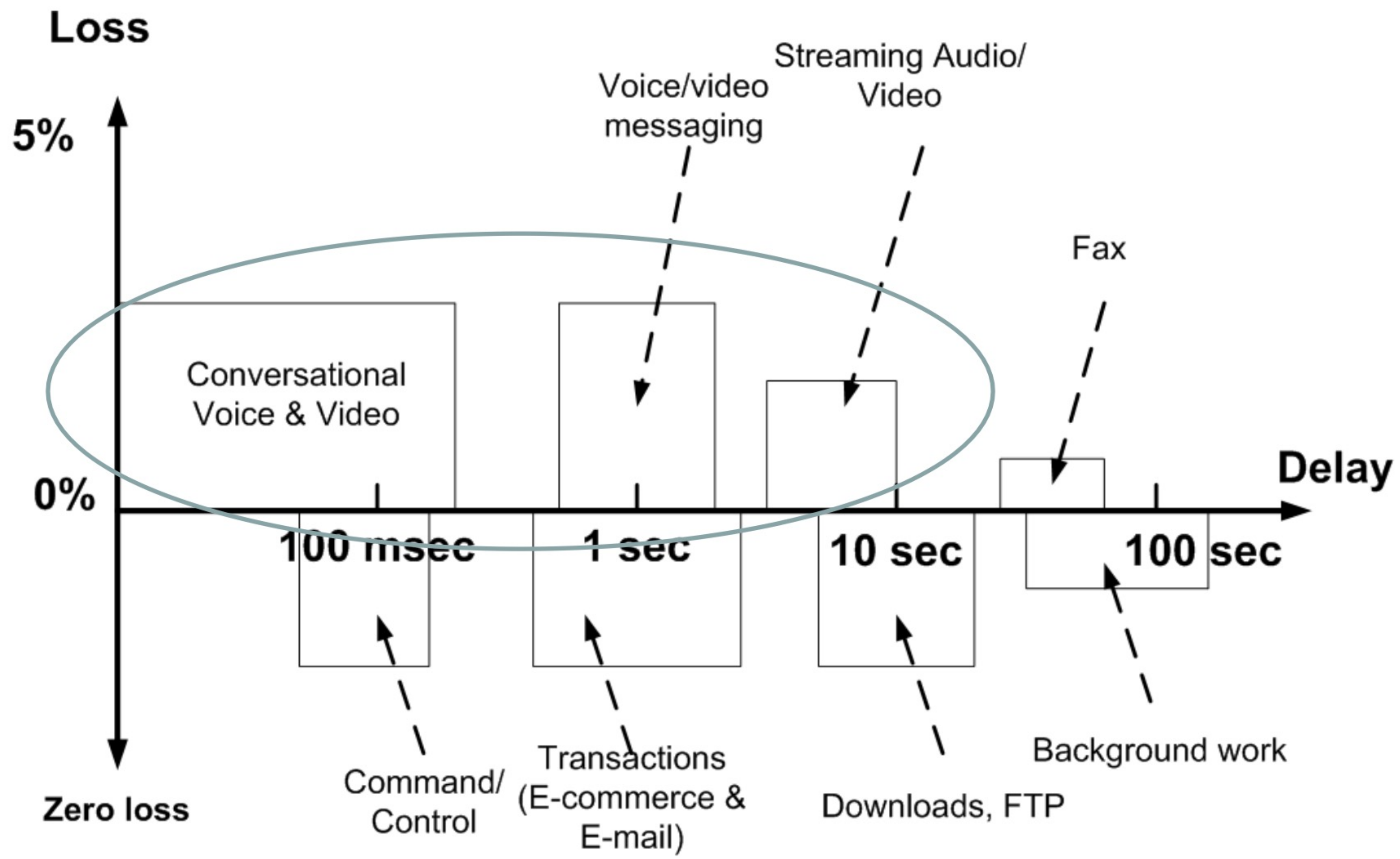
- Soft real-time systems
- QoS
- Real-time Multimedia Technologies
    - Real-Time Streaming
    - VoIP (Voice over IP)
    - Real-Time Gaming

Soft RTS

- A soft RTS is one in which performance is degraded but not destroyed by failure to meet response time constraints (Laplante)
- Example -> Multimedia systems
  - We consider multimedia device & infrastructure as soft RTS
  - E.g. VoIP
  - Here the missing of deadlines is not a safety issue, but a user experience / Quality of Service (QoS) issue

Application Dependent Requirements

Ollscoil na Gaillimhe
University of Galway



Source: G.1010 : End-user multimedia QoS categories - ITU

QoS

- Perceived versus intrinsic QoS

- **Perceived QoS** reflects the subjective evaluation of service quality from the end user's perspective and overall satisfaction.
- Examples: Perceived QoS encompasses factors like application responsiveness, ease of use, consistency, reliability, and the overall user interface design.
- Measurement: Perceived QoS is more challenging to measure accurately since it involves subjective experiences and user feedback, often gathered through surveys, user studies, or feedback mechanisms.

QoS

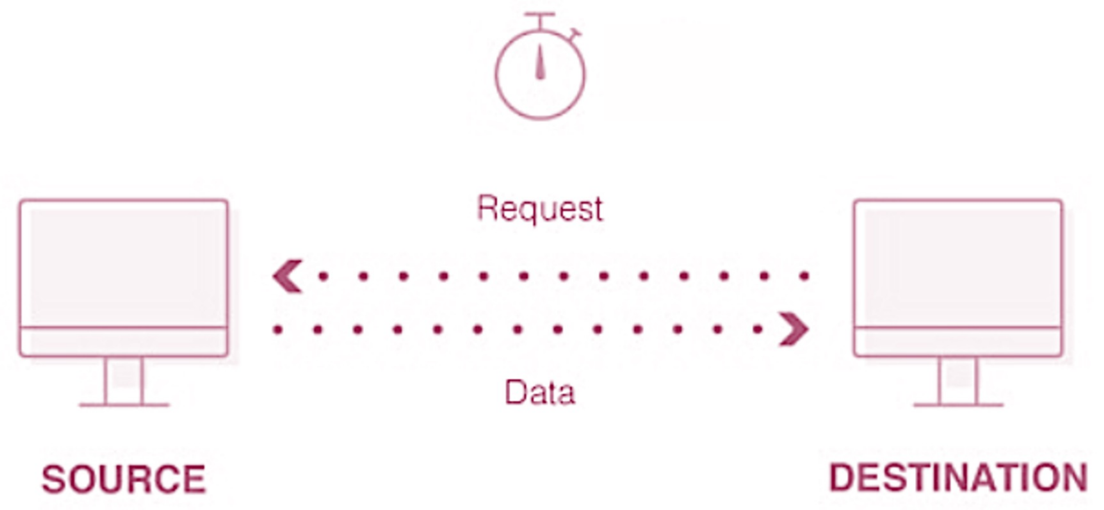
- Perceived versus intrinsic QoS

- **Intrinsic QoS** refers to the technical and measurable characteristics of a network or service that directly affect its performance and reliability.
- Examples: Intrinsic QoS parameters include bandwidth, latency, packet loss rate, jitter, and throughput.
- Measurement: Intrinsic QoS metrics are typically quantifiable and can be objectively measured using various tools and monitoring techniques.

QoS Metrics

Latency

- The time it takes to send a packet from point A (say, the client) to point B (the server).
- It is physically limited by how fast signals can travel in wires or in the open air.
- Latency depends on the physical, real-world distance between A and B.
- Typical latencies are conceptually small, between roughly 10 and 200 milliseconds.
- High latency can result in delays between user actions and system responses, leading to sluggish or unresponsive behavior in real-time applications.



QoS Metrics

Jitter

- It is the inconsistency or fluctuations in the arrival time of data packets at the receiver
- It can be caused by various factors such as network congestion, packet loss, routing changes etc.
- It can have significant implications for real-time communication applications, particularly voice and video streaming.
- Inconsistent packet arrival times can lead to disruptions, distortion, and out-of-sync audio or video playback.

QoS Metrics

Bandwidth

- Bandwidth measures the amount of data that is able to pass through a network at a given time.
- It is measured in bits per second (bps), such as megabits per second (Mbps) and gigabits per second (Gbps).
- Real-time applications with high bandwidth requirements, such as high-definition video streaming and VoIP, may experience performance issues if the available bandwidth is insufficient to accommodate the data transmission demands.

Ollscoil na Gaillimhe
University of Galway

Real-time Multimedia Technologies

Multimedia application

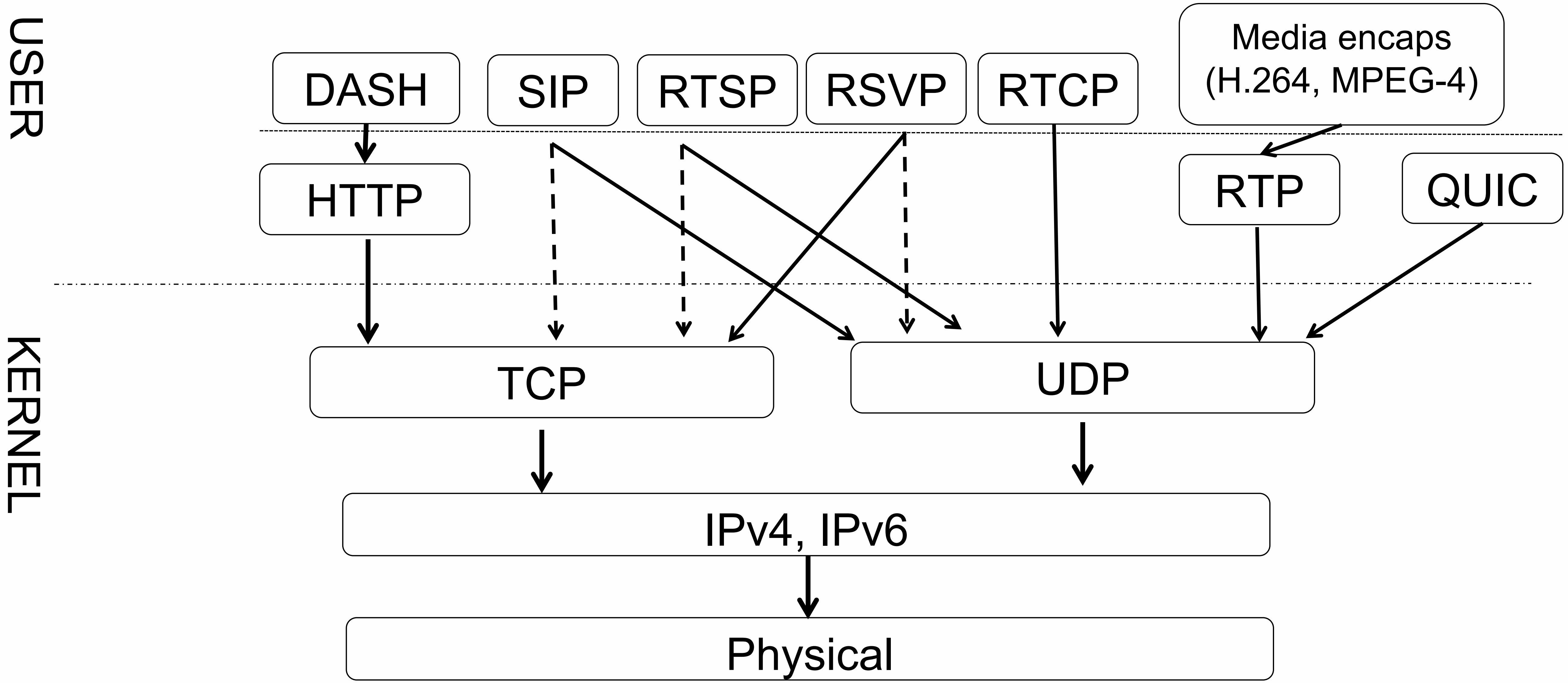
- Communication
  - WhatsApp
  - Zoom
  - MS Teams
- Streaming services
  - YouTube
  - Netflix
  - Twitch
- Cloud Gaming
  - Nvidia GeForce Now
  - Xbox Cloud Gaming
  - Amazon Luna

Transport layer for real-time media

- TCP
  - Provides reliability, ordered delivery, and congestion control
  - Retransmissions can lead to high delay and cause delay jitter
  - Not suitable for real-time
  - Does not support multicast

- UDP
  - No built-in reliability or congestion control
  - No defined technique for synchronizing
  - Low latency and minimal overhead (no handshake, no retransmissions)
  - A feedback channel must be defined for quality control

Internet Multimedia Protocol Stack



Service Requirements for Real-time Flows (Voice/Video)

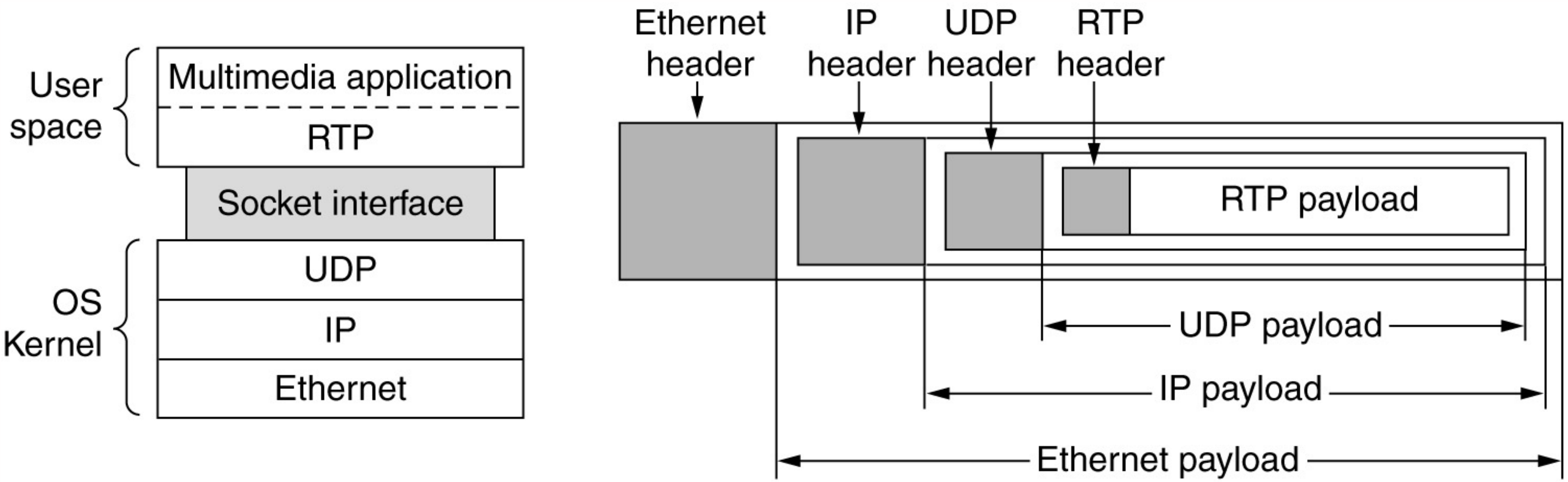
OLLSCOIL NA GAILLIMHE
UNIVERSITY OF GALWAY

- **Sequencing** - the process of maintaining the correct order of data packets during transmission and ensuring that they are reassembled correctly at the receiver's end.
- **Synchronization** - ensures that different types of data streams (such as audio and video) are aligned in time during playback.
- **Payload identification** - Different media types (MPEG1, MPEG2, H.261) may require different handling in terms of decoding or processing.
- **Frame indication** - specifying which packets belong to the same frame or video sequence and helps in decoding and rendering video frames accurately.

RTP

- Real-time Transport Protocol (RTP) provides end-to-end transport functions suitable for real-time audio/video applications over multicast and unicast network services
- It works in user space over UDP
- Working model:
  - The multimedia application generates multiple streams (audio, video, etc) that are fed into the RTP library
  - The library multiplexes the streams and encodes them in RTP packets which are fed to a UDP socket
- Secure RTP (SRTP) is used by applications including WhatsApp, Zoom, Skype etc for transporting voice and video streams

RTP

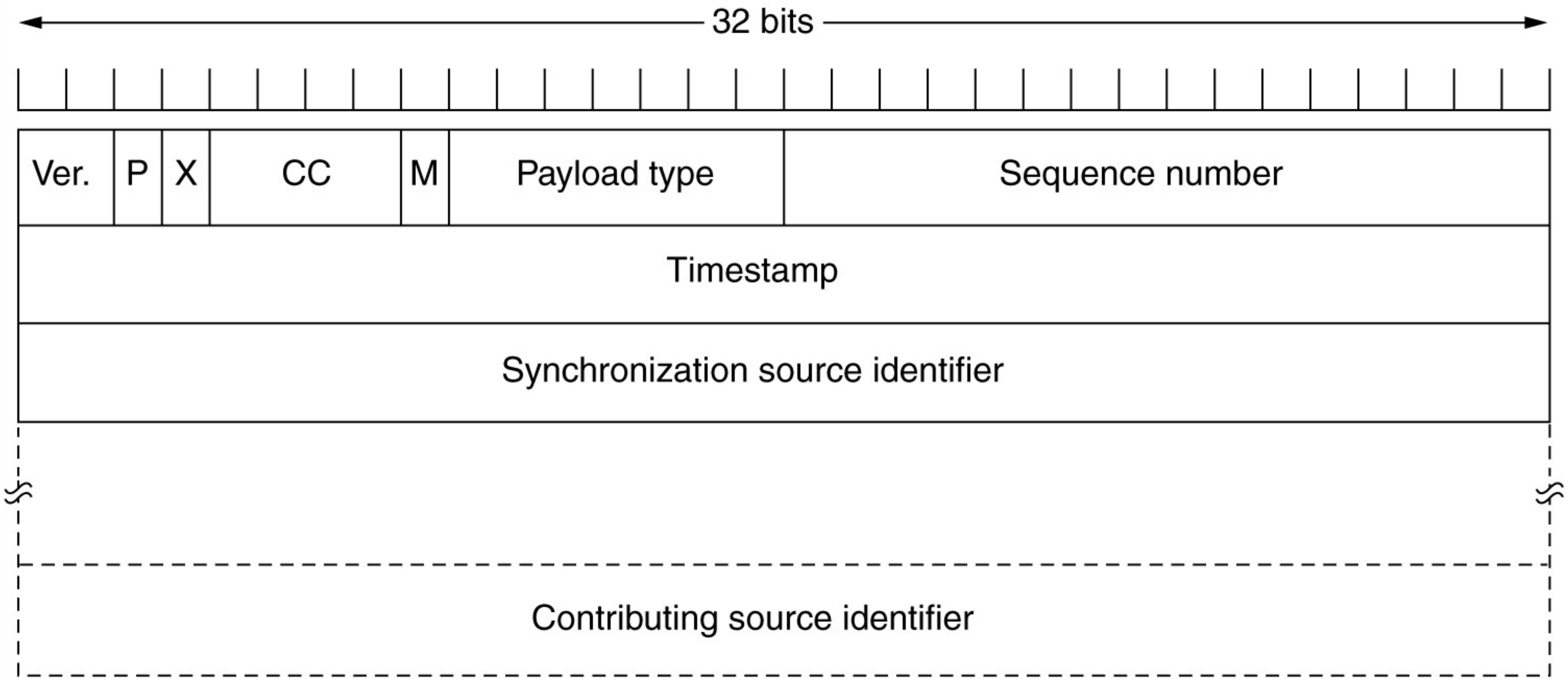


(a) Protocol Stack (b) Packet Nesting

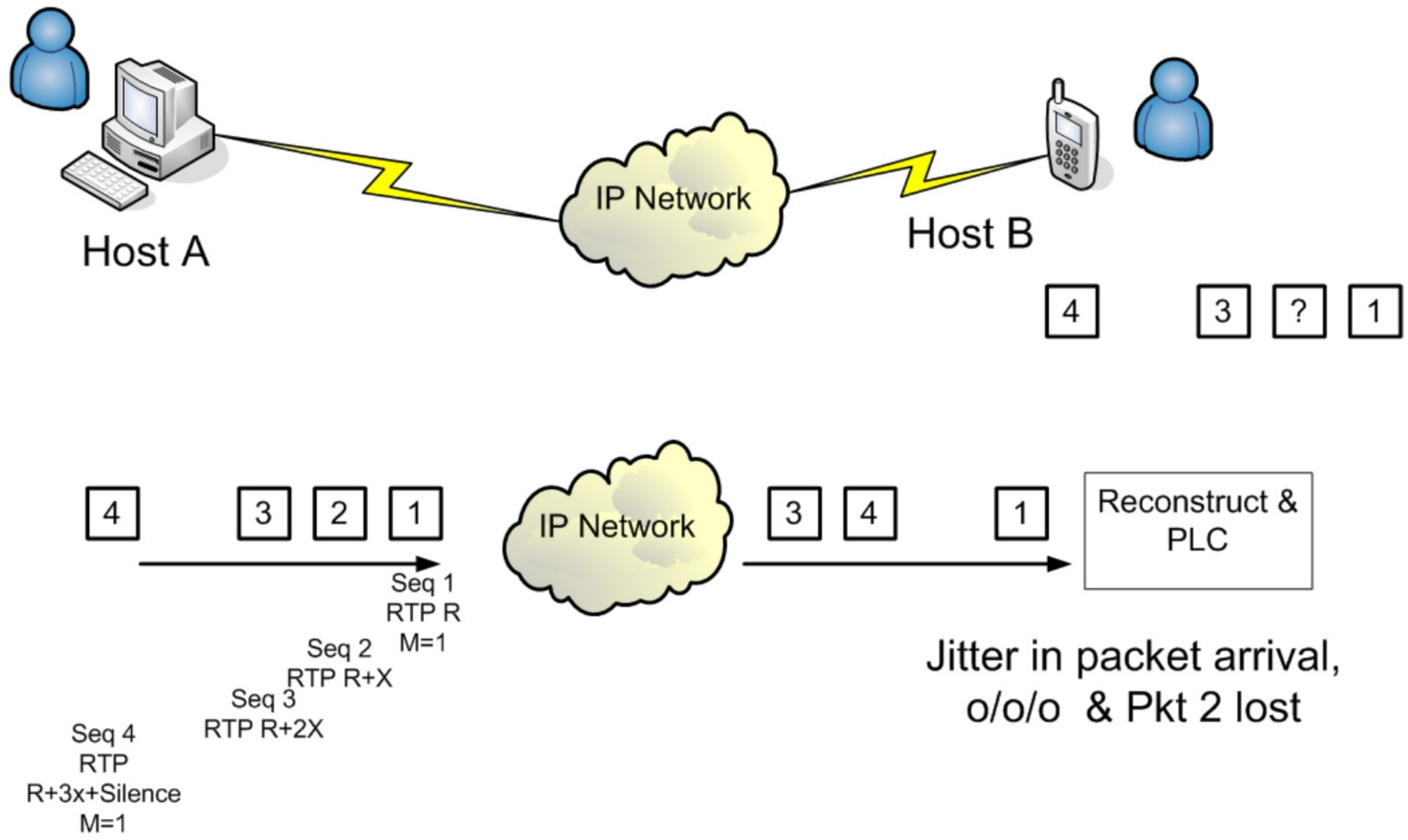
RTP Services

- Payload Type Identification
  - Determination of media encoding
- Synchronization Source identification
  - Assigned to each distinct media source (such as a microphone or a camera). Enables synchronization of multiple streams coming from the same source (e.g., lip-syncing audio and video).
- Sequence numbering
  - A counter is used that increments on each RTP packet sent; it is used to detect lost packets
- Time-stamping
  - Time monitoring, synchronization, jitter calculation
- RTP issues
  - No QoS guarantees
  - No guarantee of packet delivery

RTP Header


32 bits
Ver.
P
X
CC
M
Payload type
Sequence number
Timestamp
Synchronization source identifier
Contributing source identifier

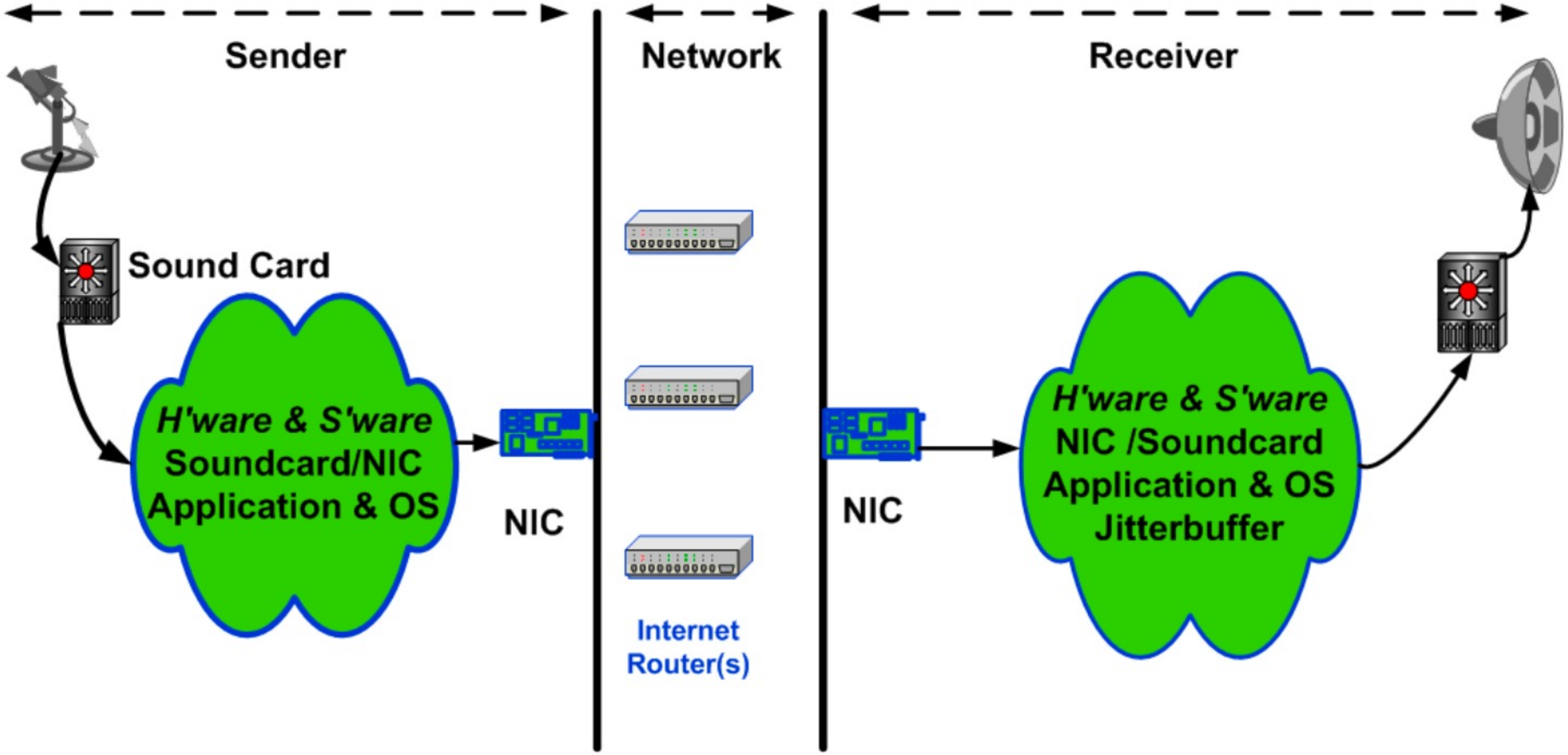
RTP Data Delivery



RTCP

- RTP Control Protocol (RTCP) is a companion control protocol to RTP
- Used periodically to transmit control packets to participants in a streaming multimedia session.
- Gathers statistics on media connection
  - Bytes sent, packets sent, lost packets, jitter, feedback and round trip delay.
- Provides feedback on the quality of service being provided by RTP but does not actually transport any data.
- Application may use this information to increase the quality of service, perhaps by limiting flow or using a different codec.

Mouth-to-Ear (M2E) delay



M2E Delay

- Sender
  - – Packetisation delay
  - – Encoding delays
  - – OS/Application/Driver software
  - – MAC delays
- Network
  - – Propagation delay
  - – Queuing delays
  - – Processing/serialisation delays in routers
- Receiver
  - – NIC delays
  - – OS/Application/Driver software
  - – Jitter buffer delays
  - – Decoding delays

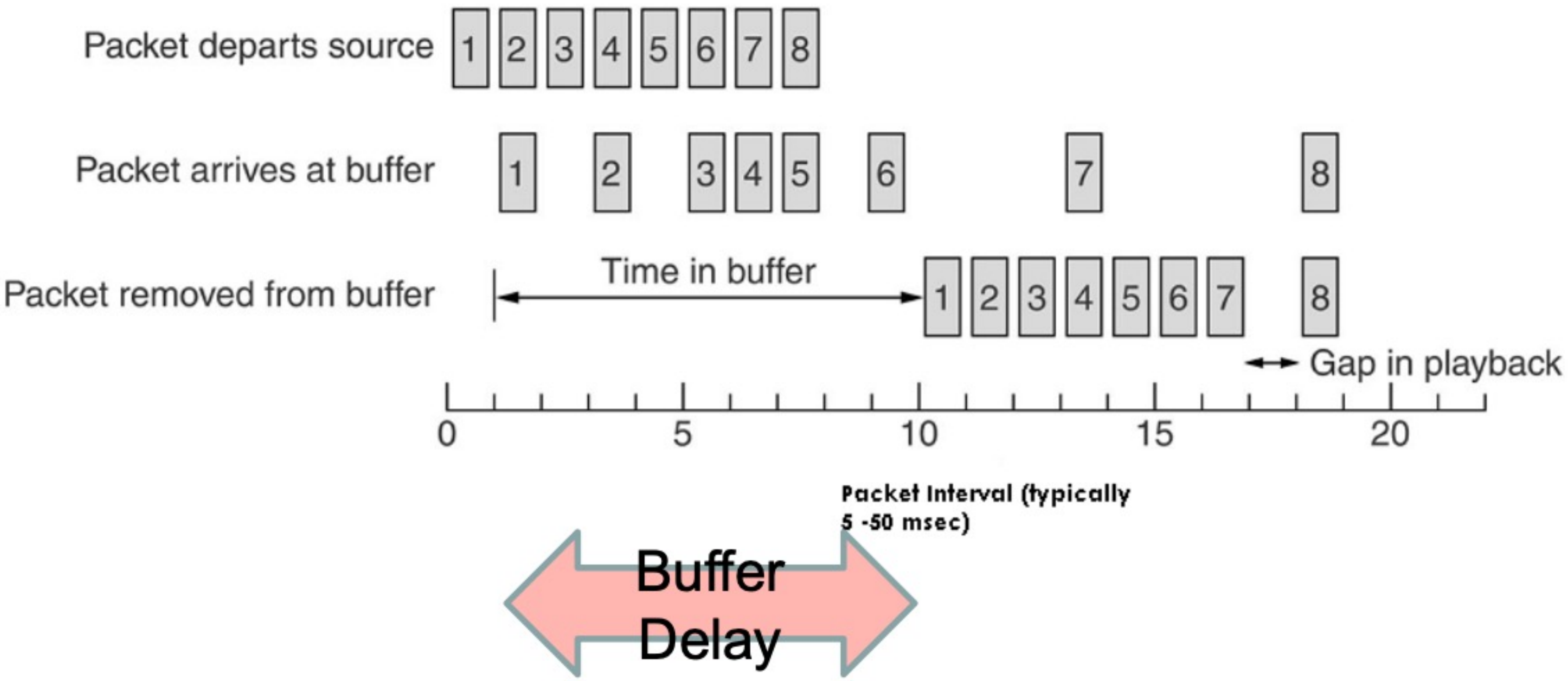
VoIP QoS Strategies

- Sender-based
  - RTCP feedback with adaptive codecs
    - If loss/delay excessive, switch to lower bandwidth codec?
    - Implement FEC – Forward Error Correction strategy
- Network-based
  - Prioritising delay sensitive traffic flows
- Receiver-based
  - Buffering strategies
    - Human ear NOT sensitive to short term variations
    - Buffer 'absorbs' variation in network queuing delays➔reconstruct voice using RTP timestamps
      - BUT.. adds to overall M2E delay➔trade-off
  - Packet Loss Concealment (PLC) and FEC

Packet Loss Strategies

- Use of UDP limits delays but can lead to packet loss
- Compensation Strategies at Sender and/or Receiver
  - Forward Error Correction (FEC)
    - Form of Information Redundancy
  - Packet Loss Concealment (PLC)
    - Silence: simplest
    - Repetition: repeat last packet
    - Interpolation

Receiver-based: Jitter Buffer Strategies



- Buffer Playout Delay adds to M2E delay
- Above strategy
  - – Pkt 8 arrives too late for playout
    - ➔ Drop the packet?
    - ➔ Increase size of buffer in response to increasing delays?

Receiver-based: Jitter Buffer Strategies

- Fixed buffer size: limitations
  - Too large➔Extends overall delay
  - Too small➔Additional late packet losses due to late arrival
- Adaptive buffer size
  - Adapt to network conditions
  - Per talkspurt (PT)
    - Operate by elongating/shortening inter talkspurt silence periods
    - Less noticeable
  - Per packet scaling (PPS)
    - Speed up/slow down speech
    - Skype/WebRTC

Network-based Strategies

- LAN Environment
  - Switched LANs typically QoS enabled
  - Fast/Gigabit Ethernet links rarely congested
- WAN Environment
  - Increase bandwidth
    - Costly and temporary solution?
  - Reservation policy & Traffic categorisation & prioritisation
    - Requires admission control policy

Cloud Gaming

Cloud Gaming

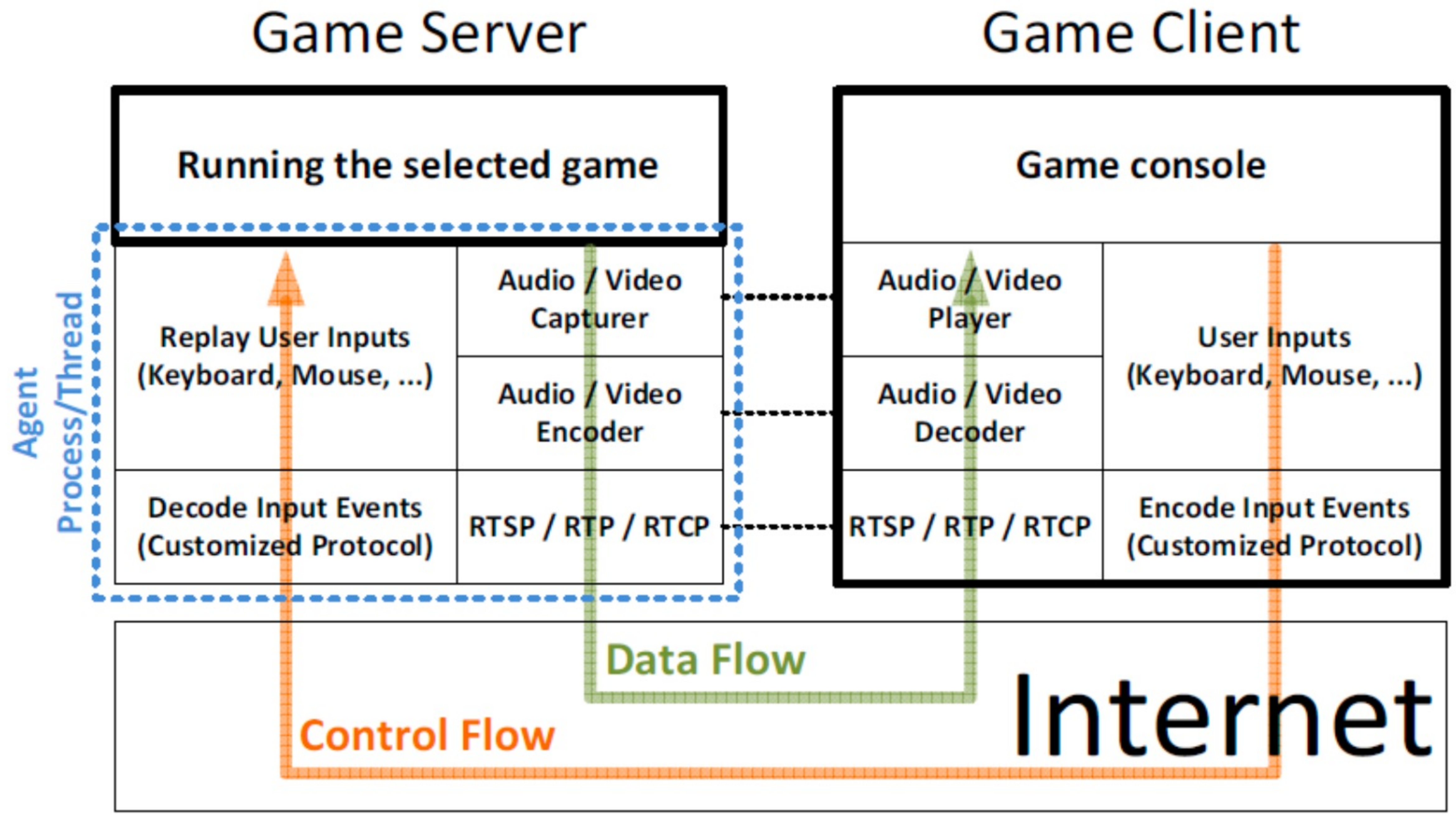
Ollscoil na Gaillimhe
University of Galway

- The game is installed and executed on a powerful remote server located in the cloud.
- The game rendering (processing graphics, physics, and game logic) is done on the server.
- Once the game is rendered on the server, the video frames and audio are compressed and streamed to the player's device via the internet.
- The player’s inputs, such as controller buttons, mouse movements, or keyboard presses, are sent back to the cloud server over the internet in real time.
- Used in Massively Multiplayer Online Game (MMOG)
- Server dealing with multiple players

Cloud Gaming

- Cloud gaming solves several issues of the traditional approach
  - Tedious installation and patching process
  - Expensive hardware and GPUs
- RTC application
  - need to minimise delays (lag)
  - need time sync to measure delays
- Simultaneity – all players see same game representation

Cloud Gaming



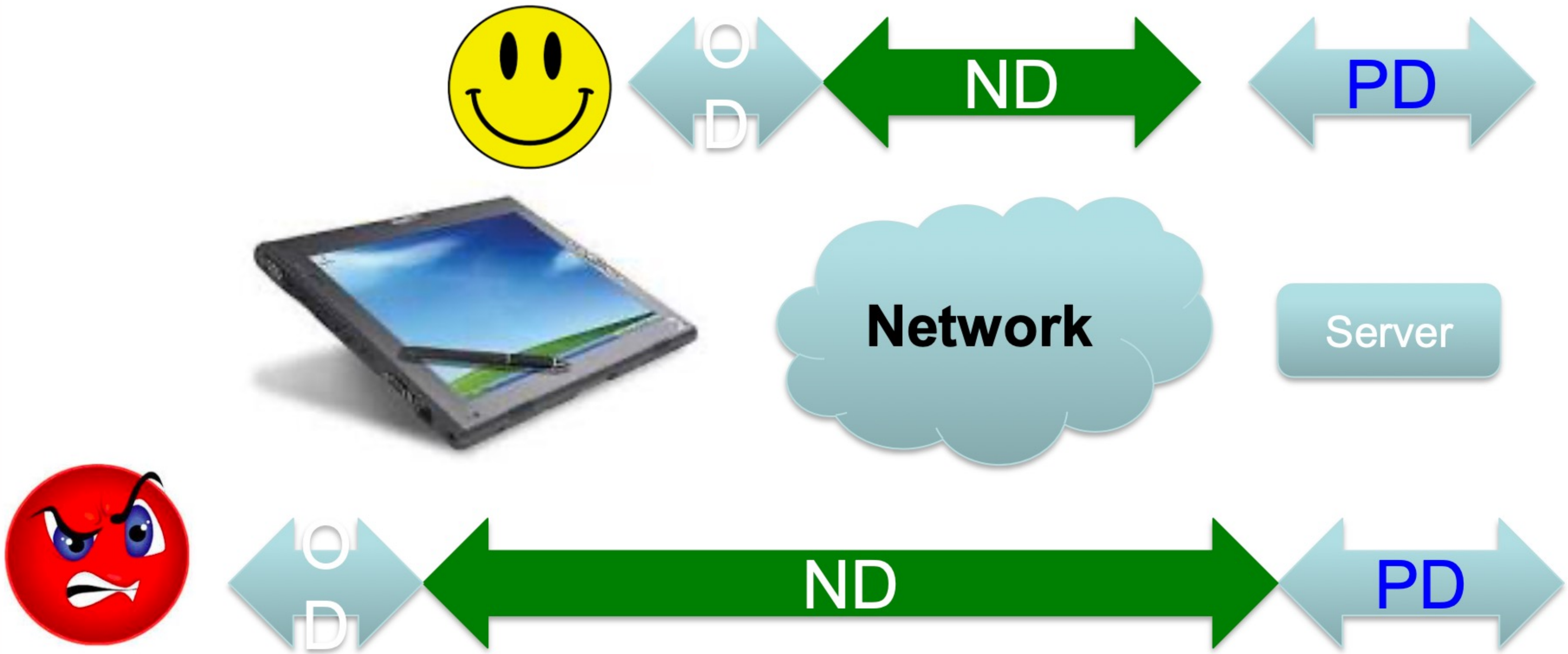
Client/Server architecture of GamingAnywhere
http://gaminganywhere.org/

Response Delay

- Response delay (RD)
  - time diff between a user submitting a command and the corresponding in-game action appearing on the screen
- Processing delay (PD)
  - time required for the server to receive/process a player's command, encode/ transmit the corresponding frame
- Playout delay (OD)
  - time required for the client to receive, decode, and render a frame on the display
- Network delay (ND)
  - Round Trip Delay

- RD = PD + OD + ND

Challenge: Lag & QoE



Ollscoil na Gaillimhe
University of Galway

WebRTC

WebRTC

OLLSCOIL NA GAILLIMHE
UNIVERSITY OF GALWAY

- WebRTC (Web Real-Time Communication) is a free, open-source project that provides web browsers and mobile applications with real-time communication (RTC) via simple APIs
- It allows audio and video communication to work inside web pages via direct peer-to-peer communication using JavaScript and HTML, i.e. eliminating the need to install plugins or download native apps
- Uses DTLS for encryption and SRTP for secure media streams
- Works on major browsers (Chrome, Firefox, Safari, Edge) and mobile platforms.
- Use case: Web version of Google Meet, Discord, Facebook Messenger, Peer-to-peer file sharing etc.

WebRTC Support

- Who supports WebRTC

- Combination of Groups

  http://www.w3.org/2011/04/webrtc-charter.html

  http://tools.ietf.org/wg/rtcweb/charters

  http://www.webrtc.org/

WebRTC working Model

- Media Capture
  - Accesses camera and microphone using browser APIs
- Connection Management
  - Exchanges session control messages (via WebSockets, SIP, or custom methods) to set up connections.
- Data Transmission
  - WebRTC uses the Interactive Connectivity Establishment (ICE) techniques to overcome the complexities of real-world networking like NAT
    - STUN: Finds public IP addresses.
    - TURN: Relays media if direct connections fail.
  - RTP/SRTP: Streams media, ensuring low-latency delivery

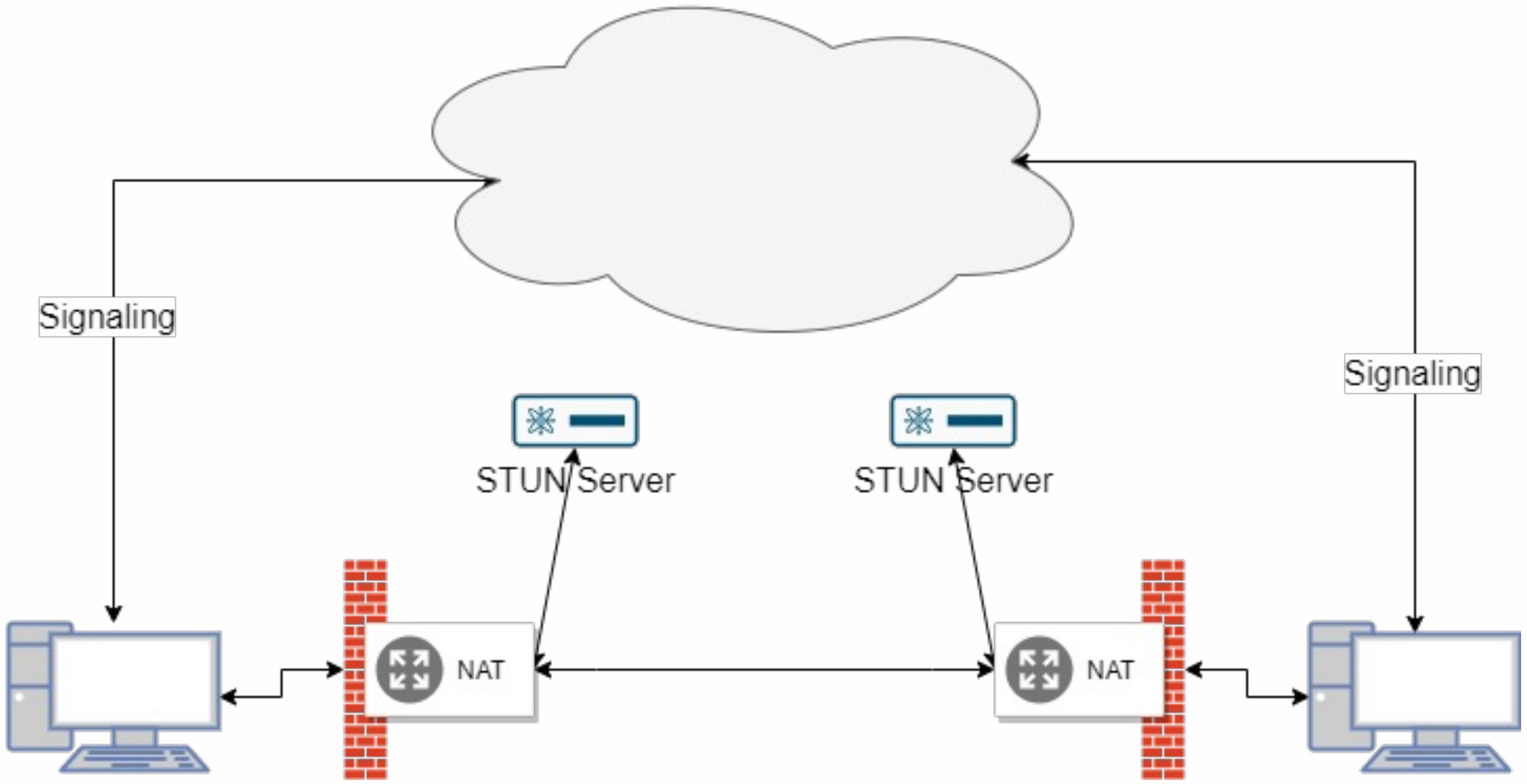
WebRTC working Model

OLLSCOIL NA GAILLIMHE
UNIVERSITY OF GALWAY

- Peer Discovery
  - In P2P communication, the peers need to be able to identify or locate each other over the wire
  - Peer discovery mechanisms are not defined by WebRTC, although the process can be as simple as sharing a URL that peers can use to communicate.
- ICE Techniques
  - ICE will first try to make a connection using the host address obtained from a device’s operating system.
  - If network if unsuccessful ICE will obtain an external address using the STUN server.
  - If that fails traffic is routed via a TURN relay server.

STUN

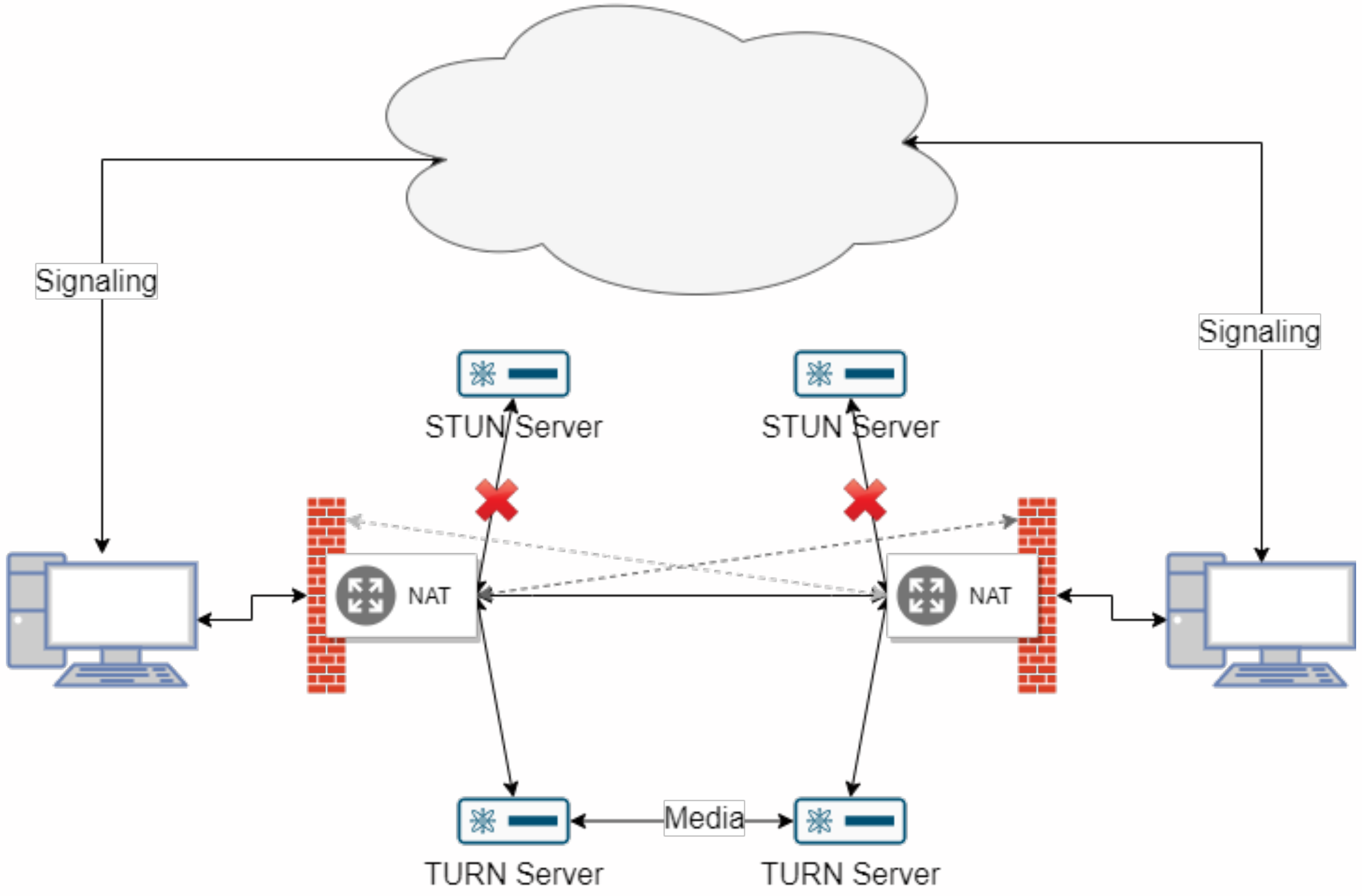
- STUN (Session Traversal Utilities for NAT) is a protocol that is used to discover public addresses and it determines any restrictions in your router that would prevent a direct connection with a peer.
- Clients receive their public addresses as requested from STUN servers.



https://medium.com/dvt-engineering/introduction-to-webrtc-cad0c6900b8e

TURN

- TURN (Traversal Using Relays around NAT) bypasses the Symmetric NAT restriction by opening a connection with a TURN server and relaying all information through that server.
- A connection is required with a TURN server which will tell all the peers to send packets to the server which will then be forwarded to the requester.

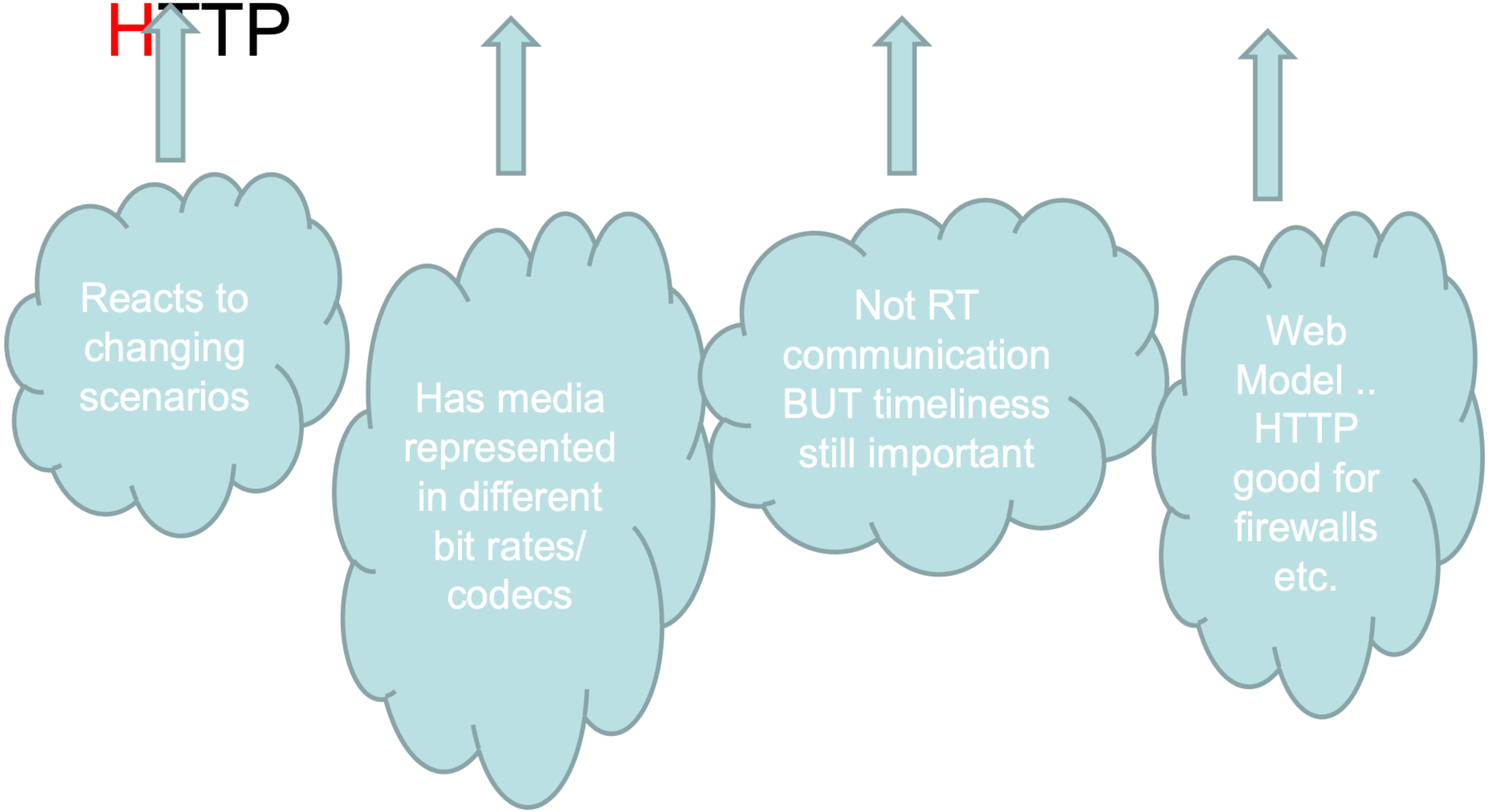


https://medium.com/dvt-engineering/introduction-to-webrtc-cad0c6900b8e

DASH

DASH

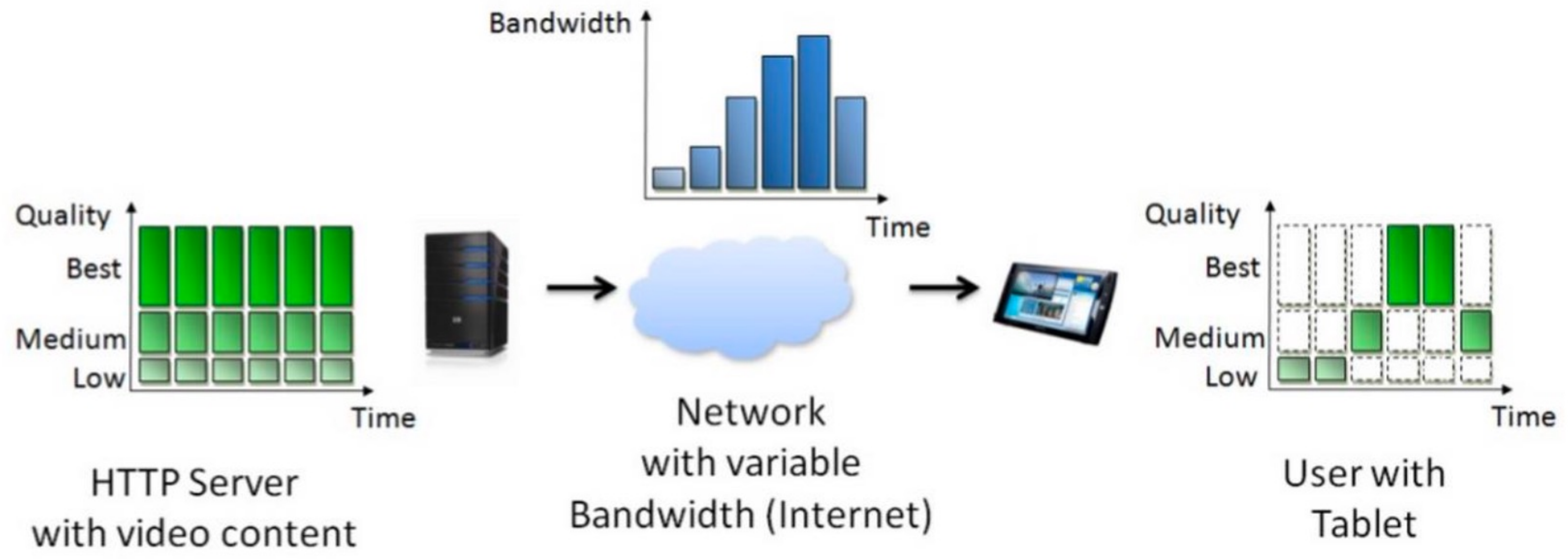
- Dynamic Adaptive Streaming over HTTP



DASH working model

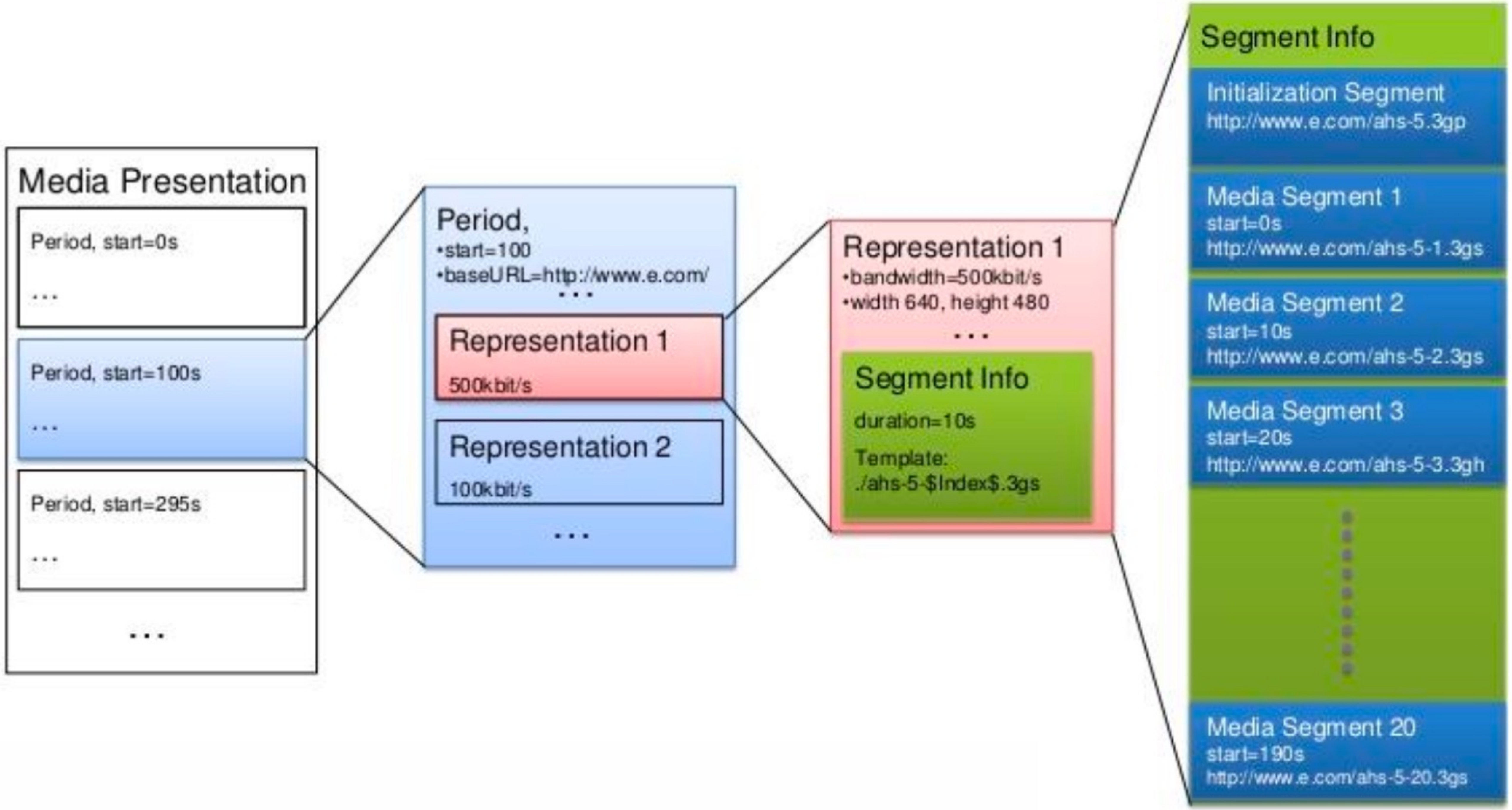
- Encoding and segmentation
  - The original video/audio content is divided into small segments (usually 2-10 seconds each).
  - Each segment is encoded at multiple bitrates and resolutions (e.g., 1080p, 720p, 480p) for adaptability.
  - A manifest file is created, containing metadata about the segments, their URLs, codecs, and timing.
- Delivery
  - Segments and the MPD file are uploaded to HTTP servers or CDNs (Content Delivery Networks).
  - The encoded video segments are pushed out to client devices over the Internet
- Playback
  - The client downloads the MPD file to understand the available content and quality options.
  - It chooses appropriate 'representation' based on network conditions, device capabilities and user preferences, decodes the chunks and plays back the video
- Quality Adjustment
  - The player continuously downloads and plays segments, adjusting the quality as network conditions change

DASH



Source: http://www.slideshare.net/christian.timmerer

DASH Data Model



Source: http://www.slideshare.net/christian.timmerer

Ollscoil na Gaillimhe
University of Galway

Thank you for your attention!

University
*of*Galway.ie