Streaming Games

Lesson I07

Dario Maggiorini (dario@di.unimi.it)
Online Game Design - A.A. 2021-2022

What is "Streaming"?

Main Entry: stream
Pronunciation: \ˈstrēm\
Function: noun

1 : a body of running water (as a river or brook) flowing on the earth; also : any body of flowing fluid (as water or gas)

2 a : a steady succession (as of words or events) <kept up an endless stream of chatter>
  b : a constantly renewed or steady supply <a stream of revenue> **c** : a continuous moving procession <a stream of traffic>

3 : an unbroken flow (as of gas or particles of matter)

4 : a ray of light

5 a : a prevailing attitude or group <has always run against the stream of current fashion>
  b : a dominant influence or line of development <the influence of two streams of inheritance: genetic and cultural — P. B. Baltes>

6 British : track

Source: Merriam-Webster

What is "Streaming"?

Main Entry: stream
Pronunciation: \ˈstrēm\
Function: noun

1 : a body of running water (as a river or brook) flowing on the earth; also : any body of flowing fluid (as water or gas)

2 a : a steady succession (as of words or events) <kept up an endless stream of chatter>
 b : a constantly renewed or steady supply <a stream of revenue> **c** : a continuous moving procession <a stream of traffic>

3 : an unbroken flow (as of gas or particles of matter)

4 : a ray of light

5 a : a prevailing attitude or group <has always run against the stream of current fashion>
 b : a dominant influence or line of development <the influence of two streams of inheritance: genetic and cultural — P. B. Baltes>

6 British : track

> We want a steady frame rate

> Data never ends, unlike a movie we could go on forever

> Without VCR functionalities, because **we are live** and, unlike a movie, we cannot pause, rewind, or fast forward

Source: Merriam-Webster

What is Game Streaming?

- To be honest, we already addressed game streaming
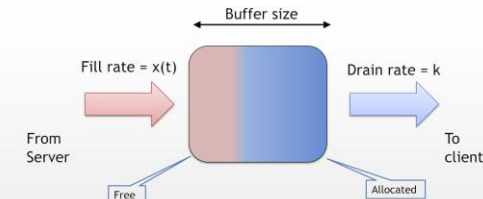

While talking about fast pacing data production
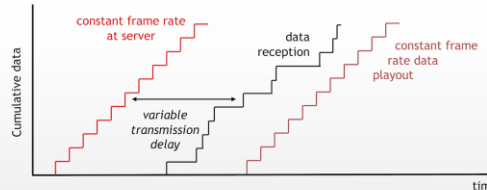
This is the same as video streaming you can have from Youtube or Netflix

... but gaming is much more than video streaming

.. and bounded delay

.. and uniform delivery

Game Streaming is More Than Video Streaming

- Why do we say that game streaming is more complicated?

- Because the transmission time must be bounded <u>on the round trip</u>!
  - In video streaming we need "just" a smoot playout
  - In game streaming we need a smoot playout of the result <u>within milliseconds from the input</u>

- **All modern access network technologies are designed to be optimized only in one direction!**

~100 ms
(depends on game)
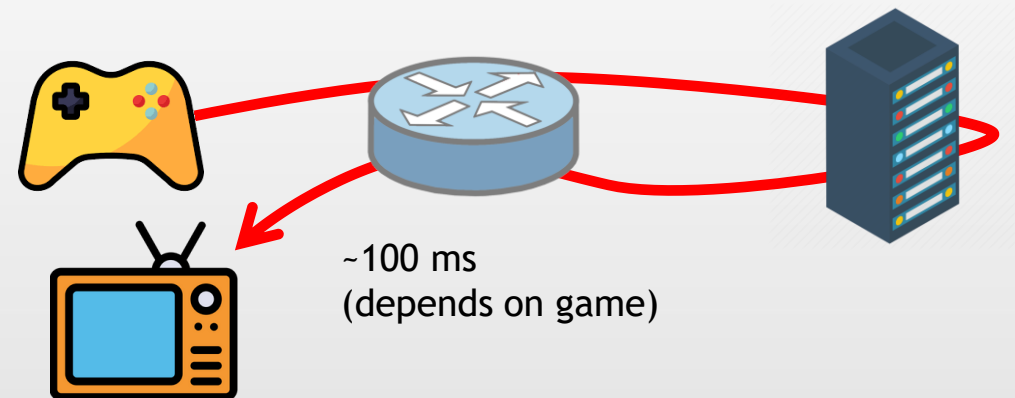
Data Streaming vs Content Streaming

- So far, in games, we considered only data streaming

- Content streaming is a completely different beast!

- With data streaming, you send information about how the client should use the assets
  - Clients are huge
  - Few bytes over the network are enough
    - Assets are already on the client

- With content streaming you send everything, because there are no assets on the client
  - Clients are small or even nonexistent
    - Like playing inside a web browser or a television
  - All assets must be provisioned with the stream
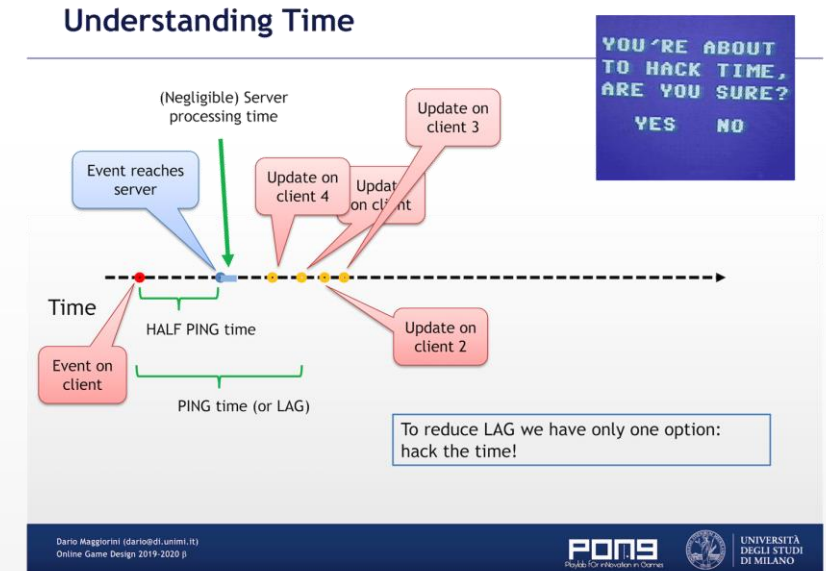
TV-Based Gaming is Coming Our Way

Samsung Announces The Samsung Gaming Hub at CES - A New Game Discovery Platform

January 02, 2022
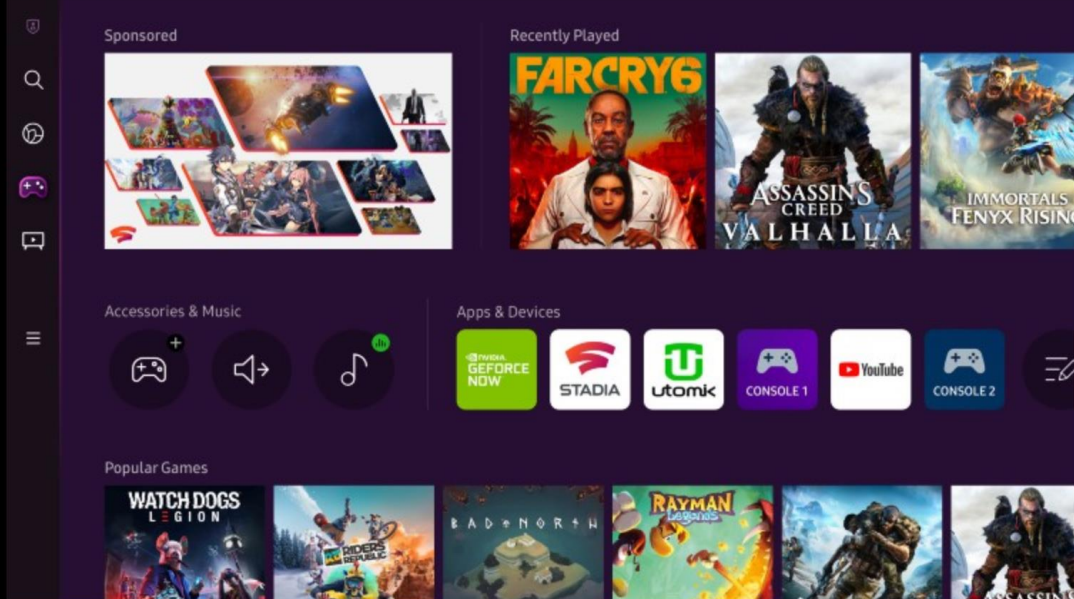
Your LG TV can now play Google Stadia if it's running webOS 5.0 or later

LG now has both Nvidia and Google cloud gaming

By Sean Hollister | @StarFire2258 | Dec 7, 2021, 7:00pm EST

Content Streaming

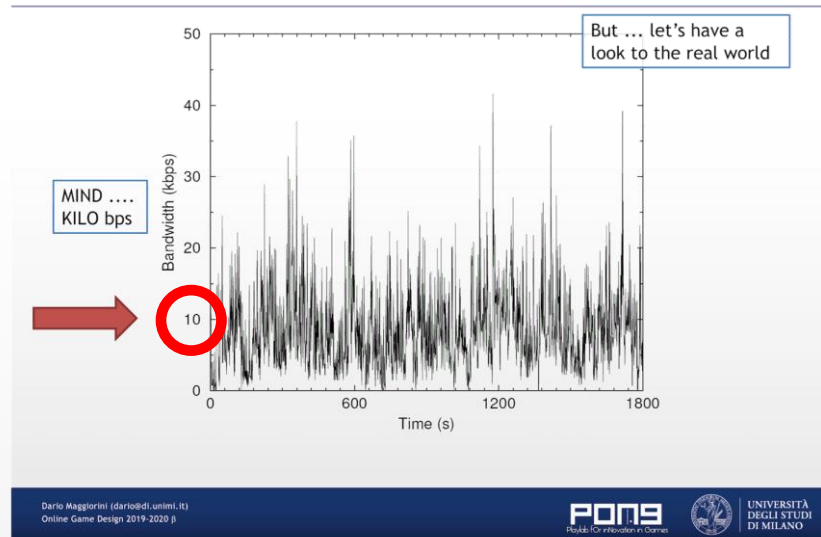- Content streaming follows exactly the same rules as data streaming

but …

- We need to use **a lot more bandwidth**, because we are also sending whatever is not preloaded on the client

  – Textures
  – Meshes
  – Shaders
  – Game logic

  – Music
  – Sounds
  – Speech
  – Animations

  – Video clips
  – A.I.
  – … and much more

Data Streaming vs Content Streaming

World of Warcraft Bandwidth Usage

But ... let's have a look to the real world

MIND .... KILO bps
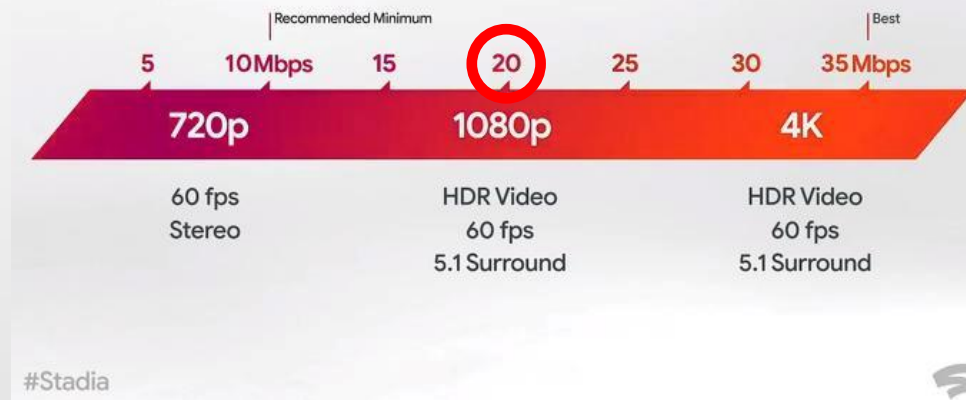
This is enough for World of Warcraft

10 Kbps vs 20 Mbps ... to achieve the same result

This is (barely) enough for Stadia

Stadia supports a range of resolutions and internet speeds

Recommended Minimum			Best
5 10 Mbps 15	20	25 30	35 Mbps
720p	1080p		4K
60 fps Stereo	HDR Video 60 fps 5.1 Surround		HDR Video 60 fps 5.1 Surround

#Stadia

Approaches to Content Streaming

- From a technical standpoint, today we have three approaches for content streaming

  1. Sending assets along the data

  – Sending the result of the actions of the players
    I.e., the screen that should be visualized
    And this can be achieved in two ways:

     2. Running the game inside a virtual machine
        and (video)streaming the screen to the user

     3. Defining a distributed execution environment
        on the cloud and running the game as a SaaS

Sending Assets

- A platform adopting this approach is Second Life
    - By Linden Lab, 2003

- The second life client is a dump application taking care of:
    - Authentication
    - Caching
    - Visualization
- Basically, it is a small program with a HUGE cash

- The surrounding environment will be requested on cache miss and displayer as soon as it will be available
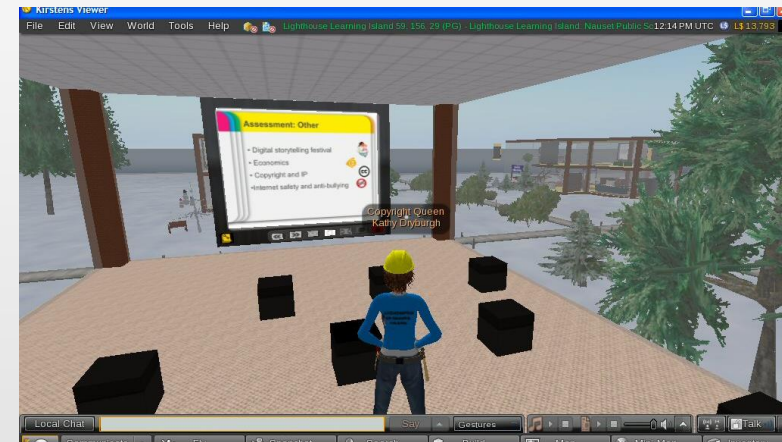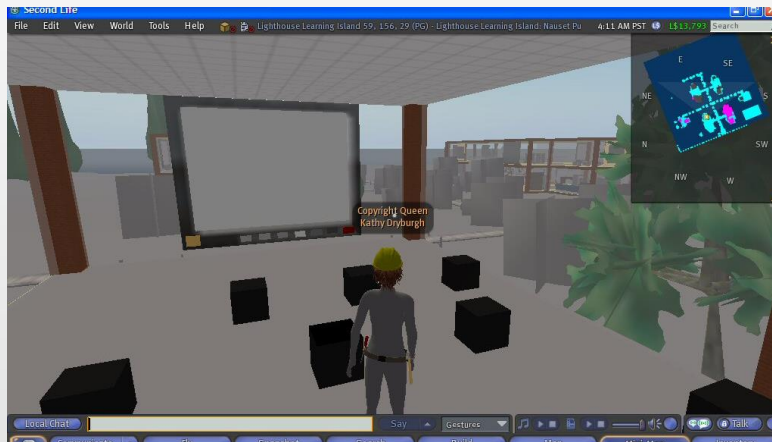
Sending Assets

- This approach is not very popular today, for many reasons
    1. The visual experience can be poor depending on network capacity and delay
    2. Hardware-dependent assets and cannot be provided "as is" but must be preprocessed (losing time)
        - Shaders are an example
          Everything looks "flat" in Second Life
    3. Continuous media streaming along assets needs to set up its own streaming channel
        - Video (we have a TV inside the virtual environment)
        - Audio (we have a concert or even just audio chat)

- Considering today's video compression standards, we can use the same resources for a video … and get much better visual results

Streaming the Screen

- It is very similar to a video streaming or a videoconference

but …

1. There is a negligible delay between generation and delivery
    - In television broadcasting you can have seconds of delay to accommodate "manual corrections" to the video

2. The quality is much higher than a standard videoconference
    - Try saying you are going to stream you webcam in native 4K in a call … and everybody will be asking what is wrong with you

UNIVERSITÀ
DEGLI STUDI
DI MILANO

Streaming an Emulated Machine

This approach has been adopted by Microsoft for its xCloud service

Internet

Player

With the additional requirements to have a "well specific" set-top-box or a windows PC

UNIVERSITÀ
DEGLI STUDI
DI MILANO
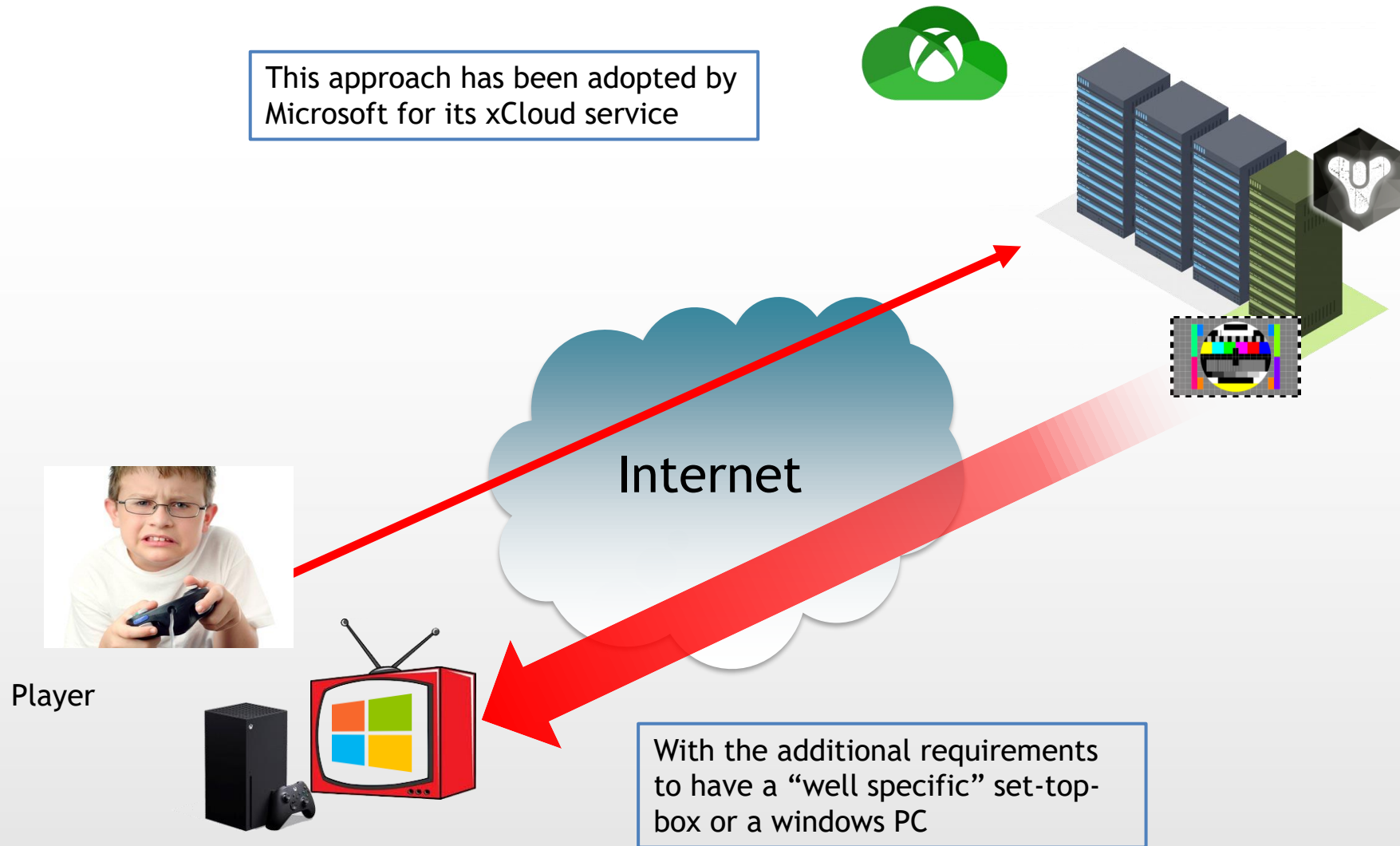
Advantages of Streaming an Emulated Machine

- ## The game software can be a retail version
  - The virtual machine is emulating a standard gaming PC
  - The game does not need to know it is in a datacenter

- ## Using a retail version will lower development costs because we can use the same code base as for the general public

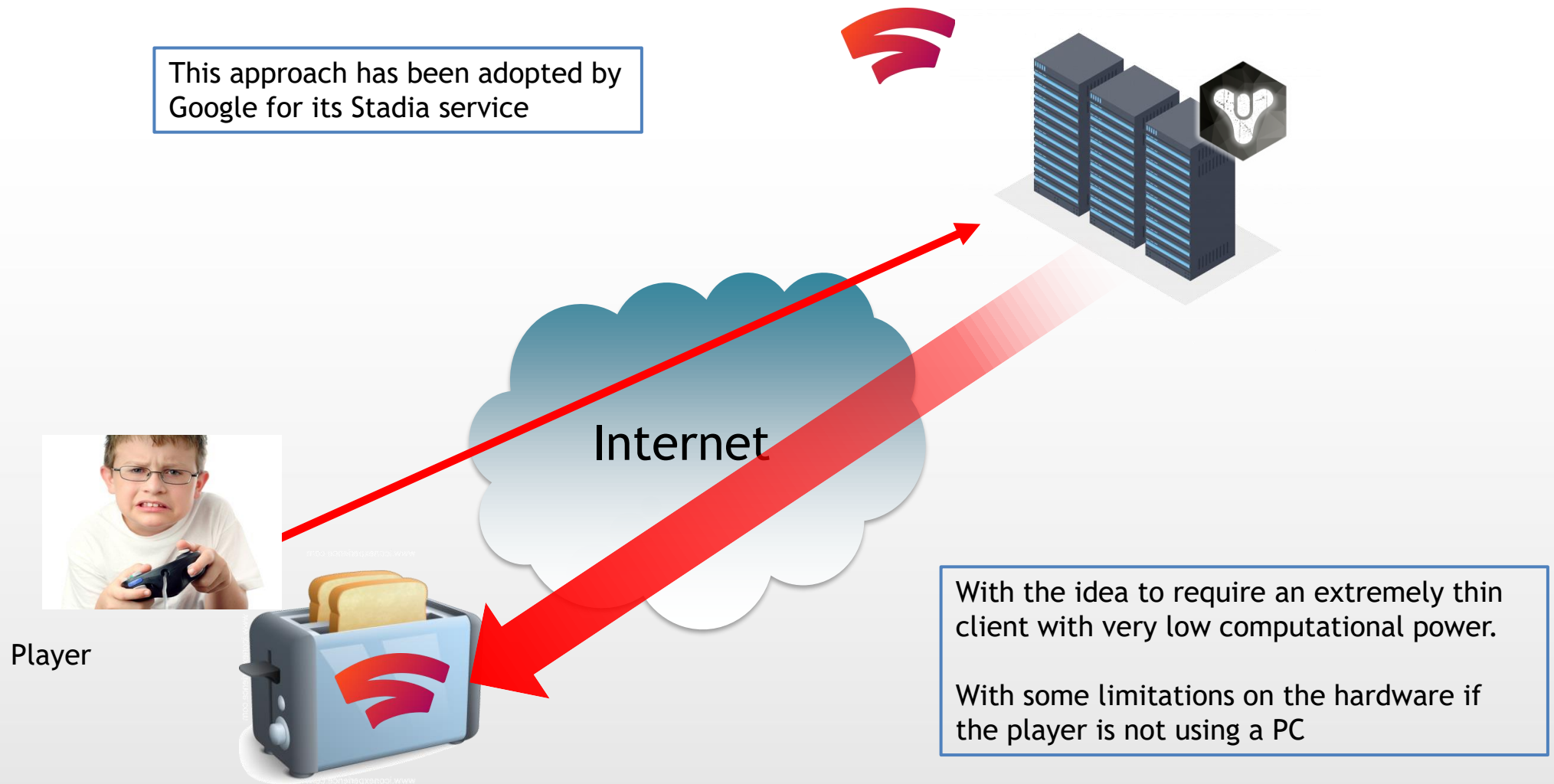- ## The streaming service can leverage on a HUGE library of existing titles already available at day one

Today, this is a selling point for Microsoft marketing department

UNIVERSITÀ DEGLI STUDI DI MILANO

Disadvantages of Streaming an Emulated Machine

- ## We are still running the game from inside a PC

  - It can be extremely powerful, but it is still a single machine

  - We can use powerful hardware, but we cannot tap on the datacenter raw computational power

- ## All network operations for online functionalities will suffer from the same constraints as of a PC at home

  - We can have fast data exchanges with other users (in the same datacenter) but the game scalability will be the same as if sitting on a desk

> From a technical standpoint we can argue that cloud technology has been designed to be "server side". In this case, we are using the cloud to run a client (or, at least, a large piece of it); undermining the initial design

Streaming From an Infrastructure

This approach has been adopted by Google for its Stadia service

Internet

Player

With the idea to require an extremely thin client with very low computational power.

With some limitations on the hardware if the player is not using a PC

Advantages of Streaming From an Infrastructure

- The infrastructure will be the only platform we will be developing for
  - Lower cost
  - Only one code base
  - There will be a middleware adapting the game on the fly to each client device

- We can tap on the datacenter raw computational power
  - Technically, it is an infinite amount of CPU

- Resources can be shared between game instances
  - Games can scale up much more than when using a single machine
  - Many players are just "there" to involve

  > Now we can finally have a brawl with THOUSANDS of players!

Disadvantages of Streaming From an Infrastructure

- The game software must be implemented explicitly for the infrastructure
  - The code base will not be portable, your project will be "here or nowhere"

- The streaming service will suffer from a limited library at day one, and will need time to gain momentum

Albeit infrastructure-based streaming is a more advanced solution, we have currently only one provider on the market (Google).

This is because having a large catalog ready at day one is much more compelling for investors and white collars

UNIVERSITÀ
DEGLI STUDI
DI MILANO

Stadia Streaming Architecture

And here is where the magic is happening!
A streamer will take care to adapt the output from the decoders to the network conditions
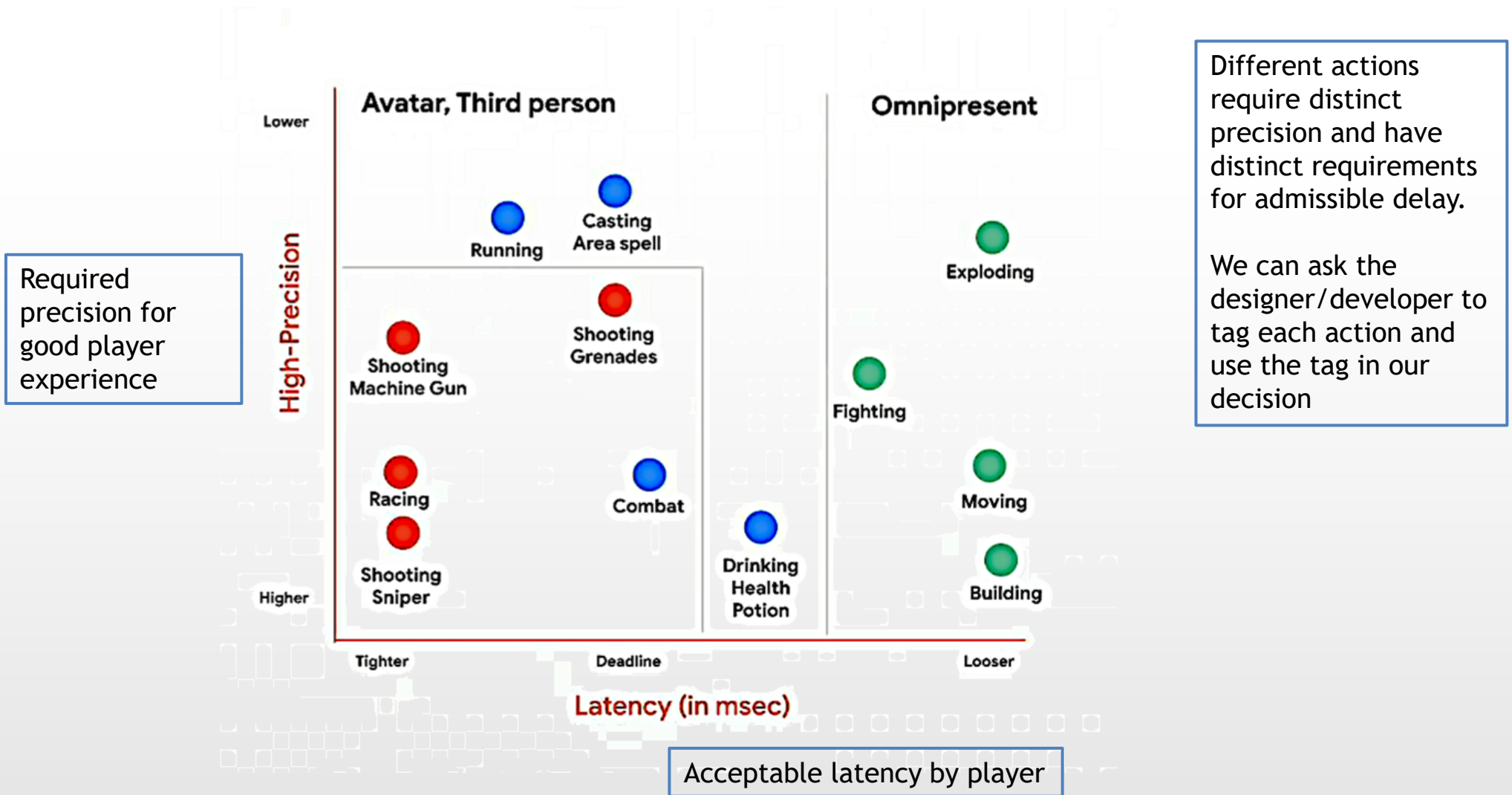
Your game is here
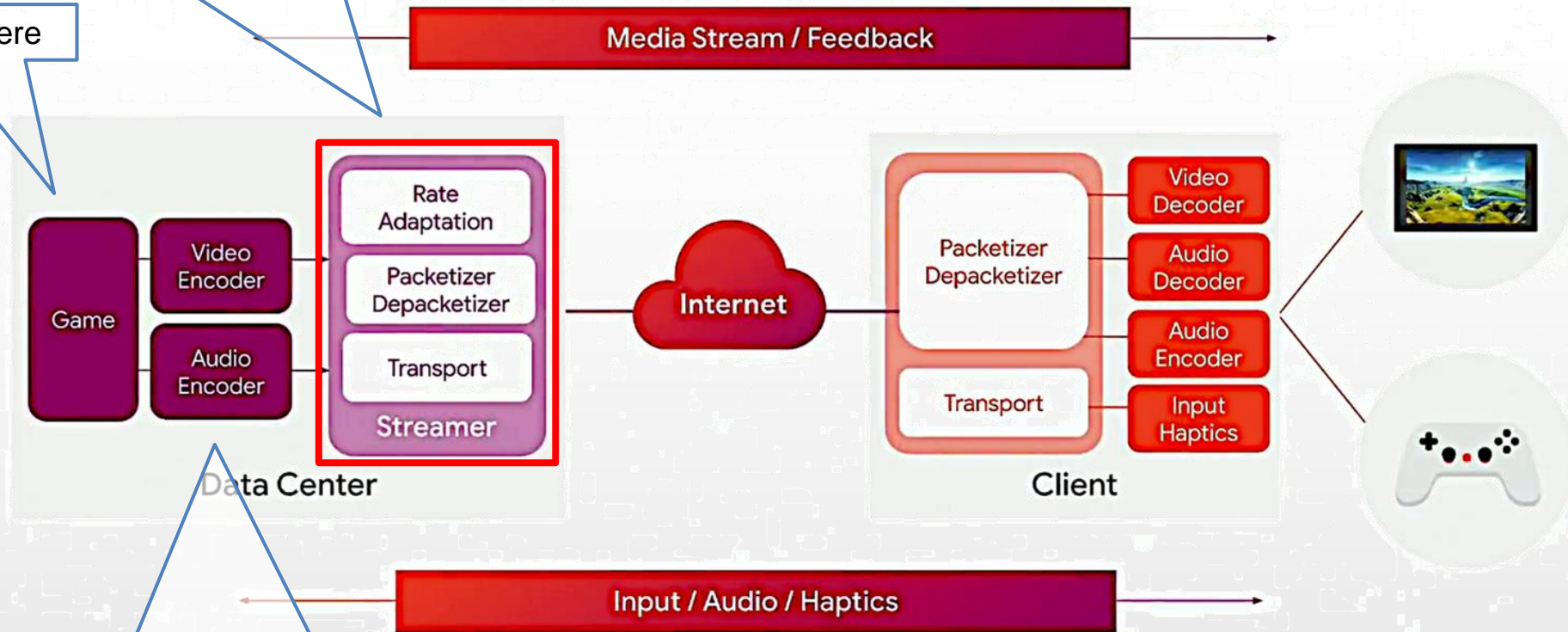
The only way your games can talk to the player is by these modules, which will encode your audio and video output
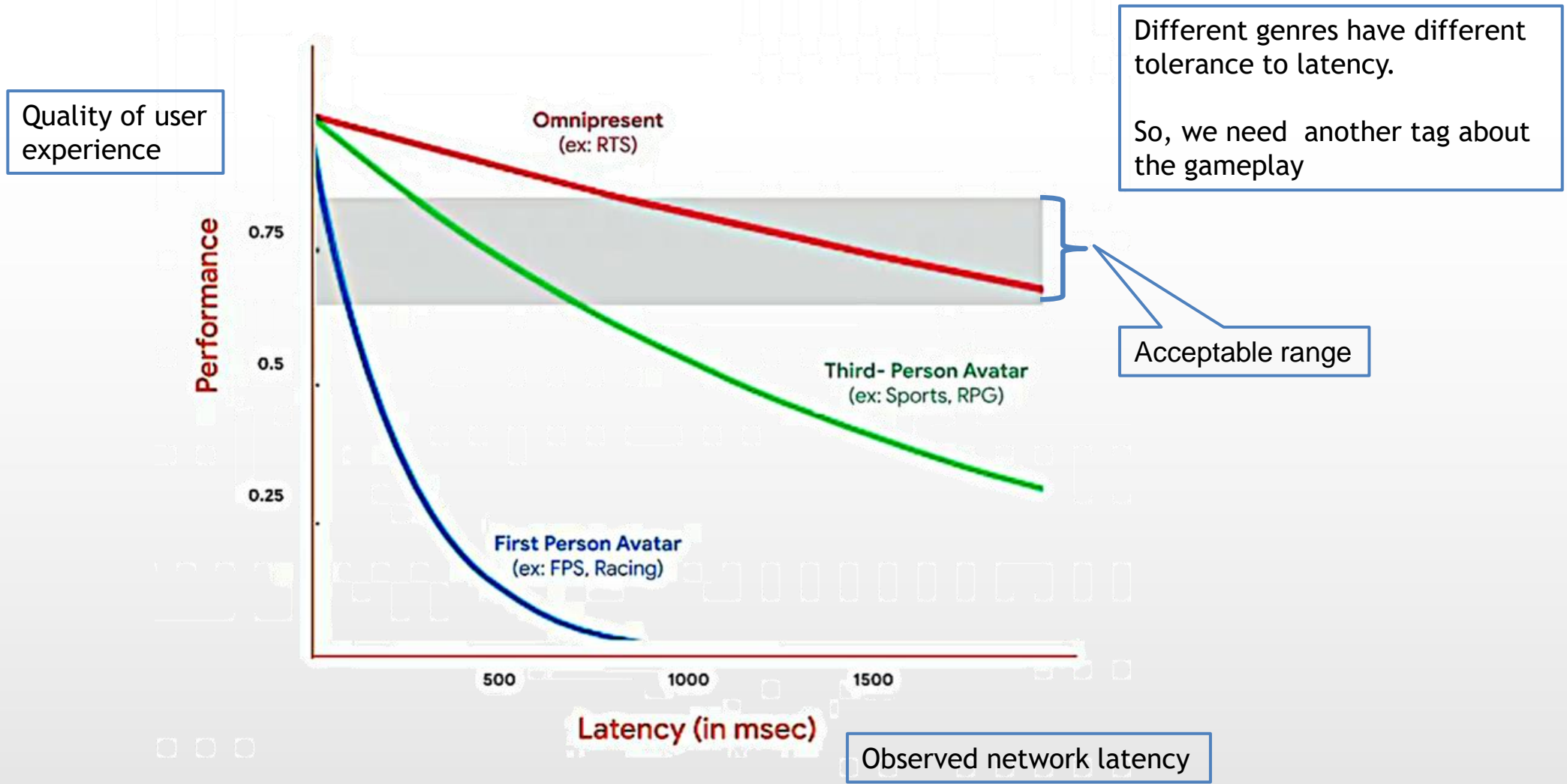
Streamer Adaptation

- This is a very complex problem
  - There is no golden rule about how to adapt a data stream

- The truth is that we cannot adapt only looking to the network and to the byte stream
  - Option 1: You can have near-zero latency at a very low bitrate
    - Our game will run smooth but will look awful
  - Option 2: You can have perfect video quality at a very low framerate
    - Our game will look great but it will be unplayable

- We are looking for option 1.5 (a tradeoff)
  - The point is: what is the correct **perceptual** tradeoff between latency and bitrate (quality)?
    - This depends also on what we are doing inside the game
    - All we can do right now is to run zounds of experiments, and hope to see a trend teaching us something useful
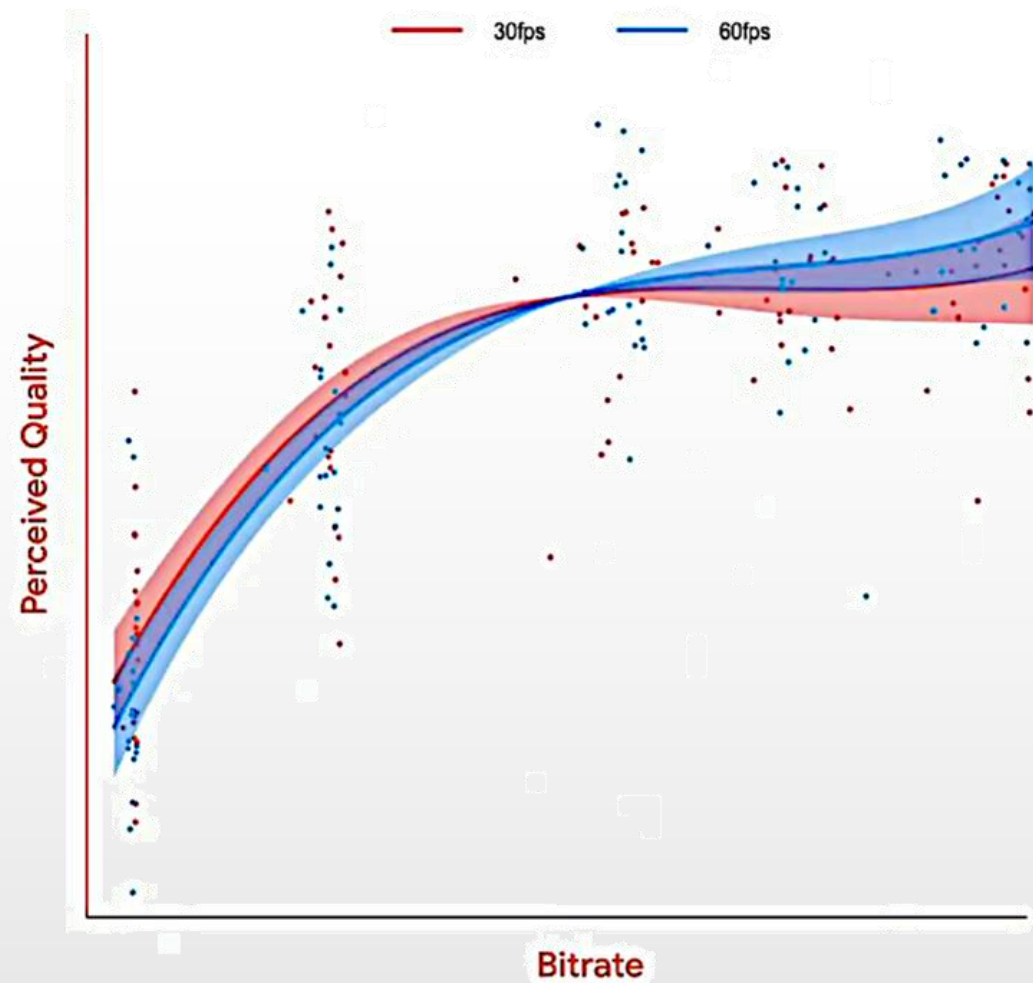
Latency vs Quality by Actions

Avatar, Third person

Lower

High-Precision

Running
Casting Area spell
Shooting Grenades
Shooting Machine Gun
Racing
Combat
Shooting Sniper
Drinking Health Potion

Higher

Tighter — Deadline

Required precision for good player experience

Omnipresent

Exploding
Fighting
Moving
Building

Looser

Latency (in msec)

Acceptable latency by player

Different actions require distinct precision and have distinct requirements for admissible delay.

We can ask the designer/developer to tag each action and use the tag in our decision

Latency vs Quality by Genres

Quality of user experience

Different genres have different tolerance to latency.

So, we need another tag about the gameplay

Acceptable range

Observed network latency

Performance

Omnipresent
(ex: RTS)

Third- Person Avatar
(ex: Sports, RPG)

First Person Avatar
(ex: FPS, Racing)

0.75

0.5

0.25

500   1000   1500
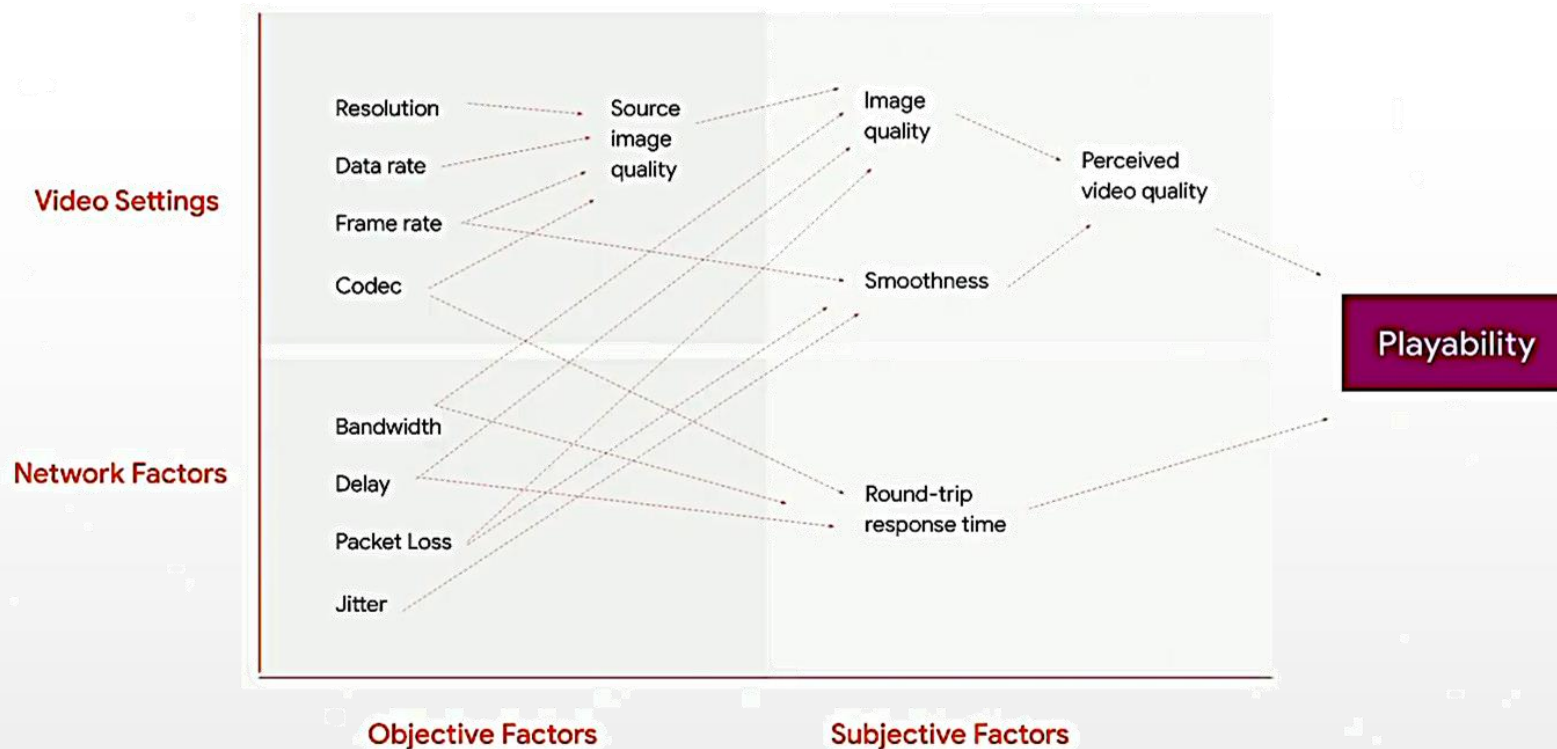
Latency (in msec)

Bitrate vs Perceived Quality

Moreover, from a perceptual standpoint, lower framerates perform better when we have a lower bitrate available

Objective vs Subjective Factors

Some parameters are difficult (if not impossible) to measure, because they are specific to each player.
What we can do here, is to establish a hierarchy to give them the right priority

Streaming in stadia

- Once we guessed a good compromise, we still have the problem to bring the video to the player

- We must monitor network conditions and react to congestion generated by cross traffic
  - We want to react to packet loss … without retransmission!

Reaction Without Retransmission

I am expecting you know how MPEG is working

- Stadia defines a streaming protocol where the MPEG GOP (Group Of Pictures) is made of the initial I-frame follower only by P-frames that will update the scene on the screen
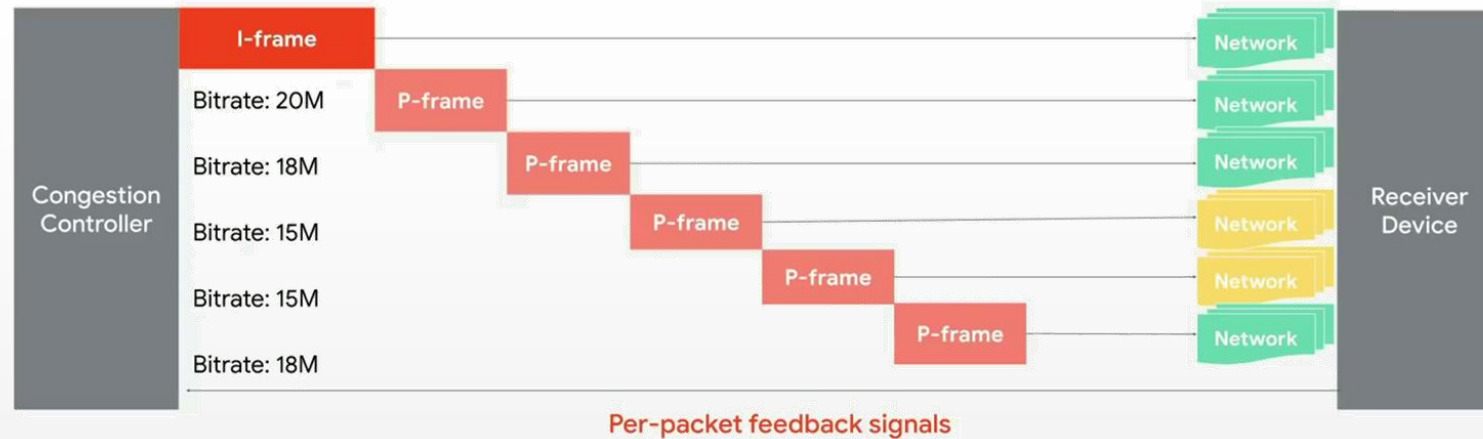
- Each P-frame will be encoded in a specific bitrate compatible with current network conditions
- The client is required to give feedback to the congestion controller for every single packet
    - Each feedback will update the available network resources estimation

Reaction Without Retransmission

- On a packet loss, the GOP will be aborted, and the next P-frame will be substituted with an I-frame to recover video quality immediately
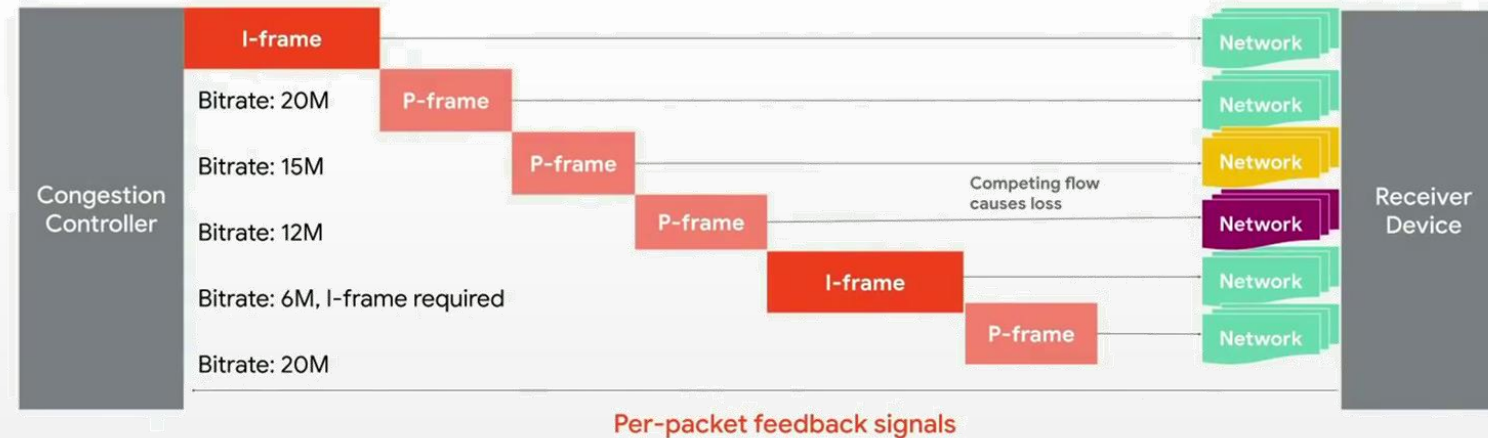
- Bitrate available to the following P-frames is adjusted to compensate the unexpected (larger) size of the injected I-frame

References

- Stadia Streaming Tech: A Deep Dive (Google I/O'19)
online: https://www.youtube.com/watch?v=9Htdhz6Op1I
McCool et all

- **NOTE**: written technical documentation on this topic is still very limited.
At the interview, I will ask questions about the content of this video
  - So, consider it like all other references