The Tension Between High Video Rate and No Rebuffering

Te-Yuan (TY) Huang
Stanford University

IRTF Open Meeting @IETF 87
July 30th, 2013

Joint work Prof. Nick McKeown
and Prof. Ramesh Johari

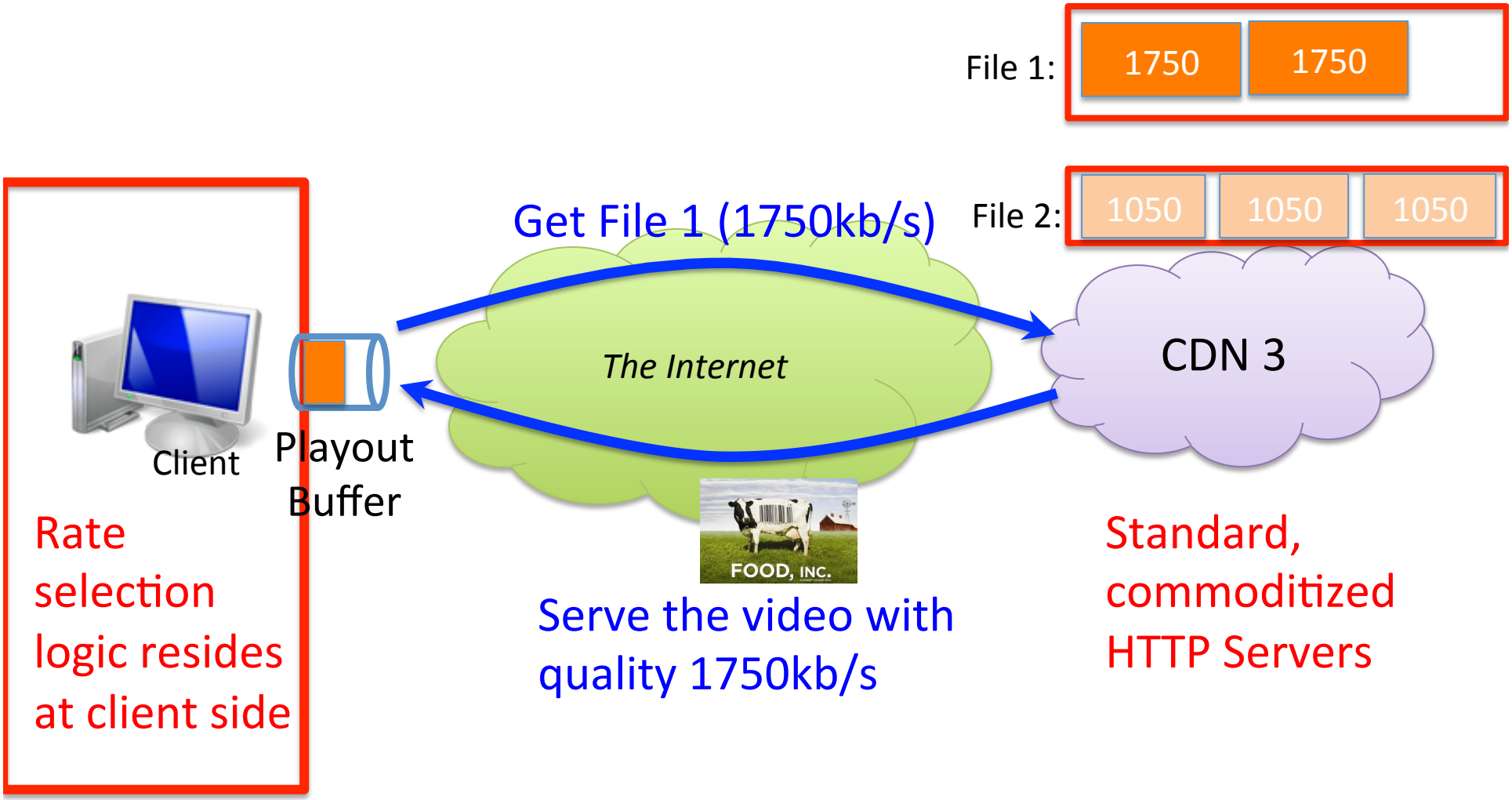
Outline

- How does streaming video work today?
  - Video streaming over HTTP
  - Video rate selection over HTTP
  - The goals of rate selection
- The tension between the goals
- Initial thoughts on how to break the tension

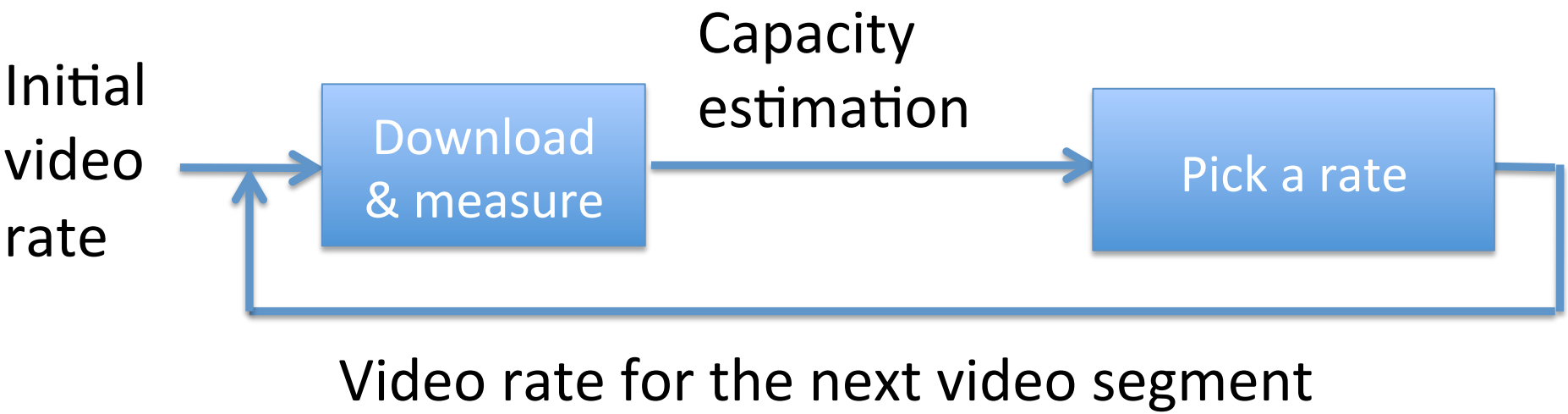
Video is the BIG thing on the Internet

- Video is more than 50% of peak traffic in the US
- Trend: Streaming over HTTP
  - Content Distribution Networks (CDNs)
    - Well-provisioned HTTP servers at the edge of the Internet
    - Cheap (1-2 cents per GB in 2013)
  - Firewall friendliness

How does rate selection over HTTP work?



The Rate Selection Process



The Common Goals

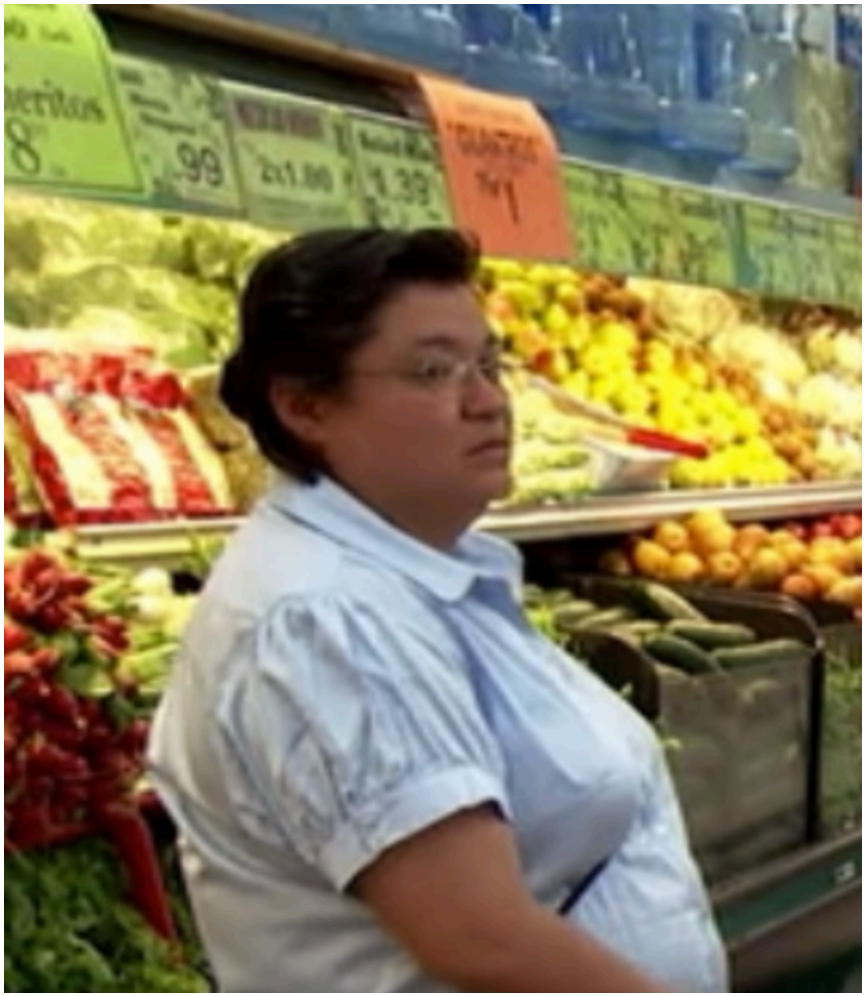
1. Achieve the highest possible video rate
   - Video rate represents video quality
2. Avoid "rebuffer" as much as possible
   - "Rebuffer" means under-running playout buffer
     - Unavoidable: Network or Service Outage
       - Necessary rebuffers
     - Avoidable: Requesting a video rate that is too high
       - Unnecessary rebuffers

The Tension Between The Goals

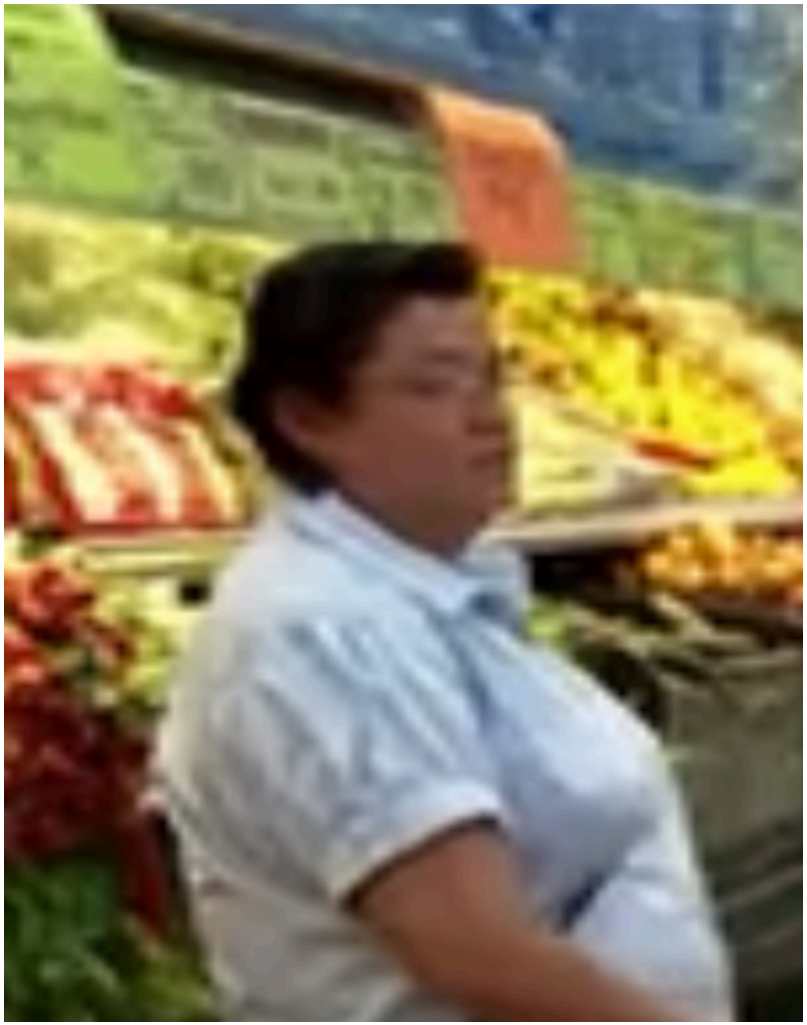
- The actual capacity is unknown and varies
  - Accurate estimation is hard
- Underestimate the capacity
  - Picking a rate that is too low
  - leads to sub-optimal video quality (Fail Goal #1!)
- Over-estimate the capacity
  - Picking a rate that is too high
  - leads to rebuffering (Fail Goal #2!)


53:33
Food, Inc.
HD

Before download started



After download started

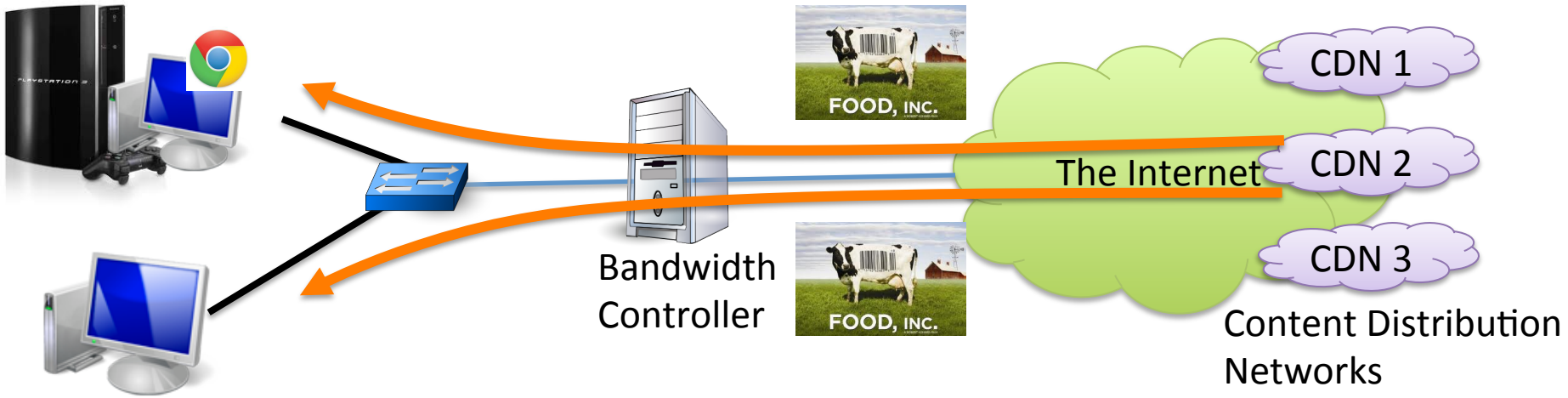


What happened?

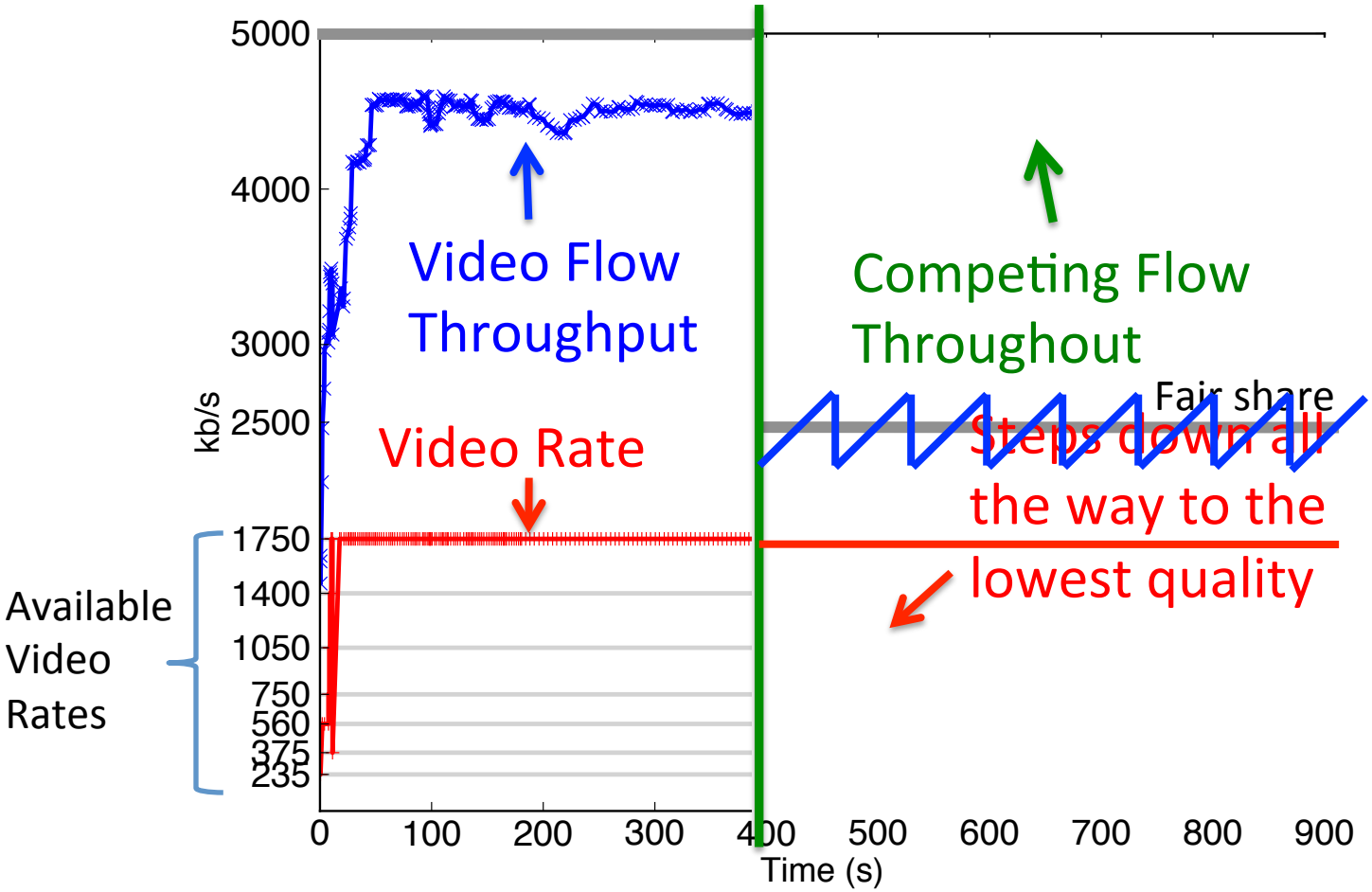
- Both the download and video are over HTTP
  - TCP shares my home link equally among all flows

What is the problem?

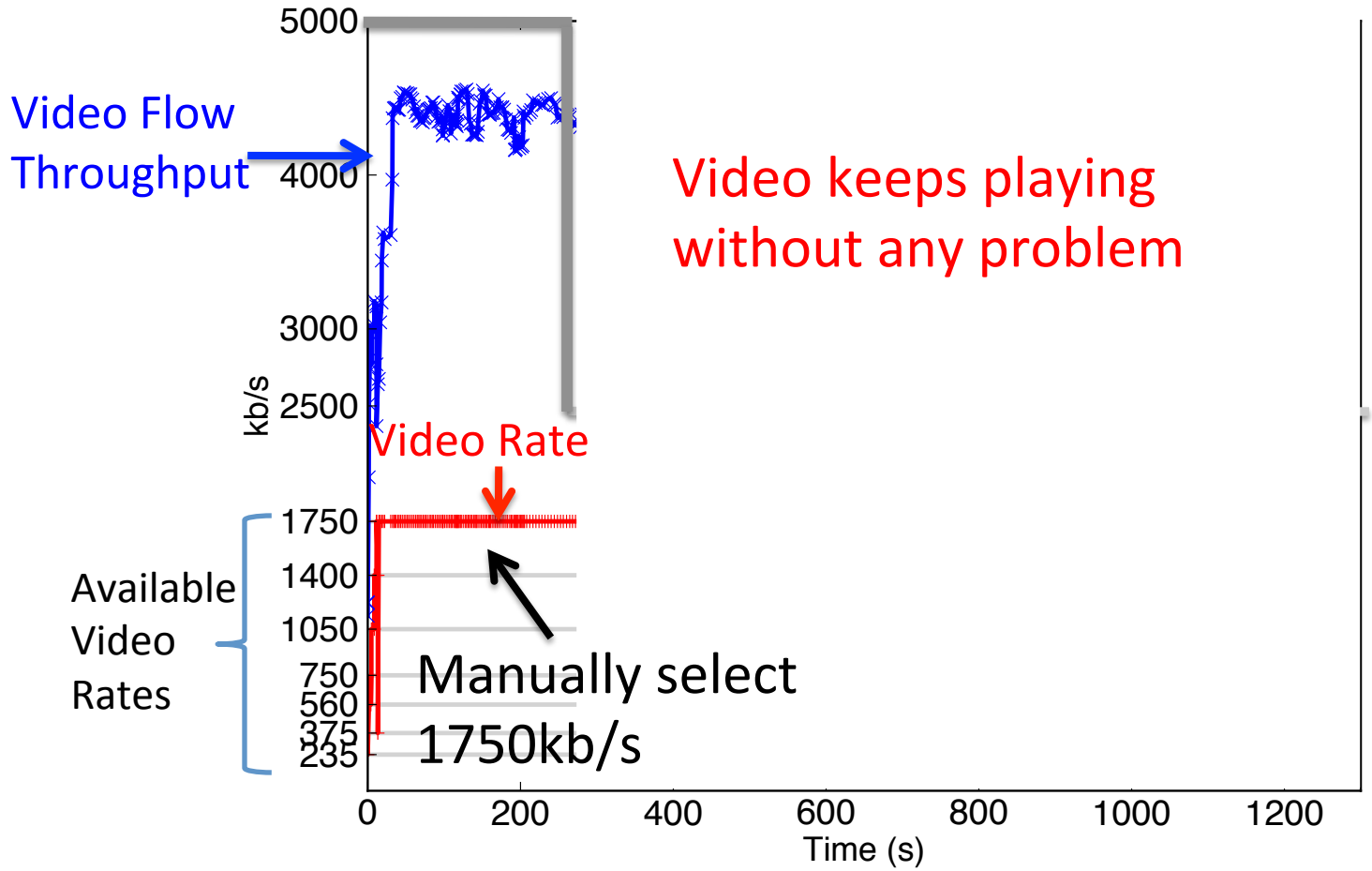
Experiment Setup



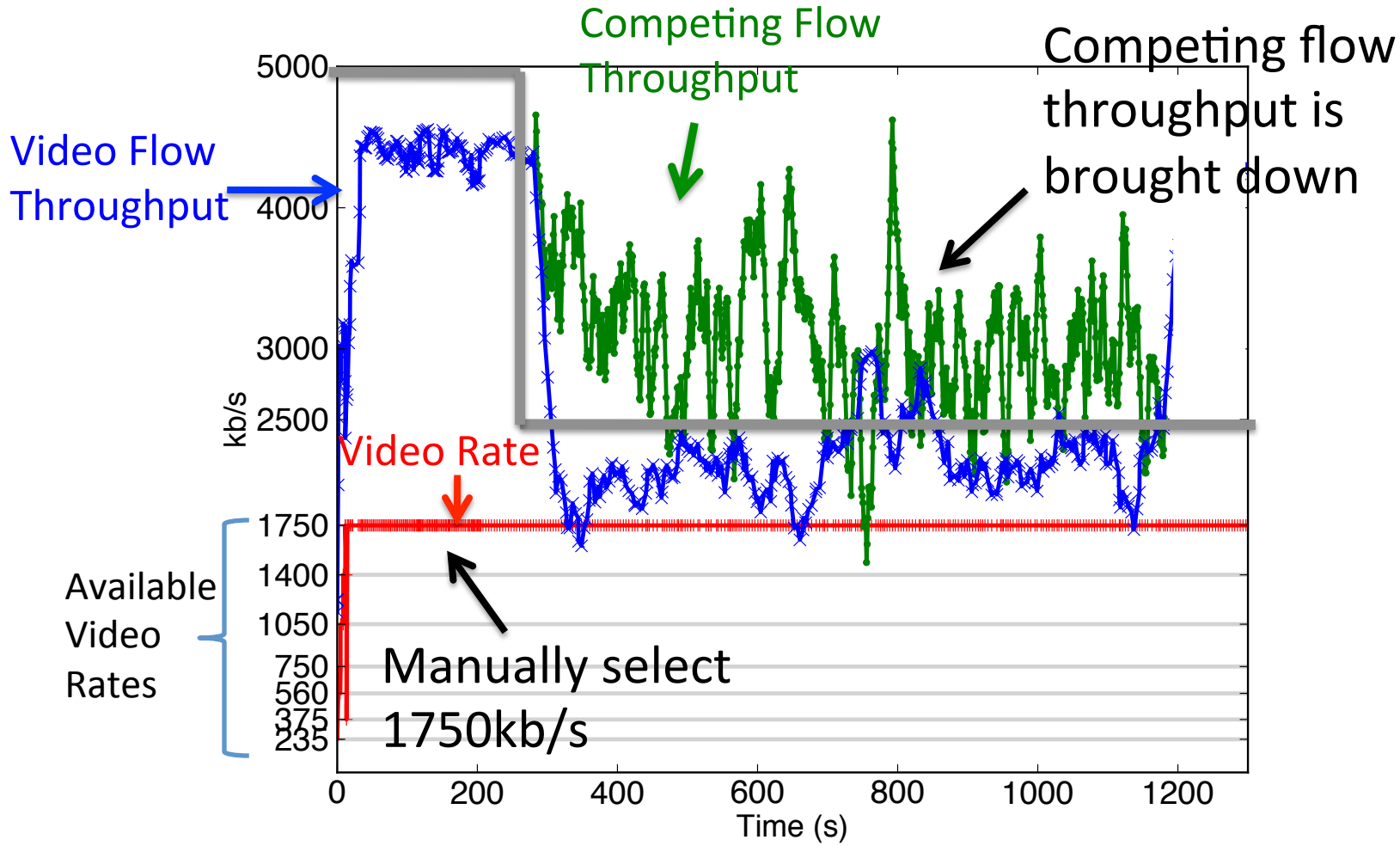
Video Rate
in the Presence of a Competing Flow



What If we manually select a video rate?

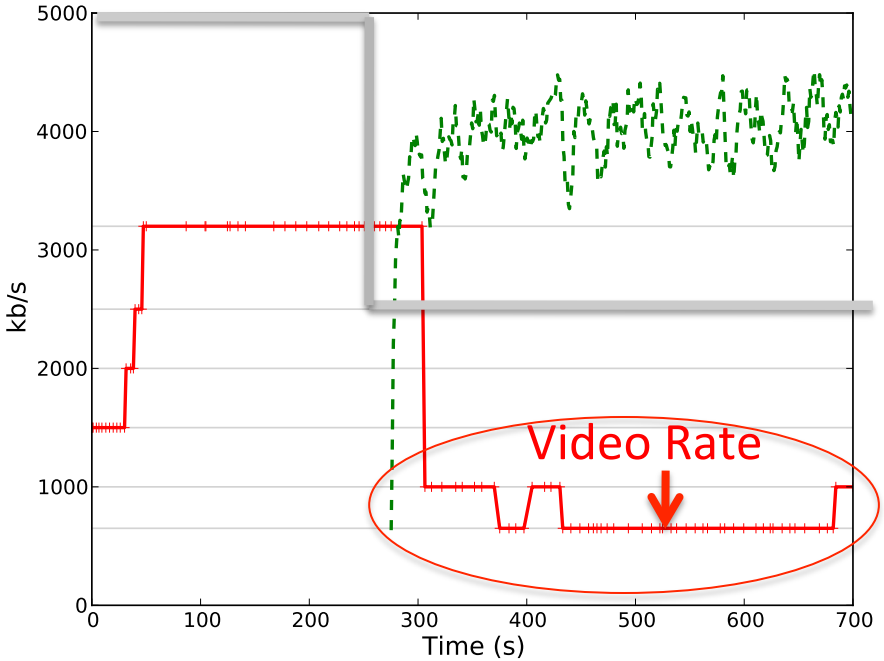


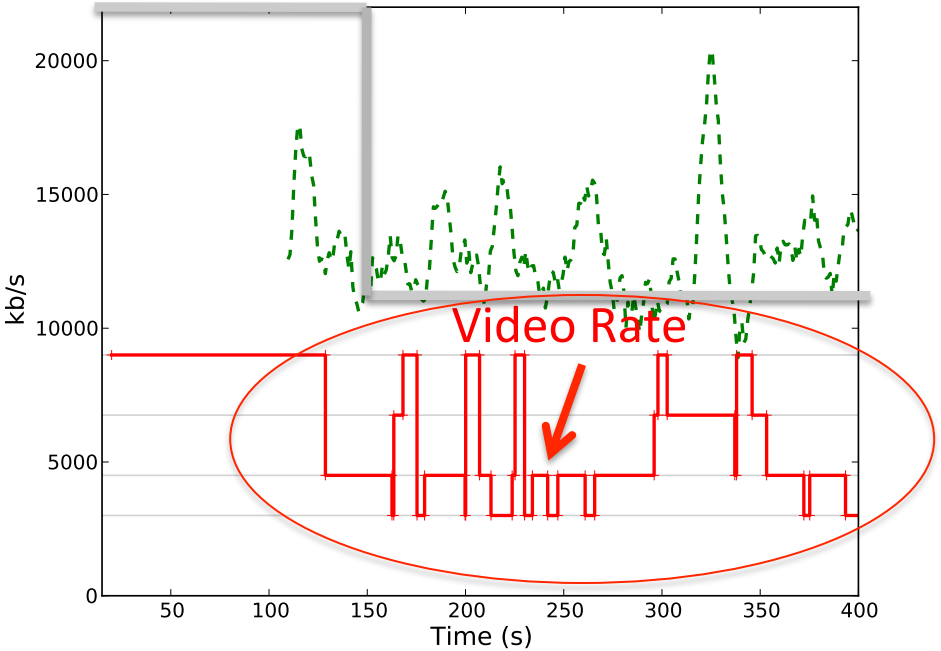
What If we manually select a video rate?



Not Just One Service's Problem

- This happens in all the three services we measured
  - Hulu, Netflix and Vudu





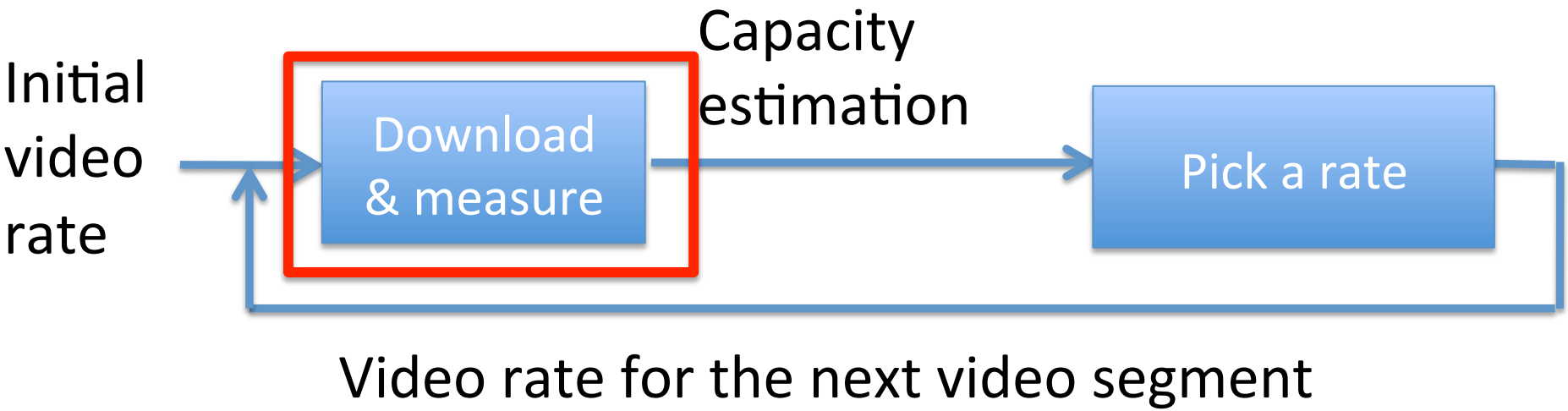
The Problem

Video client ends up with much less throughput than its fair share

It picks a video rate that is much too low

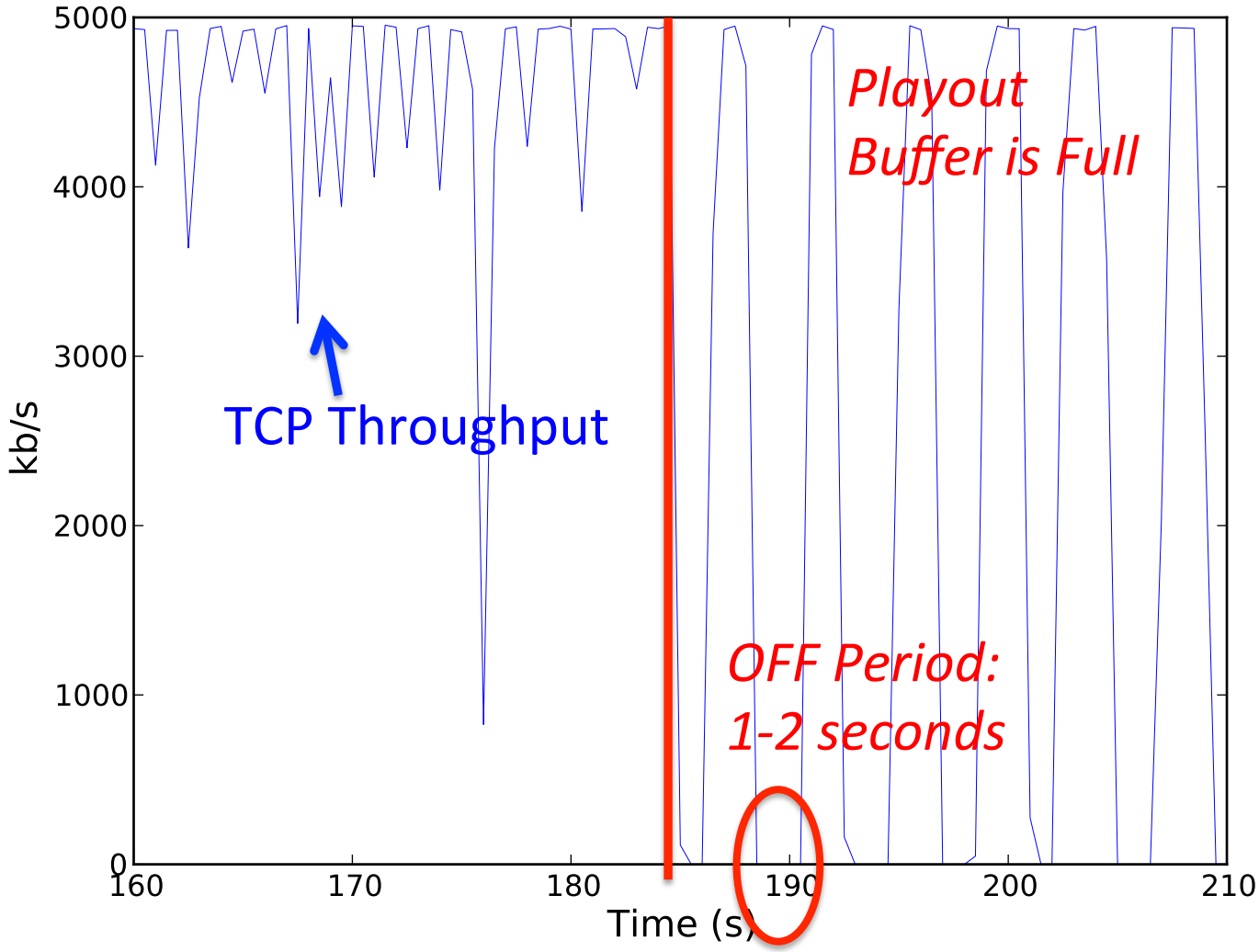
Why?

The Rate Selection Process



What goes wrong?

TCP Throughput of the Video Flow

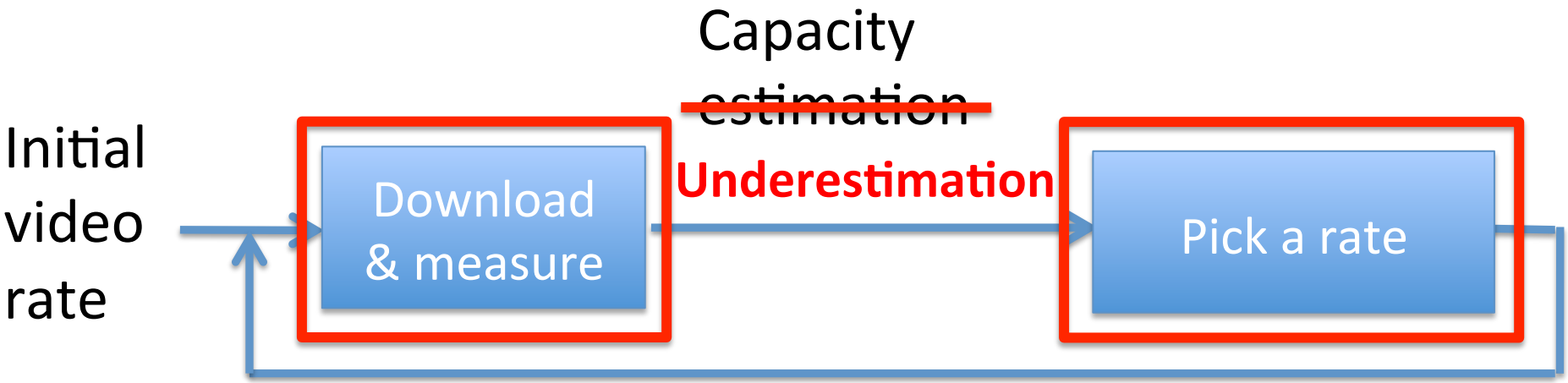


Impact of OFF period on TCP

- TCP sender resets its congestion window
  - When idle more than one RTO (200ms)
  - Slow-start restart, RFC 2581/5681
  - Linux 3.x (tcp_output.c, line 163)
- Throughput will be affected
  - Worse with a competing flow
    - Experience packet loss during slow start

50% of the segments get < 1.8Mb/s
(Fair Share is 2.5Mb/s)

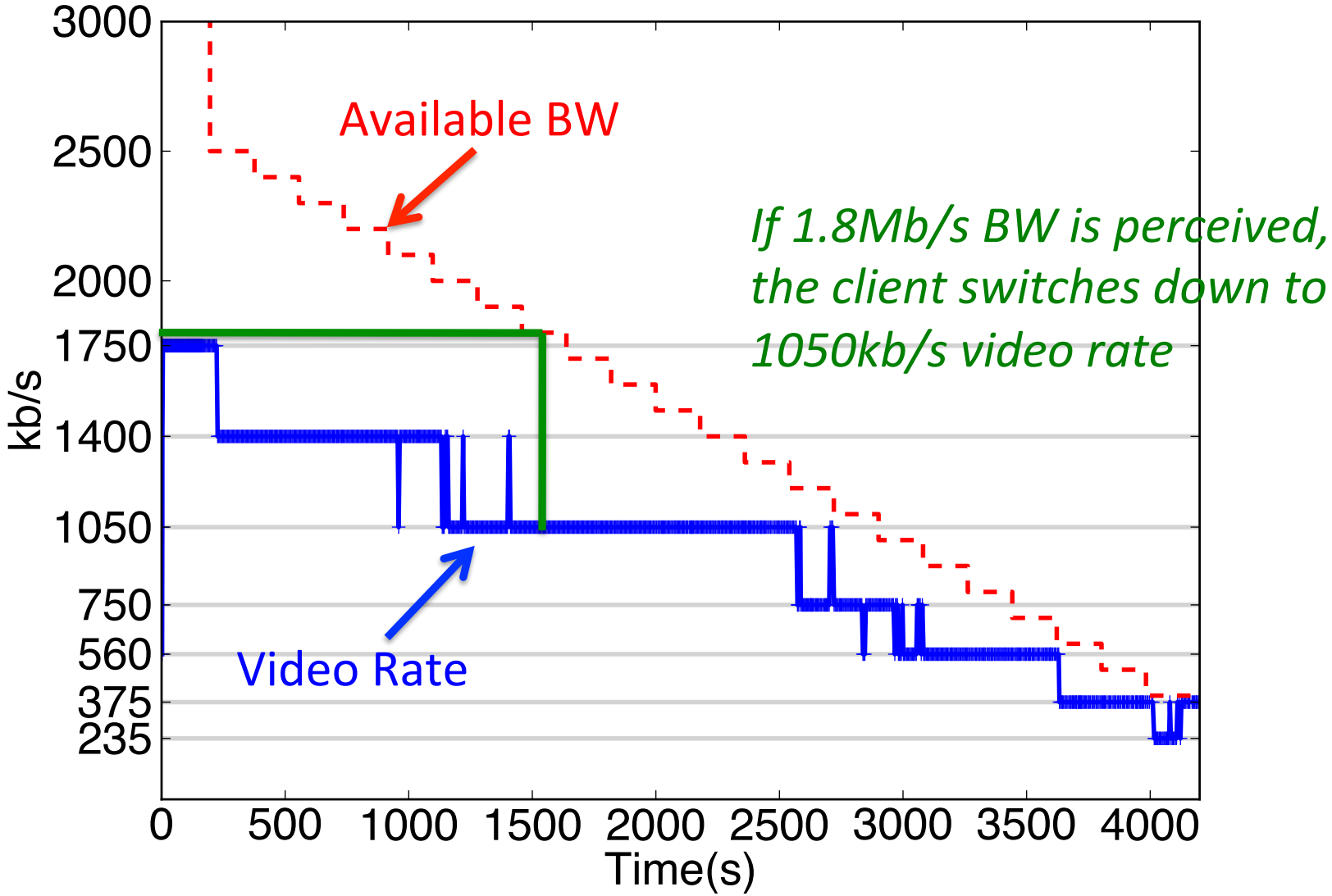
The Rate Selection Process



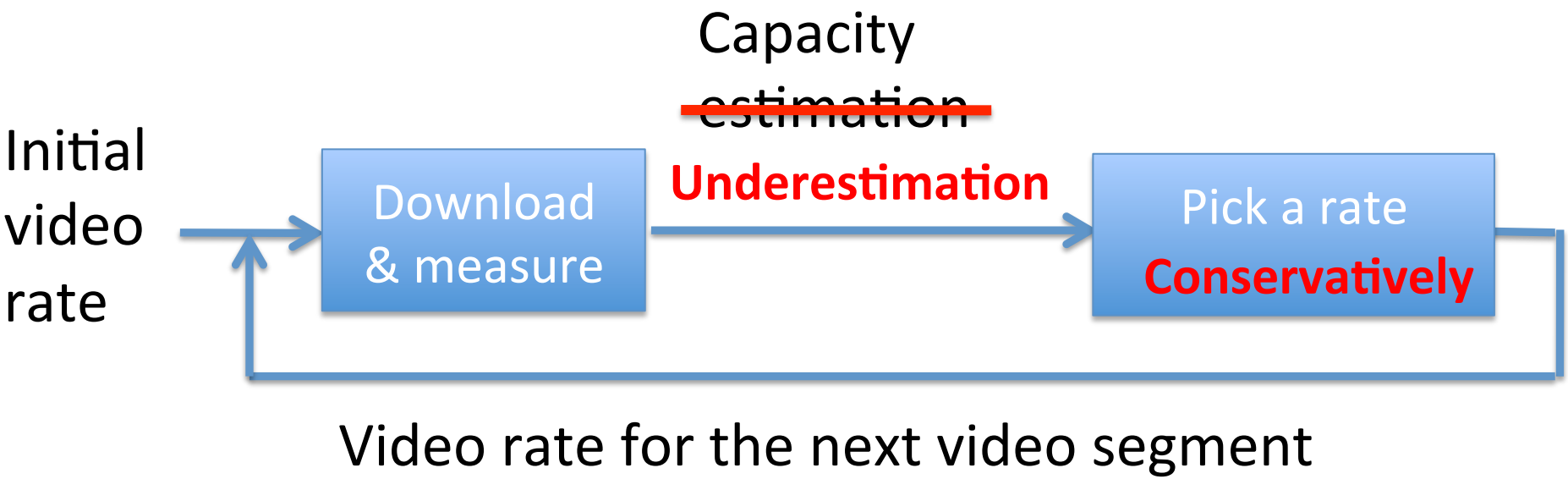
If perceived 1.8 Mb/s,
which video rate would it pick?

1750 kb/s? 1400 kb/s? Even lower?

Conservative Rate Selection

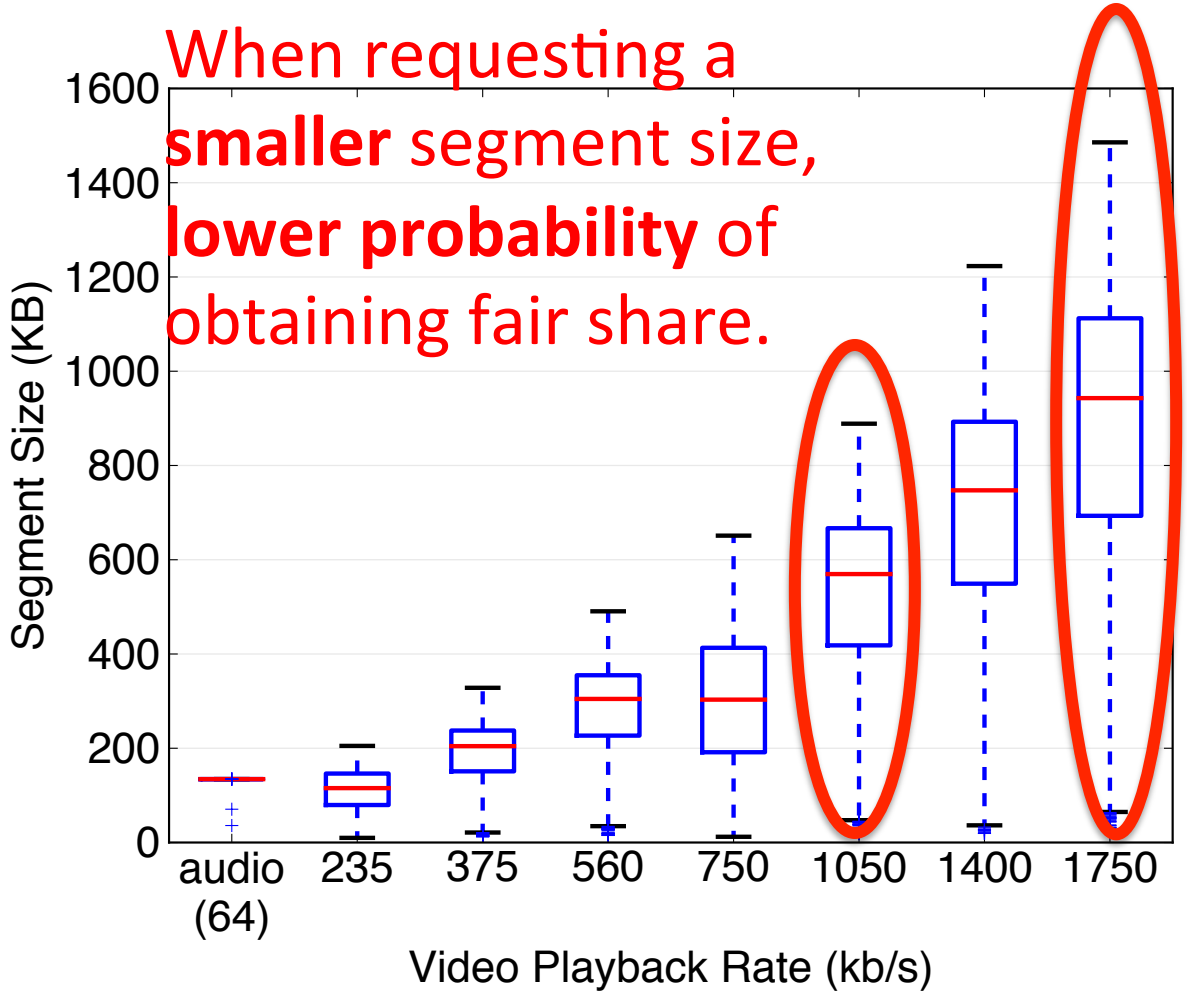


The Rate Selection Process

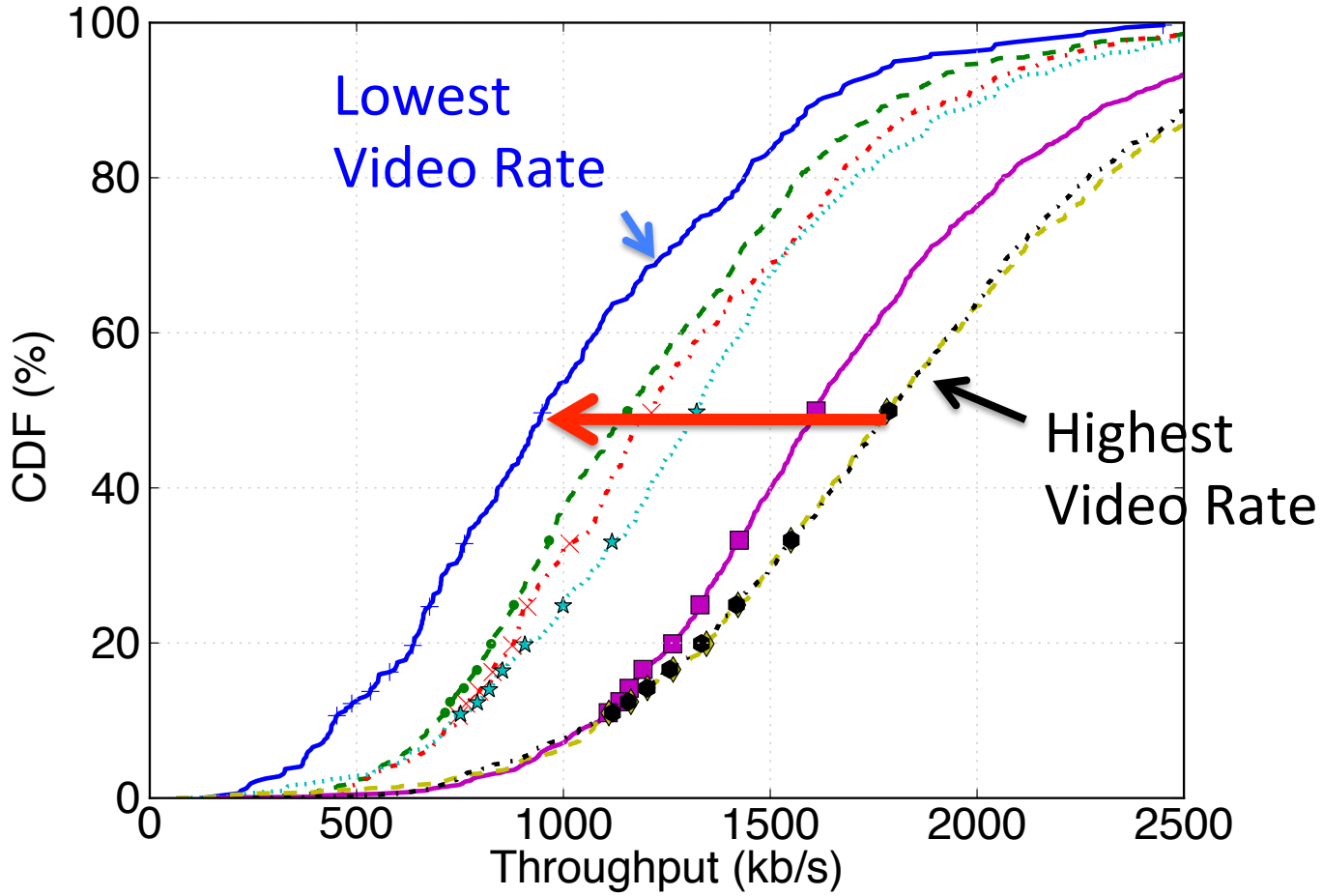


Is there any consequence of being conservative?

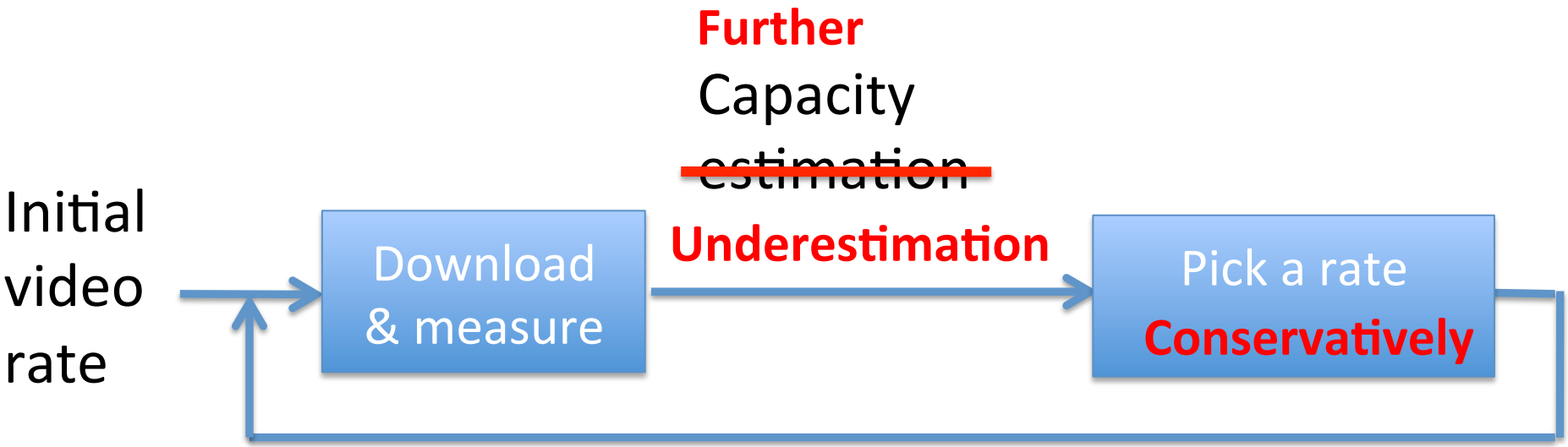
Smaller Segment Size for Lower Video Rate



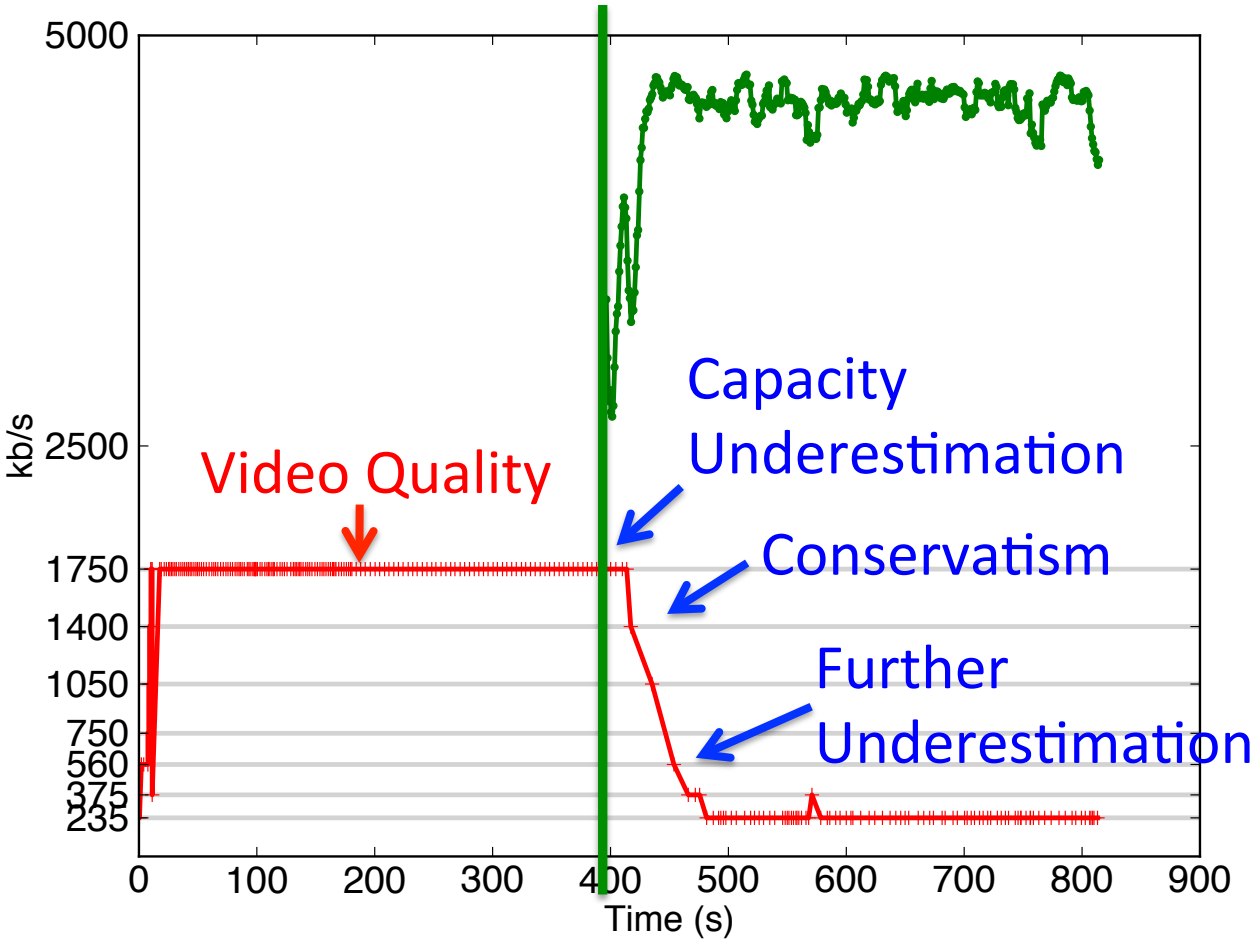
Lower video rate leads to further bandwidth underestimation



The Rate Selection Process



The Complete Story



Being conservative can trigger a vicious cycle!

But the available capacity is unknown & varies,

So being conservative is understandable.

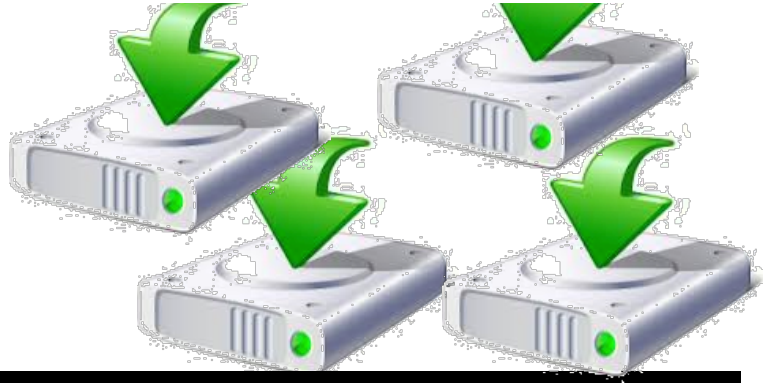
Although this leads to sub-optimal quality (Fails Goal #1),

at least it will avoid rebuffer (Goal #2) …

Right?

The Tension Between The Goals

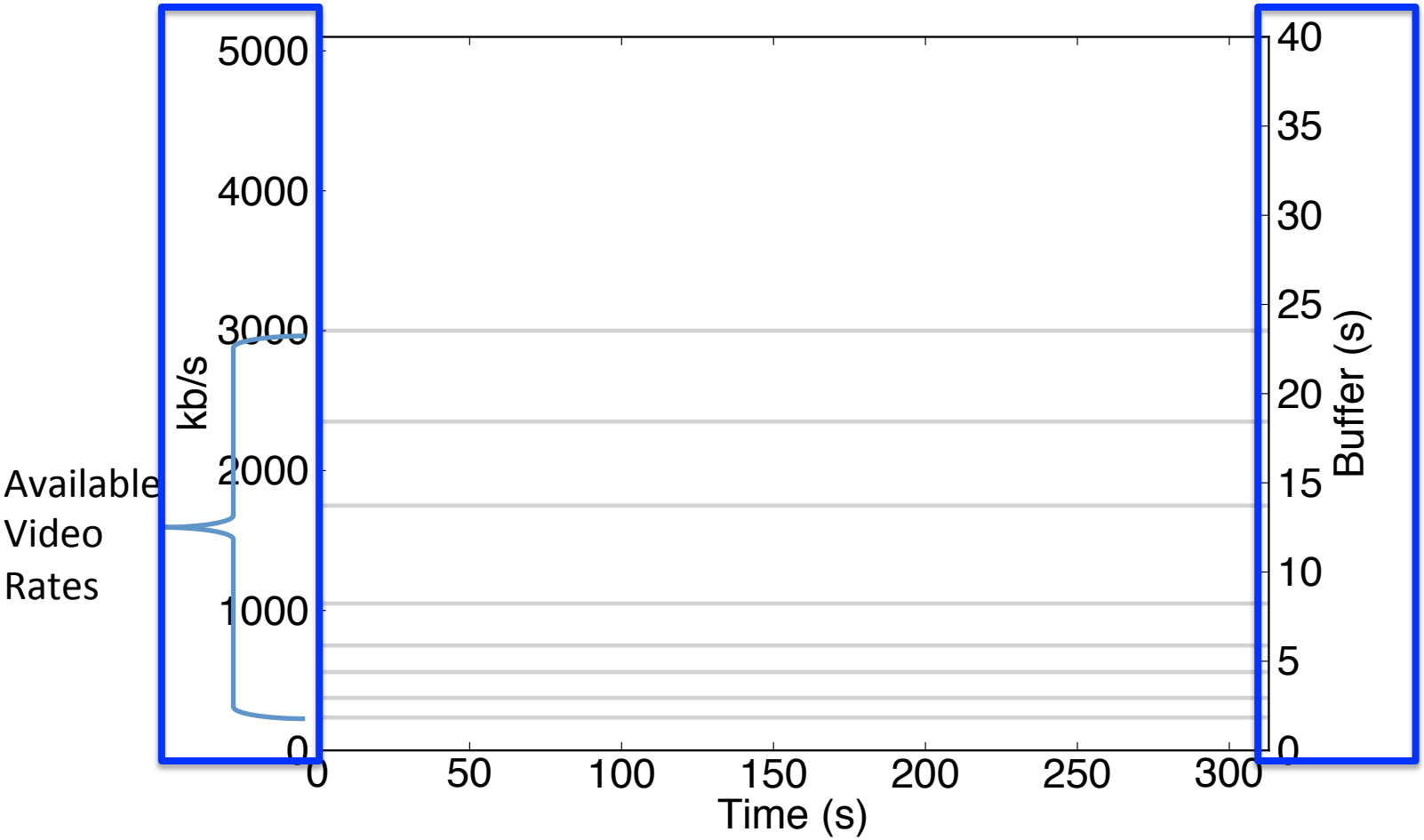
- The actual capacity is unknown and varies
  - Accurate estimation is hard
- Underestimate the capacity
  - Picking a rate that is too low
  - leads to sub-optimal video quality
- Over-estimate the capacity
  - Picking a rate that is too high
  - leads to rebuffering



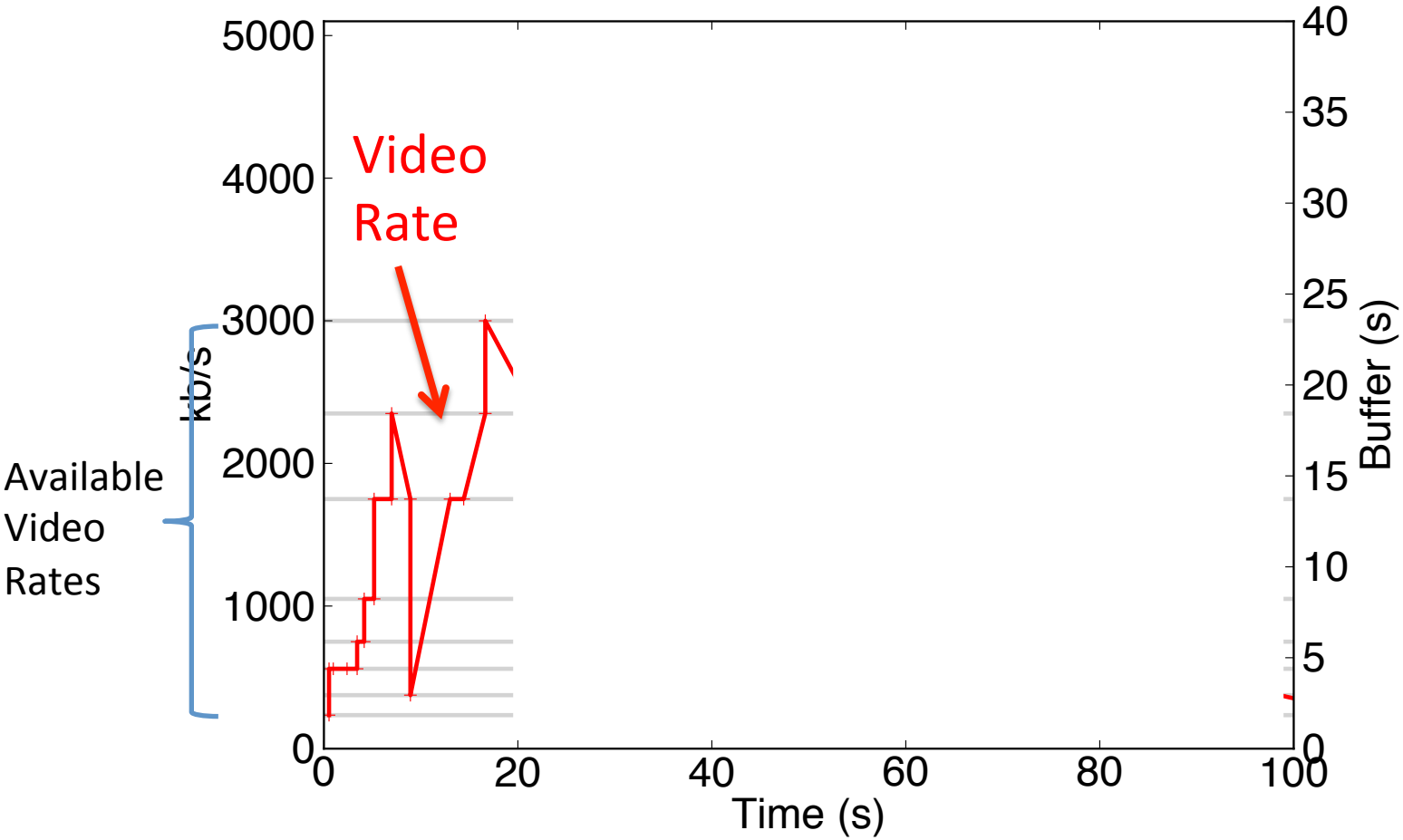

MESSAGE, SPOCK?

What happened?

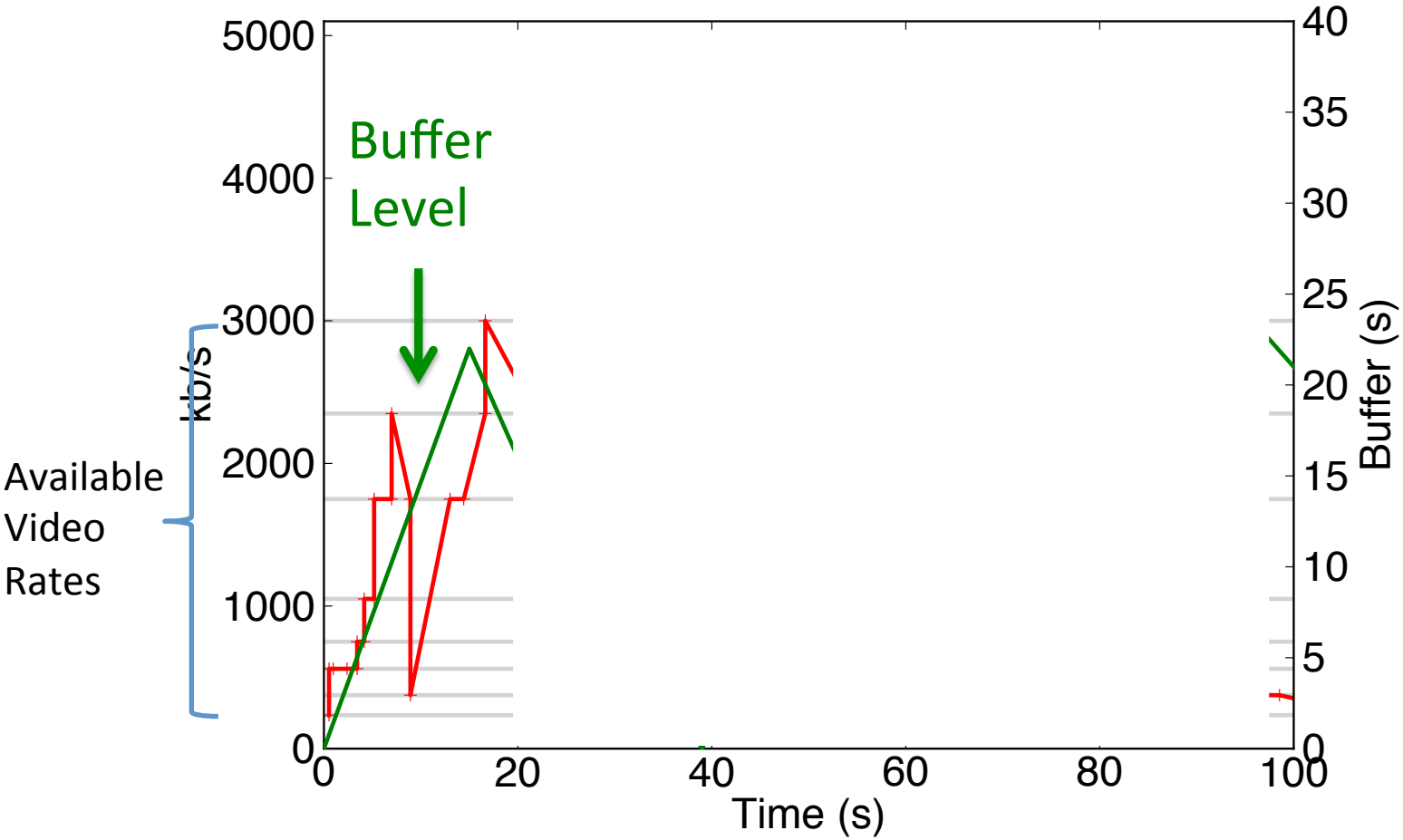
What happened?



What happened? – Cont.



What happened? – Cont.



What happened? – Cont.

Why did the stream rebuffer?

- Capacity is estimated by weighted average of recent throughput
- The drop of the actual capacity does not reflect on the estimation until later
- Request a higher rate than it should
- End up under-running the buffer

Rebuffer!

Why does the tension exist?

- Under-estimate the capacity
  - Picking a rate lower than the actual capacity
  - leads to sub-optimal video quality (Fail Goal #1!)
- Over-estimate the capacity
  - Picking a rate higher than the actual capacity
  - leads to rebuffering (Fail Goal #2!)
- To break the tension
  - Need accurate capacity estimation

Why the tension exists?

Why does the tension exist?

- Pick video rate based on capacity estimation
  - The actual capacity is unknown and varies
  - The estimation never equals to the actual capacity
- The same algorithm can both **under-estimate** and **over-estimate** the capacity

What if….

Pick the video rate
based on something we know:

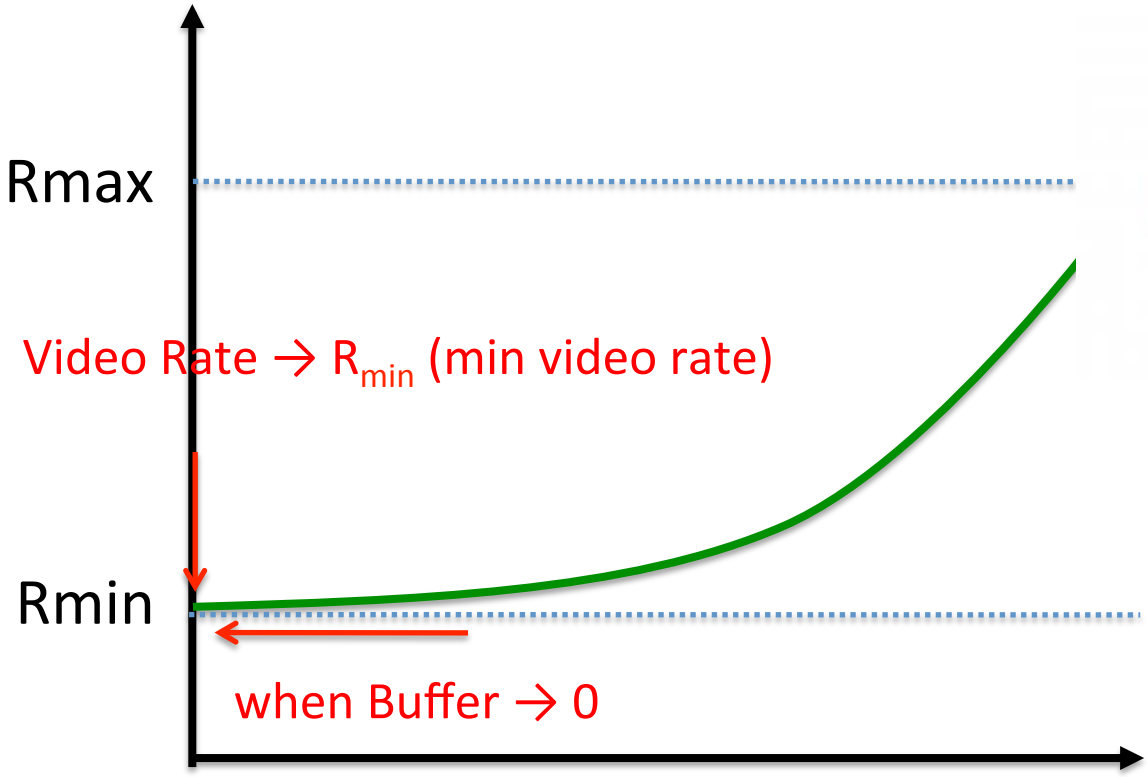
The Playout Buffer

Can we break the tension?

Break the Tension – Goal #1

- To achieve the highest possible video rate:
  - Need to fully utilize the capacity
    - Avoid ON/OFF behavior
    - Unless we have more capacity than we need
- Request for the highest video rate before the buffer is full
  - Playout Buffer will only be full:
    - When the capacity is larger than the highest video rate
    - Have more capacity than needed
  - Fully utilize the bandwidth
  - Average video rate = Average throughput

Goal #2: Avoid Rebuffers



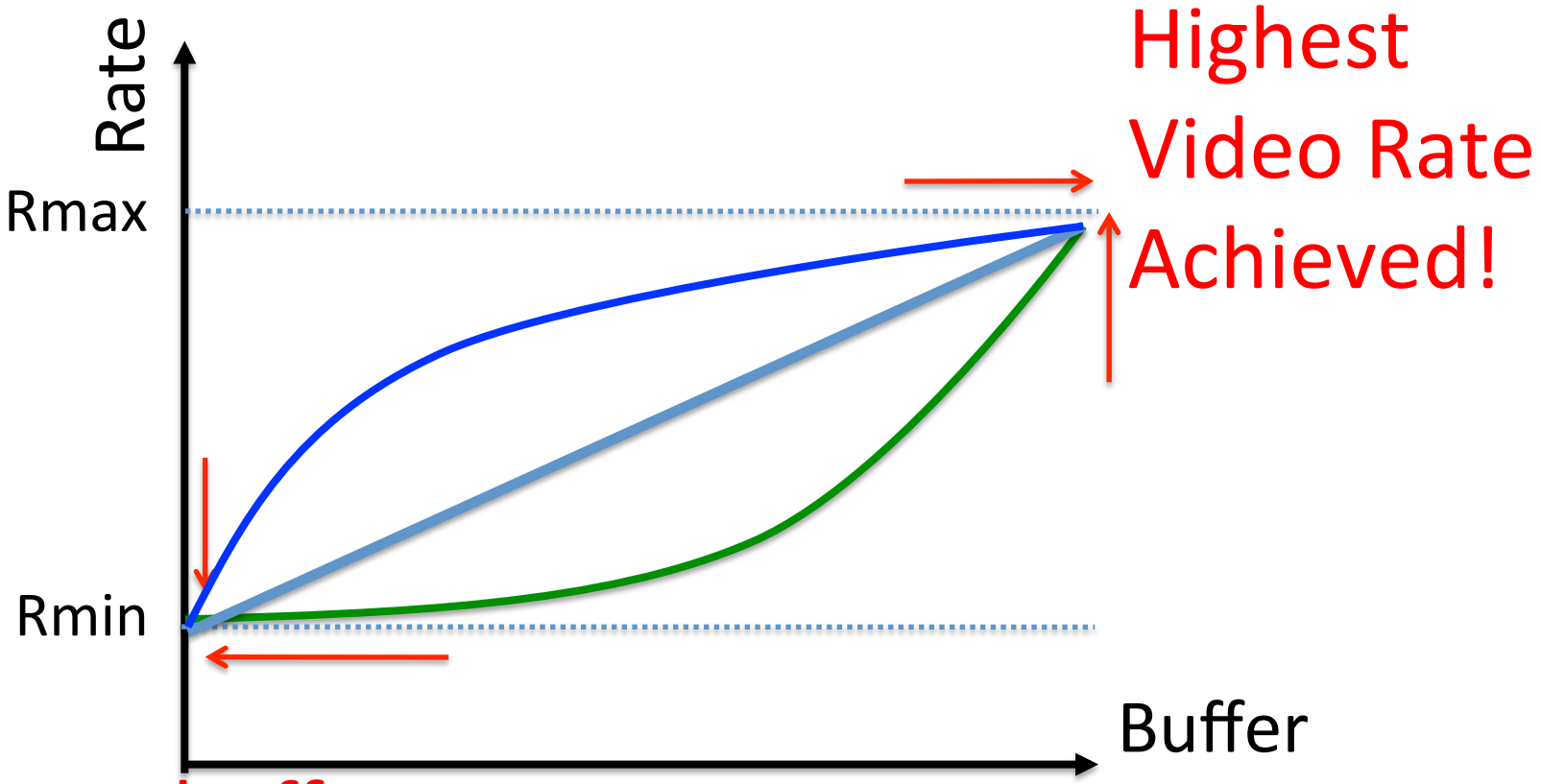
As long as Capacity > R_{min}, Buffer will grow.

Never Unnecessarily Rebuffer!

Break the Tension – Goal #1

- To avoid rebuffer:
  - Step down to the lowest video rate when buffer approaches to zero
  - Buffer will start growing once stepping down
    - As long the capacity is larger than the lowest rate
  - Playout Buffer will only under-run when:
    - The capacity is **less** the lowest video rate
    - There is nothing we can do, the rebuffer is “necessary”
    - Called “necessary Rebuffers”
- Guarantee to have zero “unnecessary rebuffers”

Break the Tension – Goal #1 & #2



Practical Concerns

- Discrete segment downloads
  - Can only pick video rate when requesting a segment
- Discrete video rates
- Many more…
- Buffer-based algorithm provides a framework to address these concerns

"*Downton Abbey without the Hiccups: Buffer-Based Rate Adaptation for HTTP Video Streaming*".
SIGCOMM FhMN Workshop, Aug. 16th, 2013

Conclusion

- Current practice of rate selection algorithm:
  - Pick a video rate based on capacity estimation
  - Two common goals:
    - Achieve highest possible video rate
    - Avoid rebuffer as much as possible
- The tension between two common goals
  - Underestimation vs. Overestimation
- The tension is caused by the estimation
  - Let's take that out from the algorithm
- Focus on the one thing we know: The Playout Buffer
  - The tension will be broken down