Assigning Responsibility for Deteriorations in Video Quality

Oleg Vasilyev, Henry Milner, Wensi Fu, Oleg White

CONVIVA

SPARK SUMMIT 2017

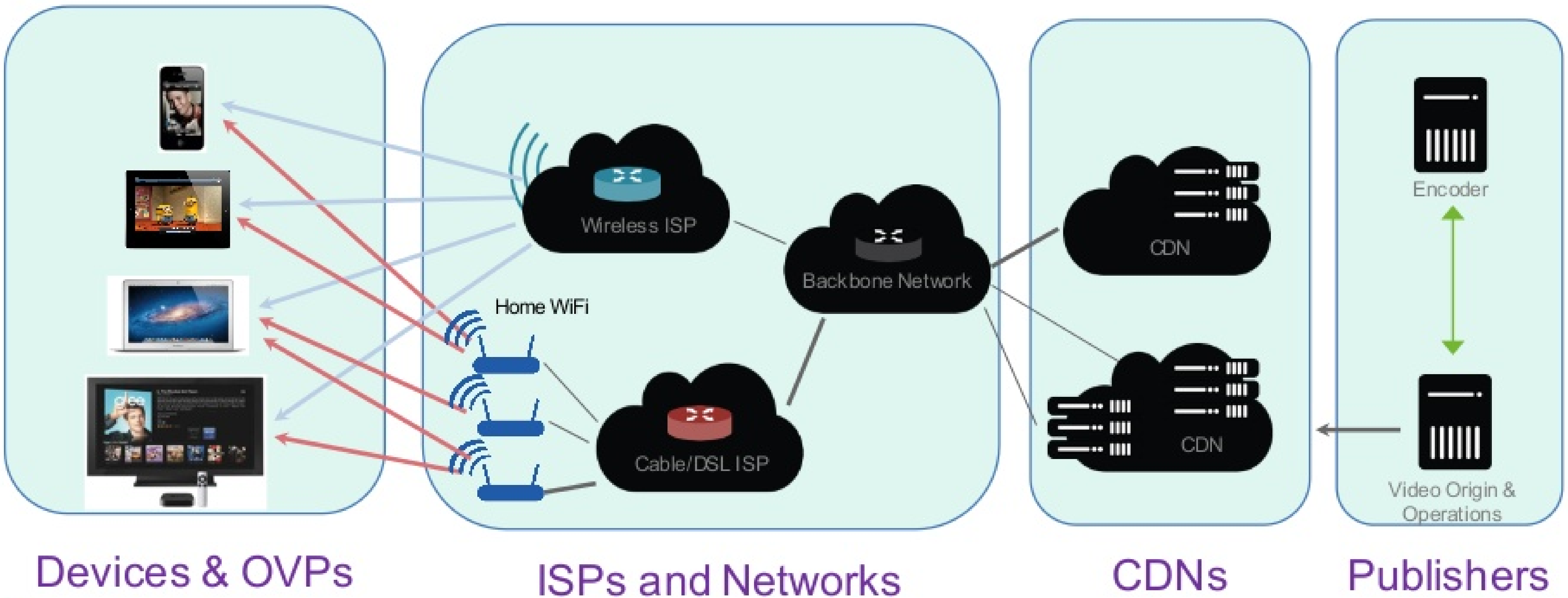
CONVIVA®

…is a video experience management platform that maximizes viewer engagement.

We provide quality metrics that give a comprehensive view into online video businesses.

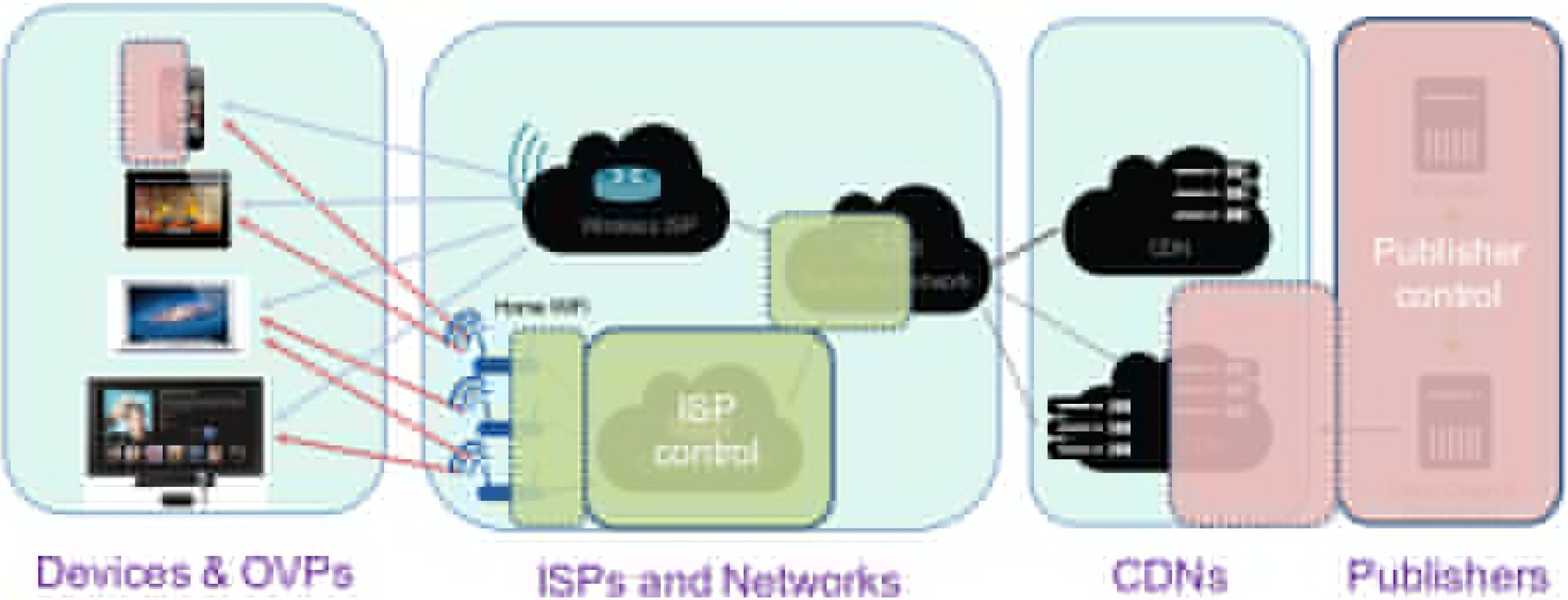
Internet Video Streaming is Hard

Many parties, many paths, but no E2E owner

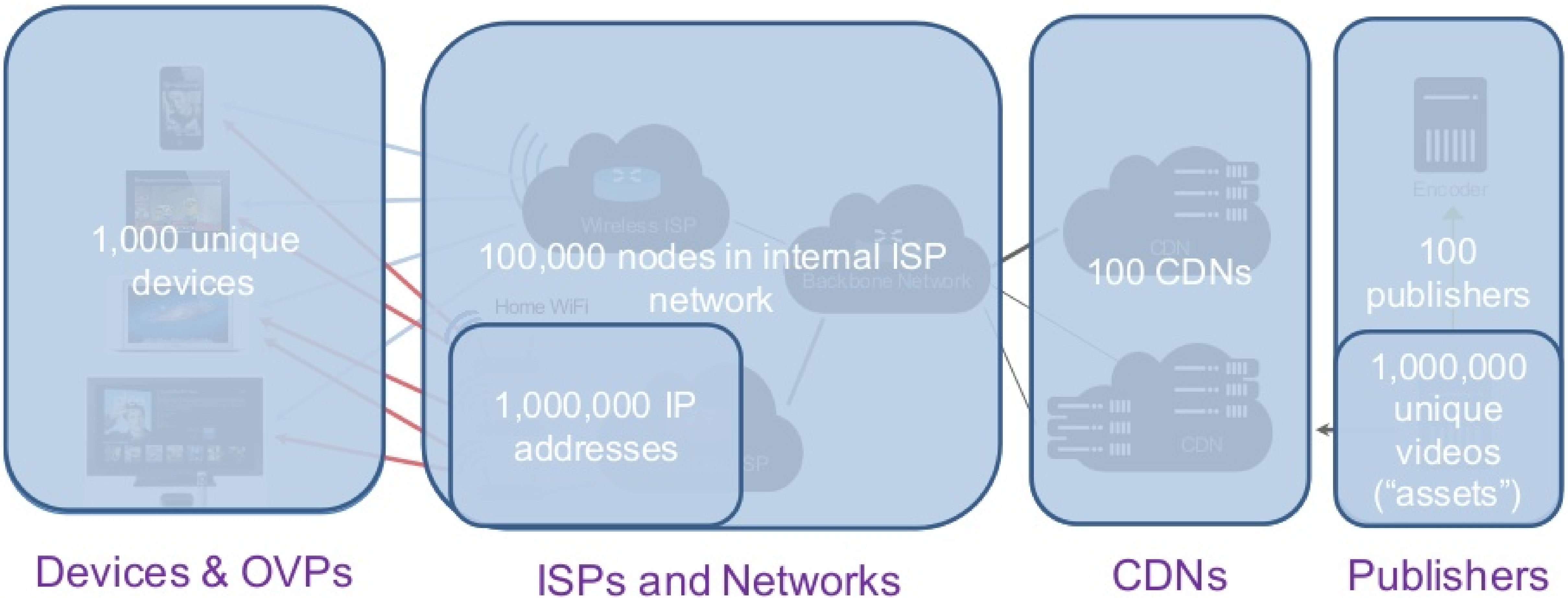


Internet Video Streaming is Hard

Many parties, many paths, but no E2E owner



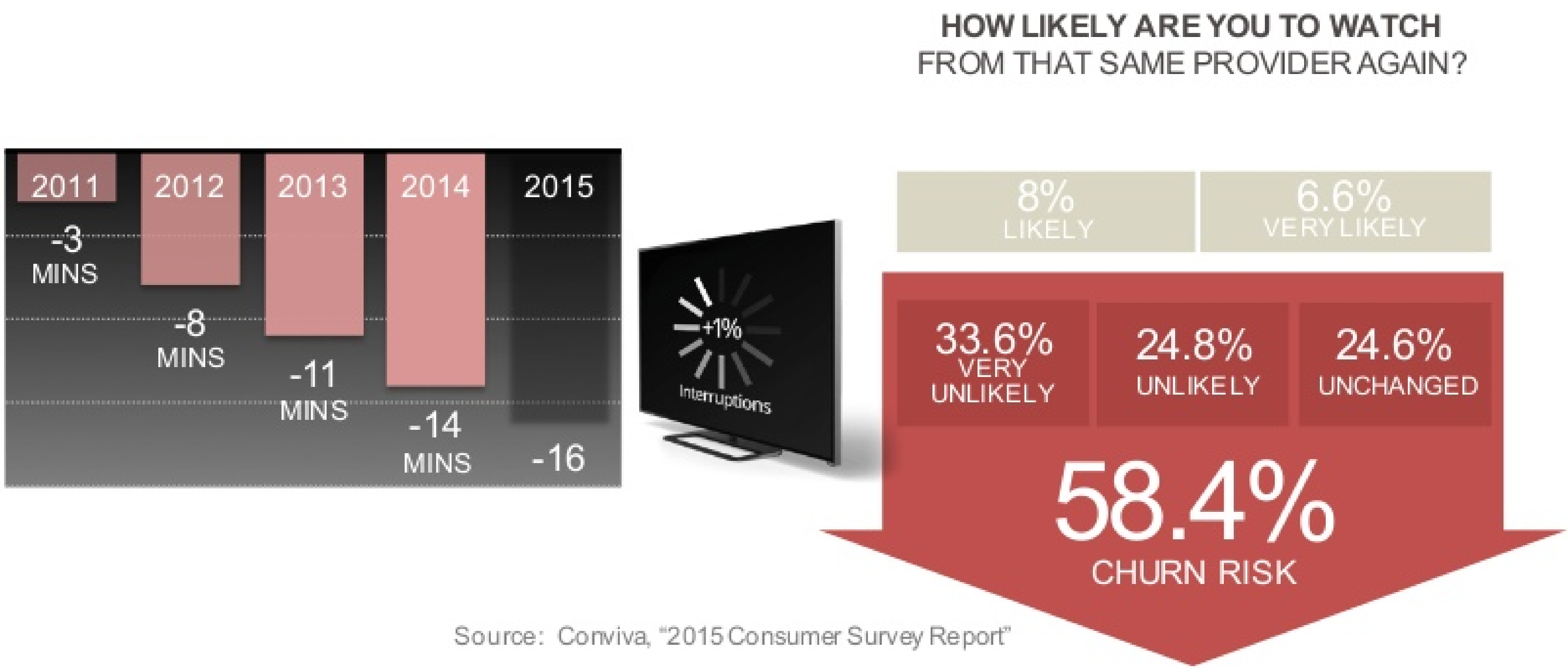
The ecosystem is big.



…in 10,000,000 video sessions from one week, at one US ISP

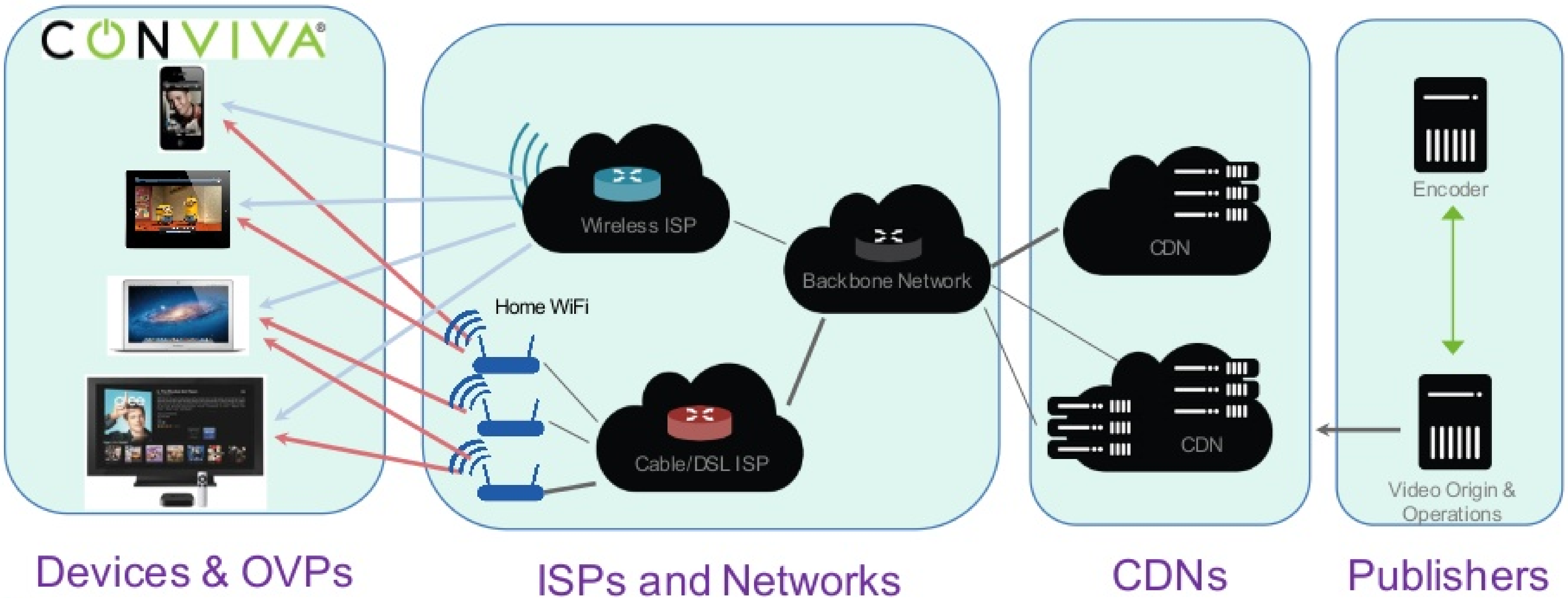
QoE is Critical to Engagement

For both video and advertisement businesses

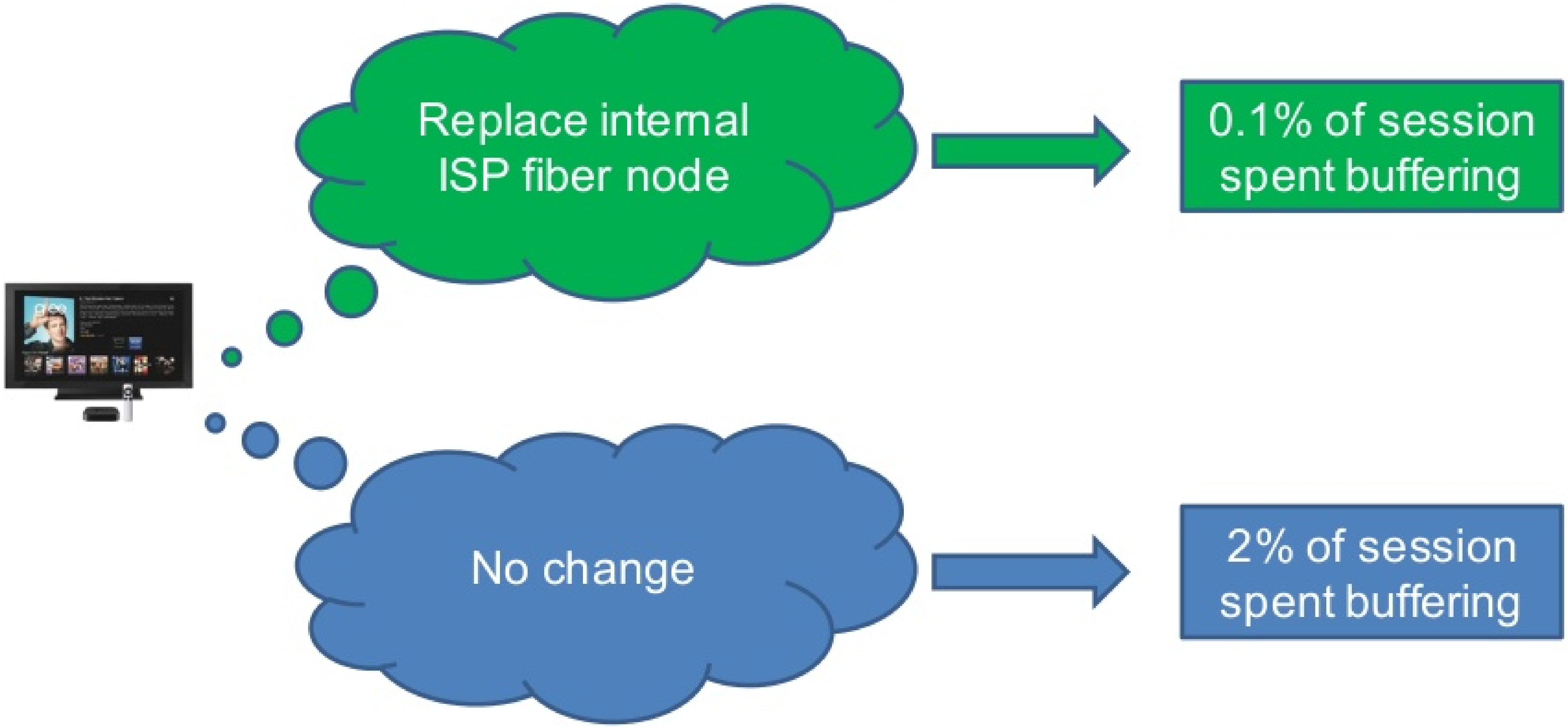


Source: Conviva, "2015 Consumer Survey Report"

Viewers are expecting TV-like quality (or better)


CONVIVA
Home WiFi
Wireless ISP
Backbone Network
Cable/DSL ISP
CDN
CDN
Encoder
Video Origin & Operations
Devices & OVPs
ISPs and Networks
CDNs
Publishers

Quality impact

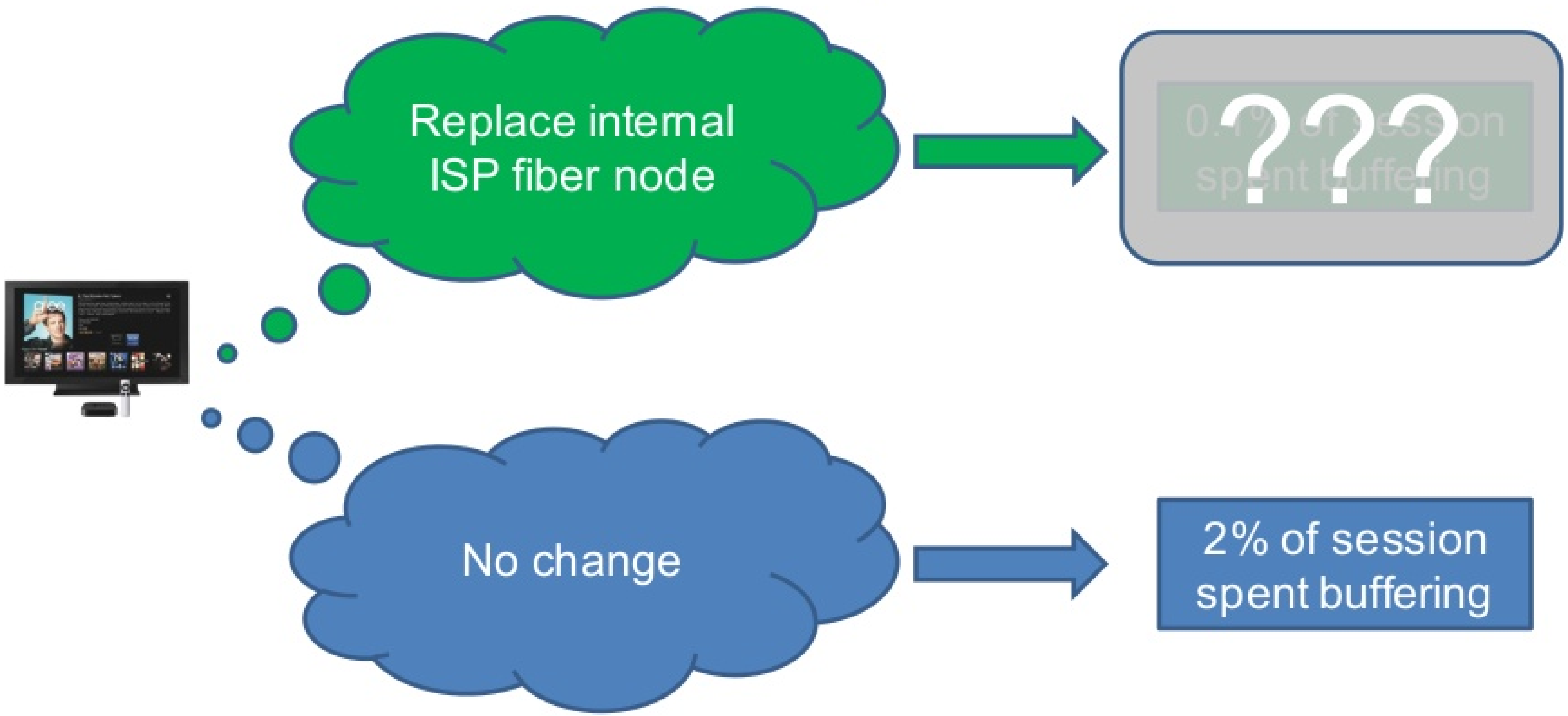


Quality impact

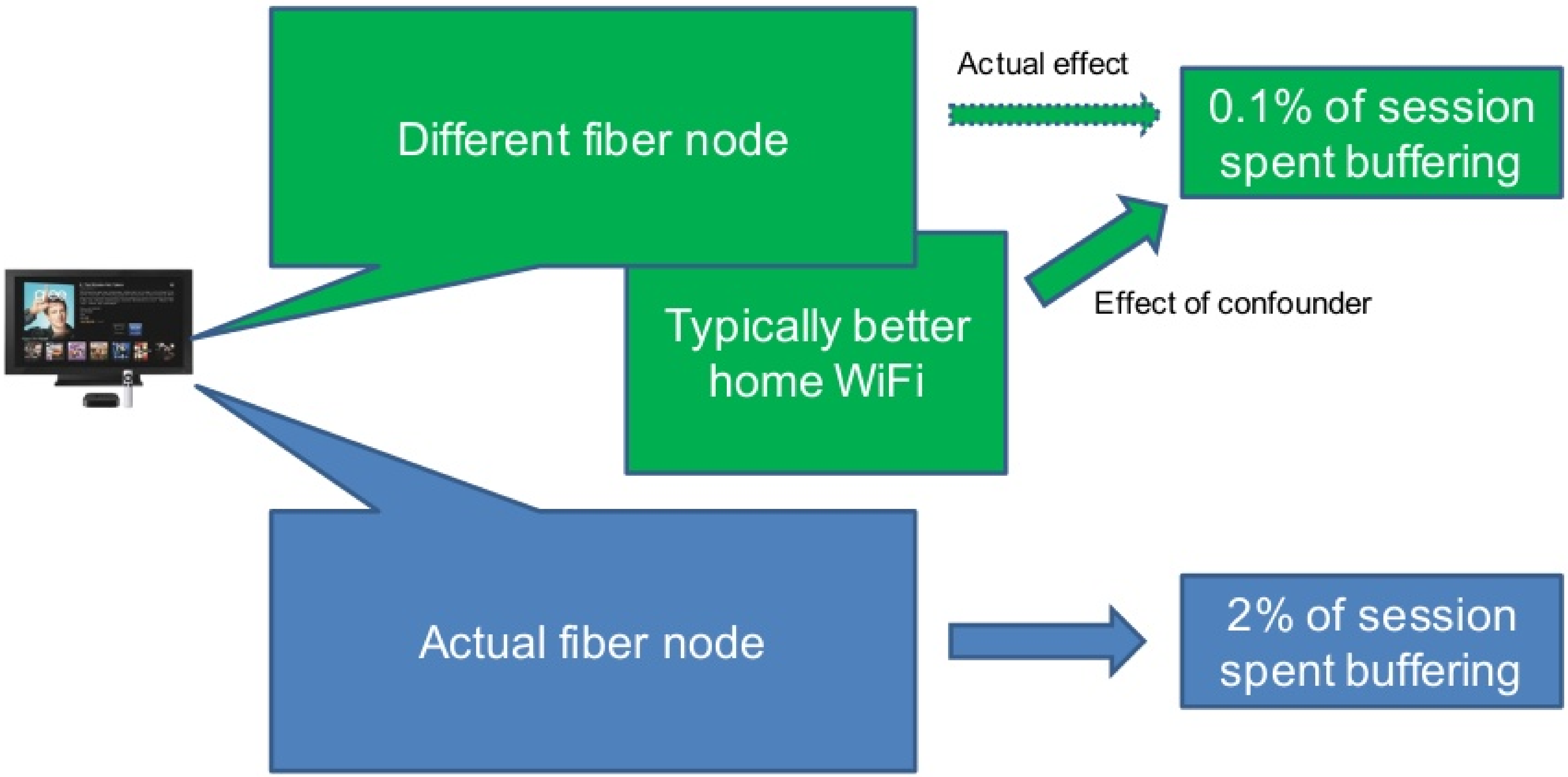
Quality impact of this fiber node on this session = 2% of session spent buffering - 0.1% of session spent buffering

Quality impact of this fiber node = $\sum$ [Quality impact of this fiber node on session 1, session 2, session 3]

Quality impact



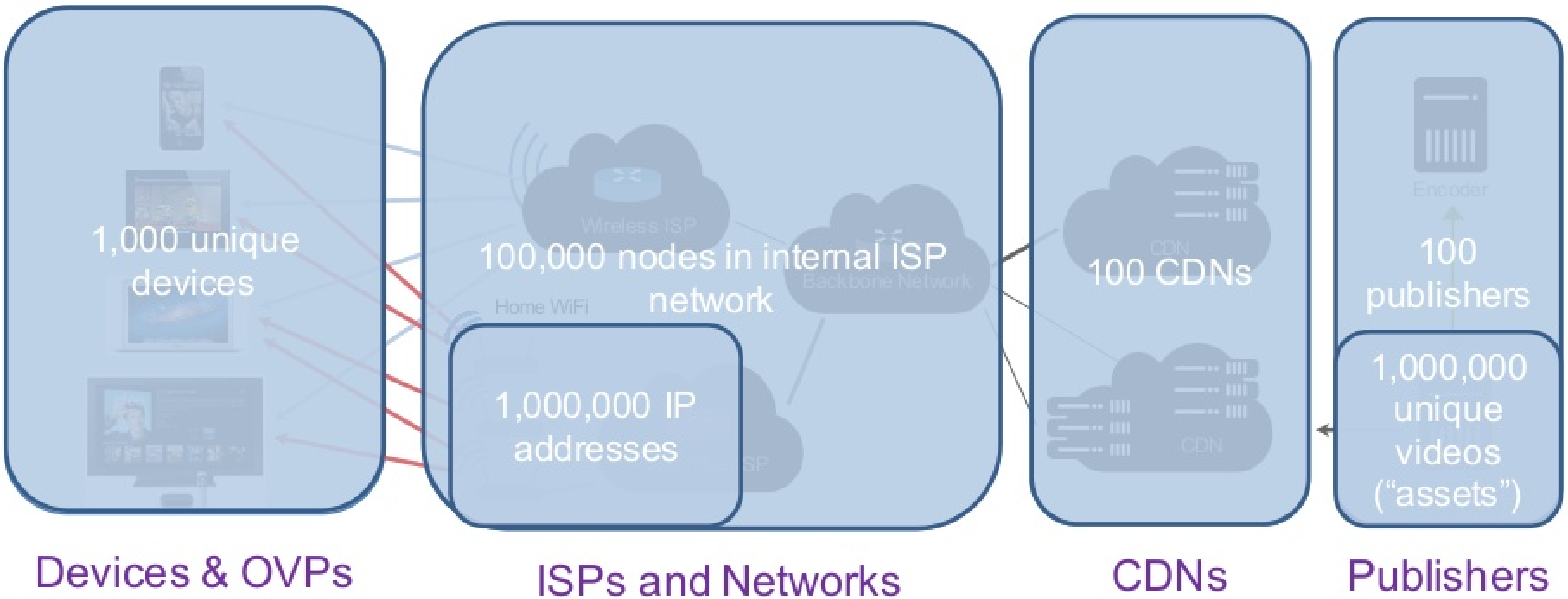
Confounding factors



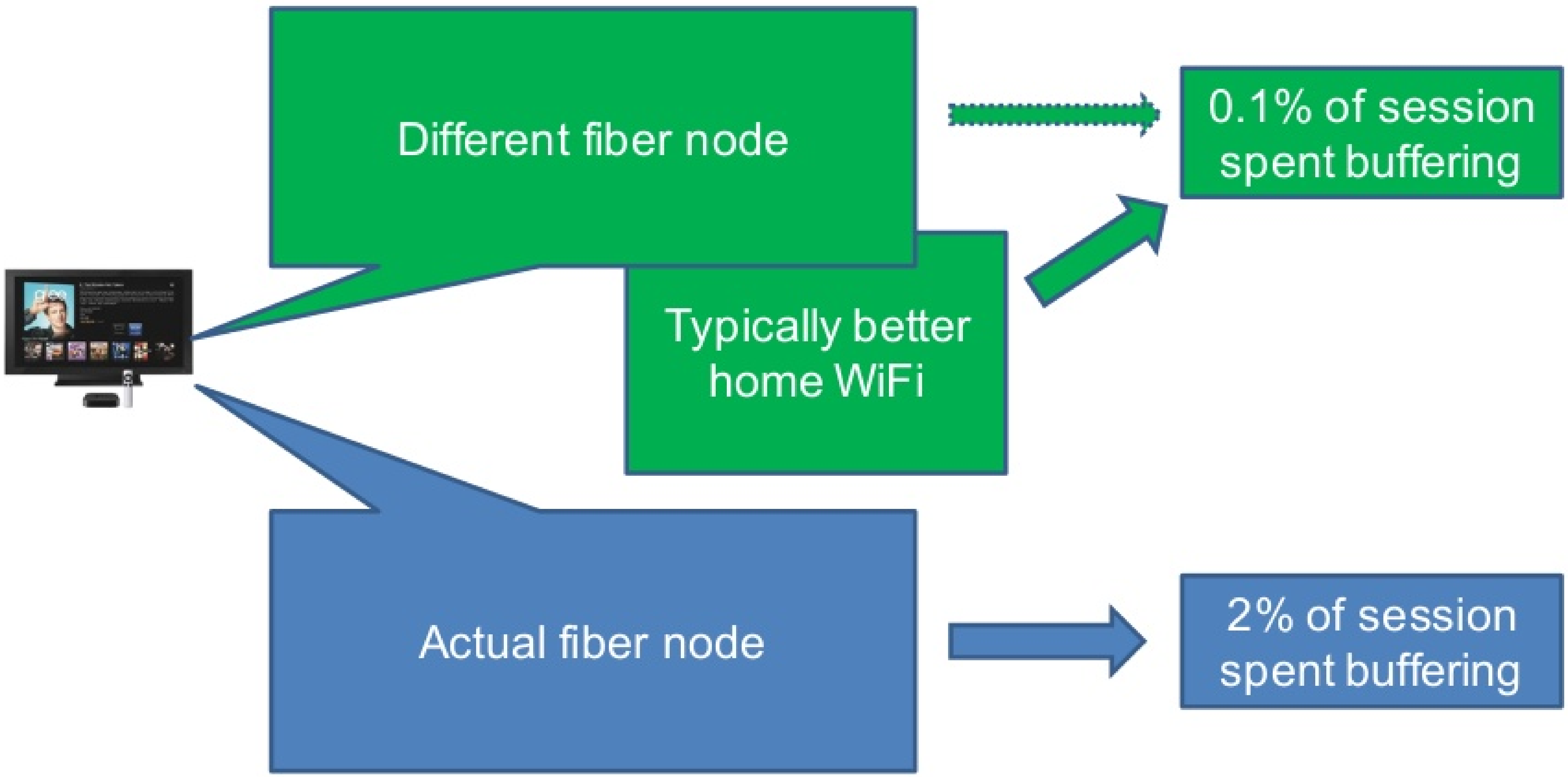
SPARK SUMMIT 2017

Many potentially-important factors

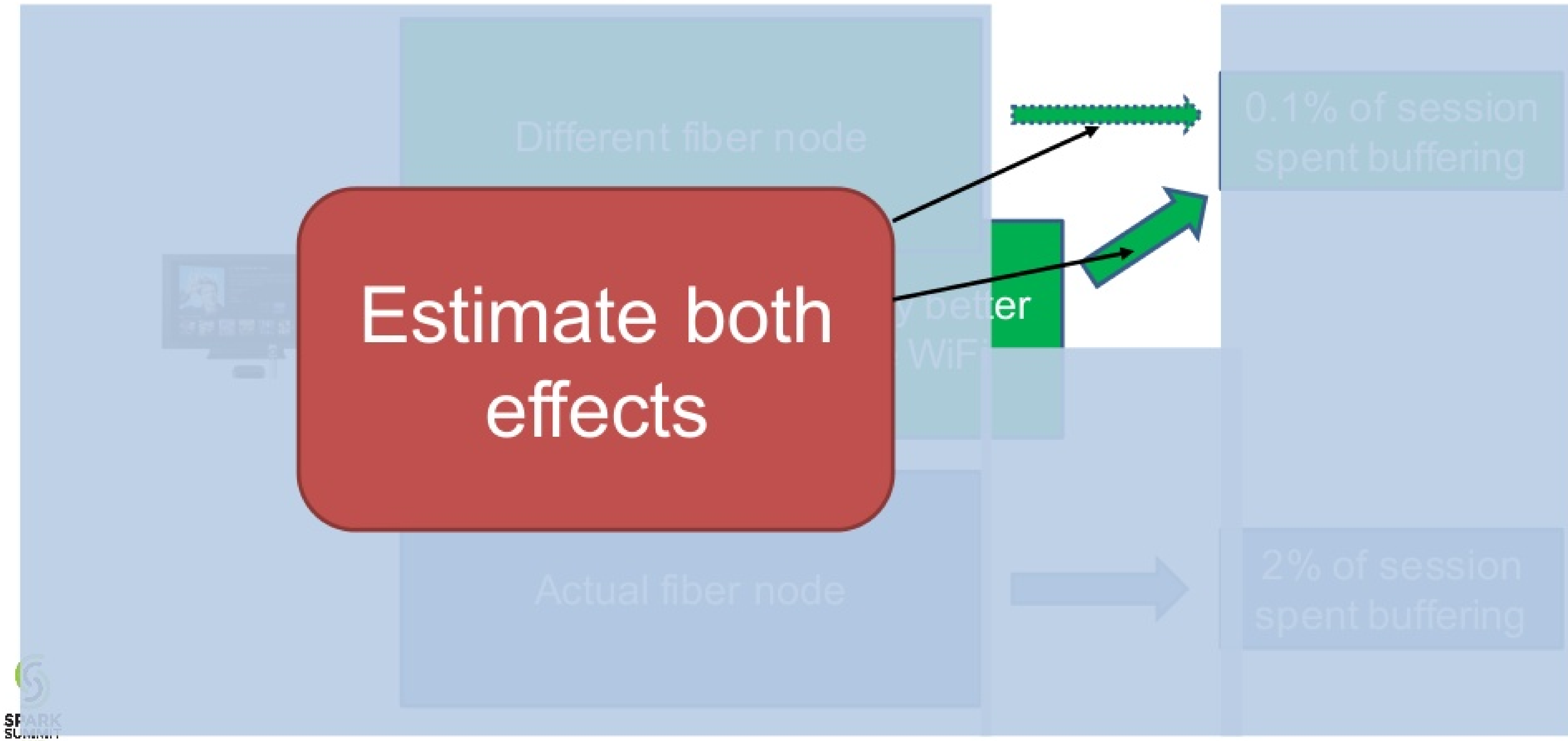
…and confounding factors!



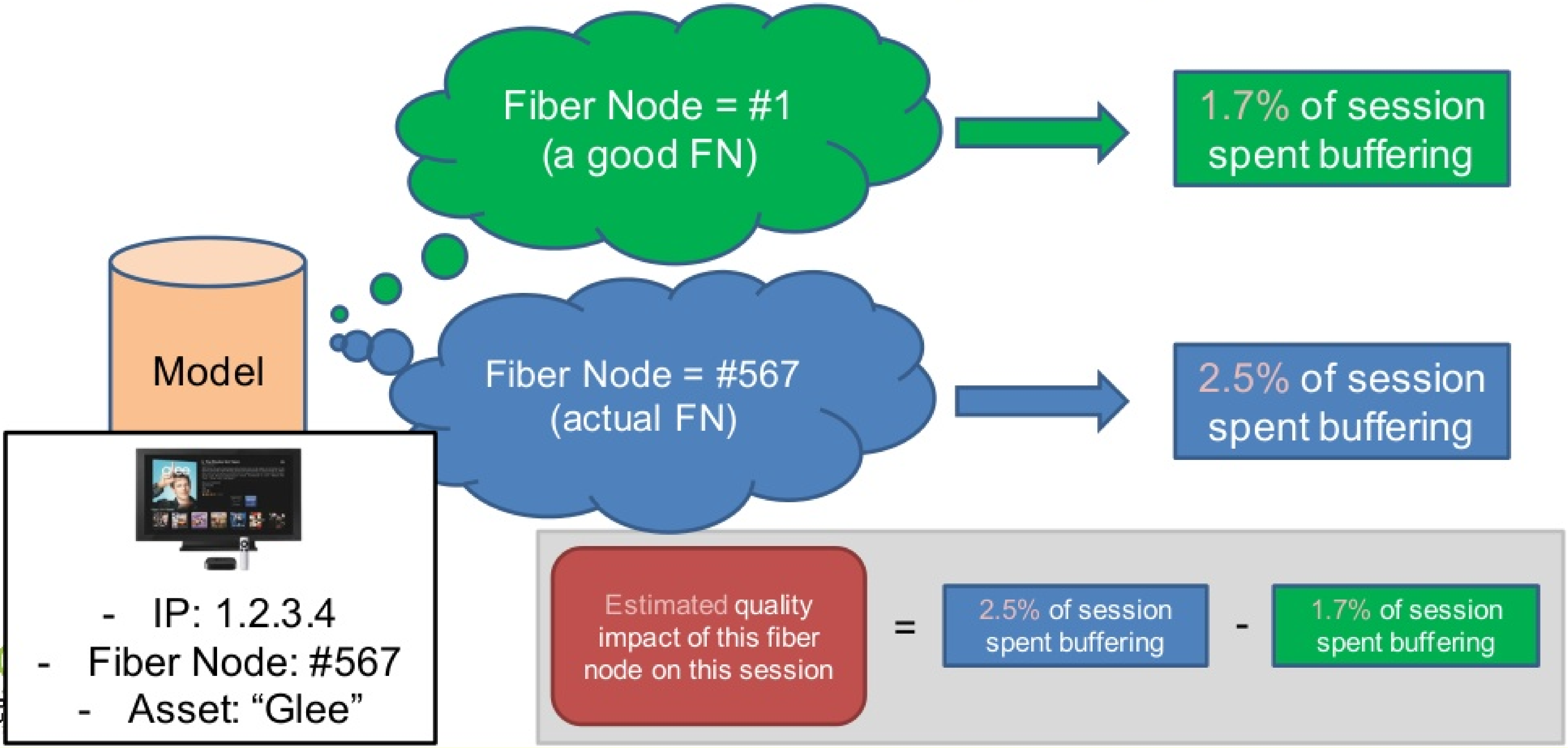
Fixing confounding



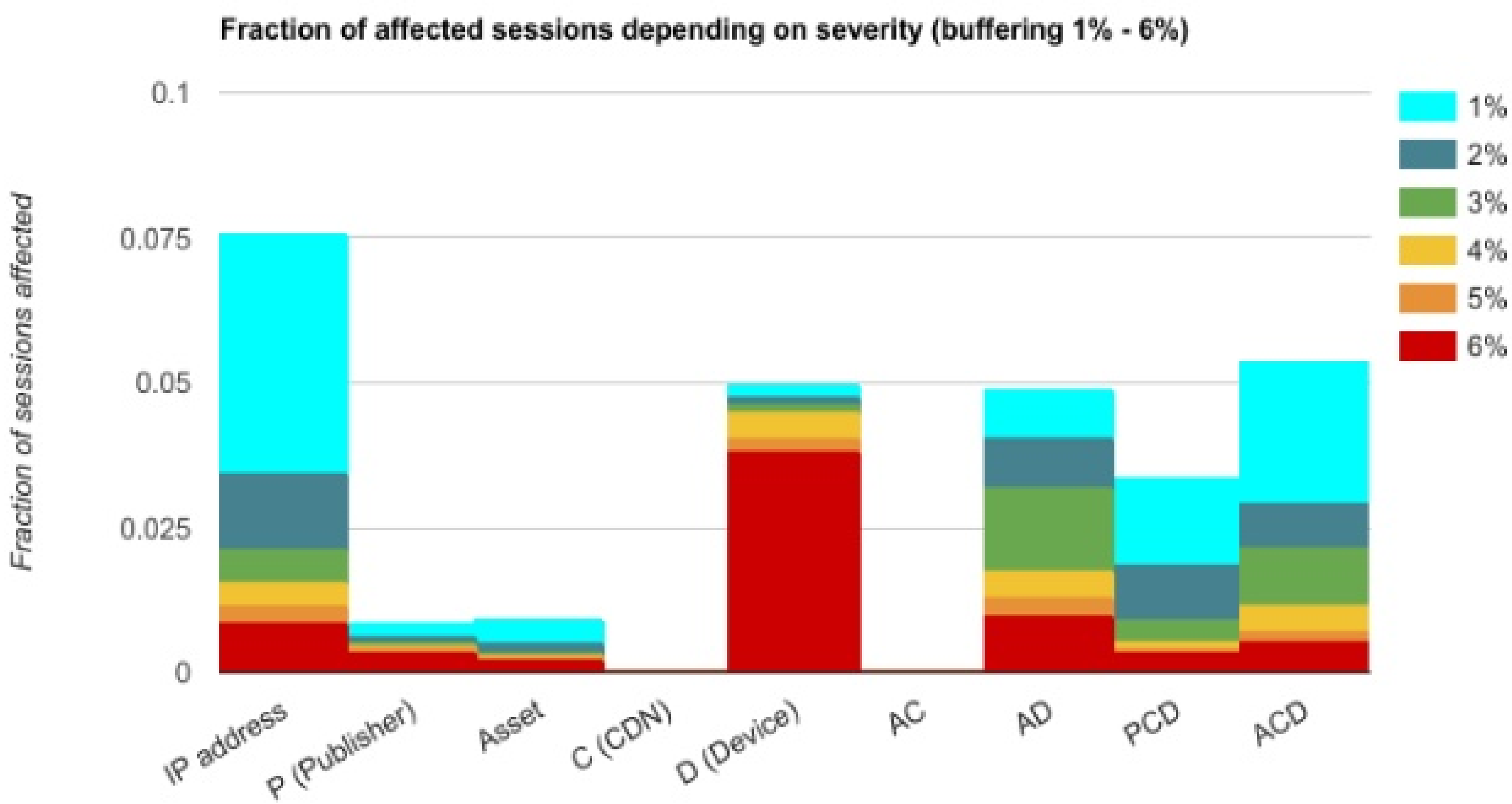
Fixing confounding



Estimating quality impact

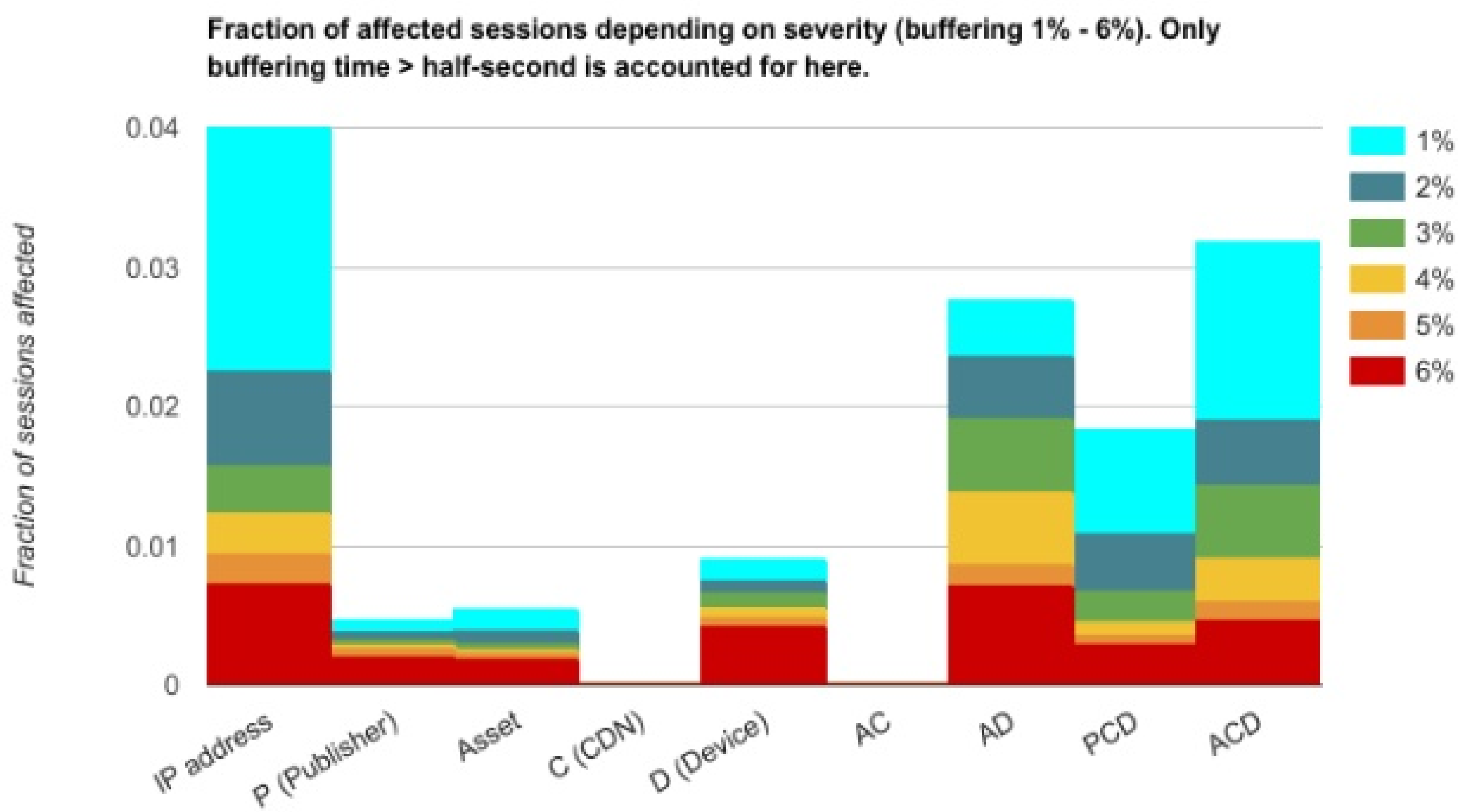


Devices produce short buffering

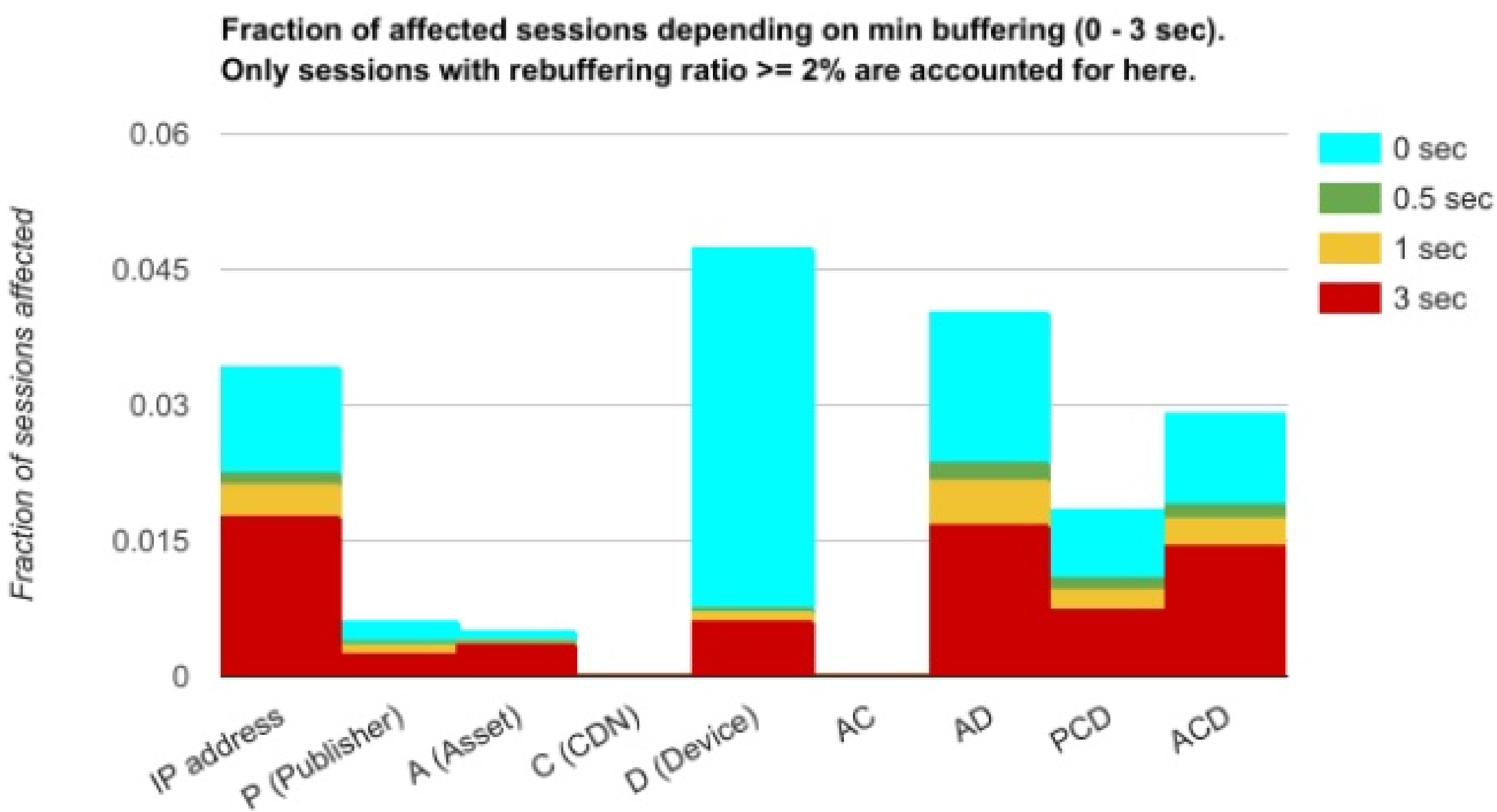


SPARK SUMMIT 2017

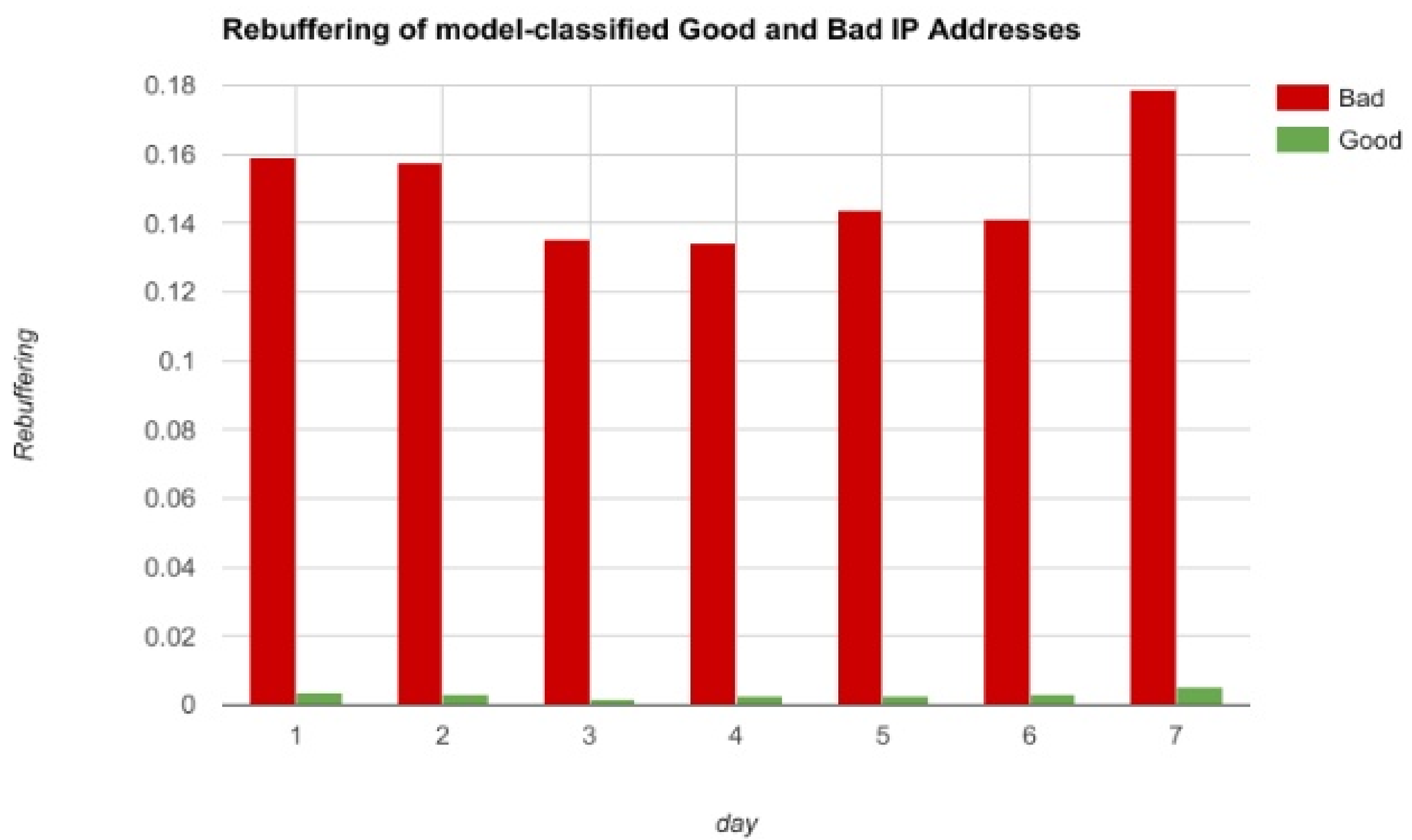
Devices produce short buffering



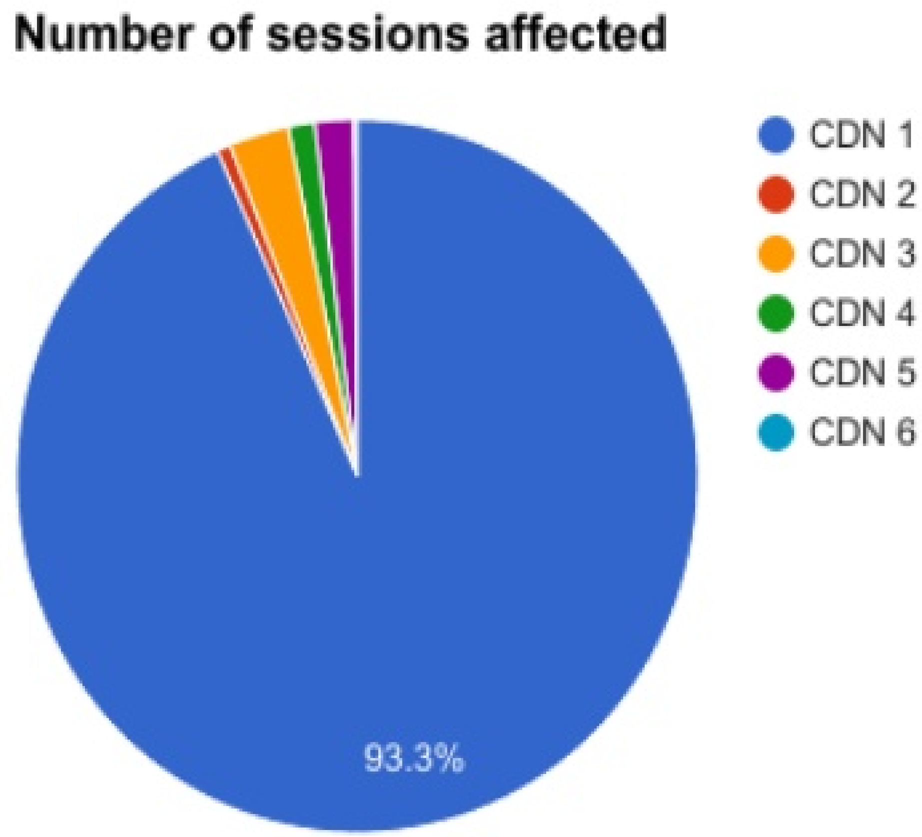
Device-level issues are responsible for most short buffering episodes



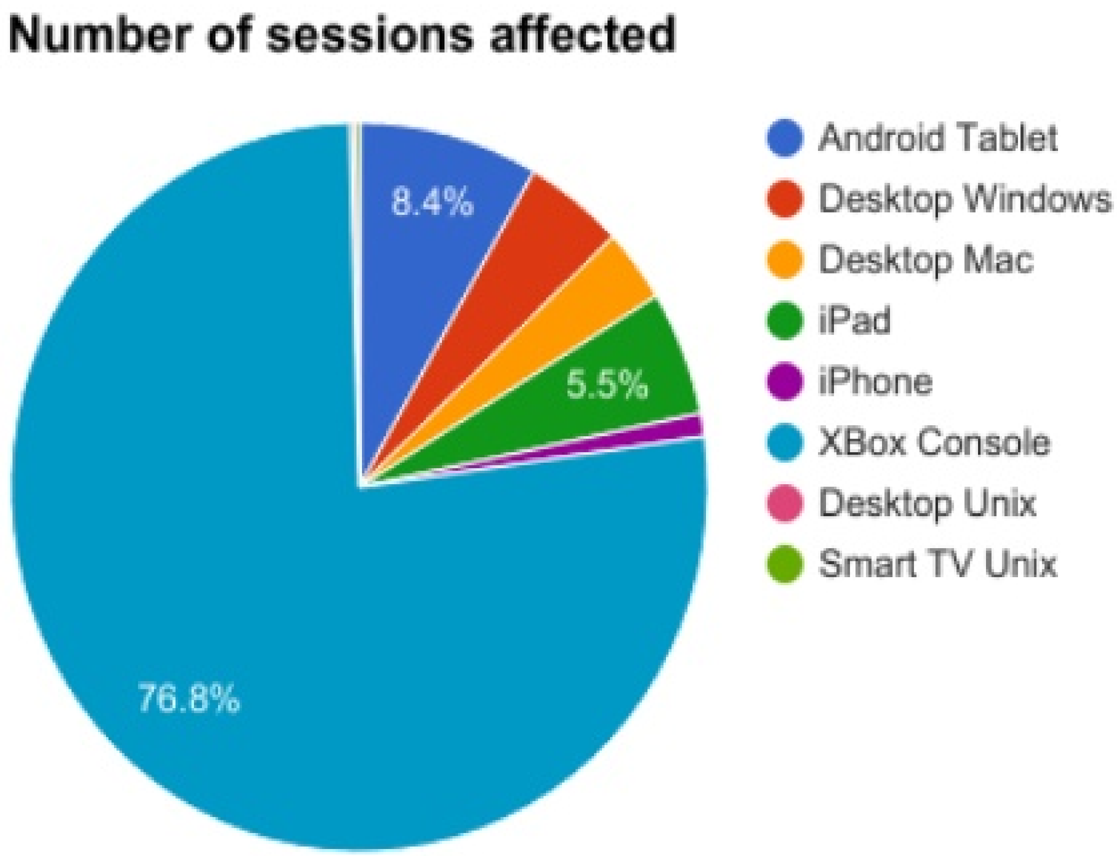
Good and bad IP addresses



From worst PCD, typical week:



From worst PCD, typical week:

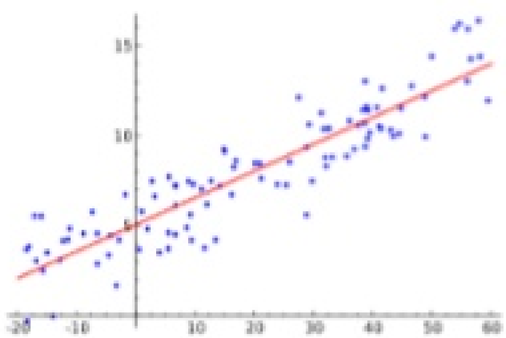


Worst assets, typical week:

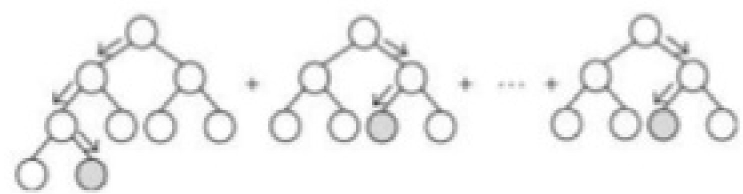
Only 6 really bad assets responsible for strong deterioration of their sessions regardless of other factors.

Of these, 4 are sport programs and 2 are obscure regularly scheduled foreign programs.

Modeling video quality



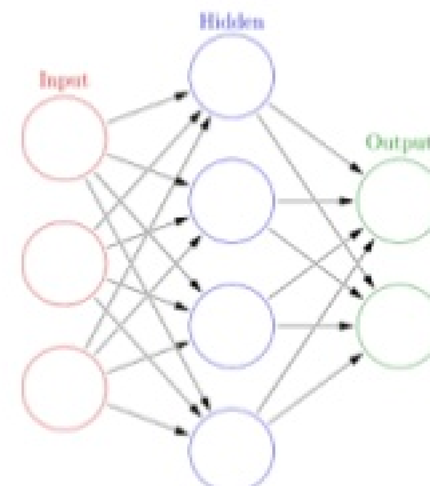
GLMs



random forests

boosted trees

Today's results



neural networks

Model for rebuffering ratio

Response: Rebuffering ratio (R)

(buffering time / (buffering time + playing time))

Categorical features:

IP address (I)
Fiber Node (N)
Service Group (S)
Publisher (P)
Asset (A)
CDN (C)
Device (D)
Live / VoD

More features:

Time, day of week, asset length.

Time or no time?

Time is strongest or one of strongest features.

Big game, popular show – many sessions suffer, regardless of IP and device or even CDN.

Time makes model more precise.
But: time steals effect from big nodes such as Asset and Publisher.

Similar question: Bitrate or no bitrate?

Model versions, practical issues

Preference: Spark cluster of 10-30 nodes, each node 4-8 cores and 30GB memory. Hopefully 1-3 hours to process 1-3 weeks of data of big ISP, whole US.

Nodes as embedded features. Boosted trees on Spark. Whole US. Learn from 1-3 weeks, apply to last week.

Nodes as one-hot encoding. Random forest on Spark. Each geographical area is processed separately.

Model versions, practical issues

Nodes as a trainable embedding first layer. Neural network on single Spark node with Tensorflow. Each geographical area separately.

Embedding size is same for all kinds of nodes.

Embedding size is ~ log of node's dictionary size.

One or two hidden layers above the embedding layer.

Adagrad or any other Ada-like optimizer performs better than no-Ada.

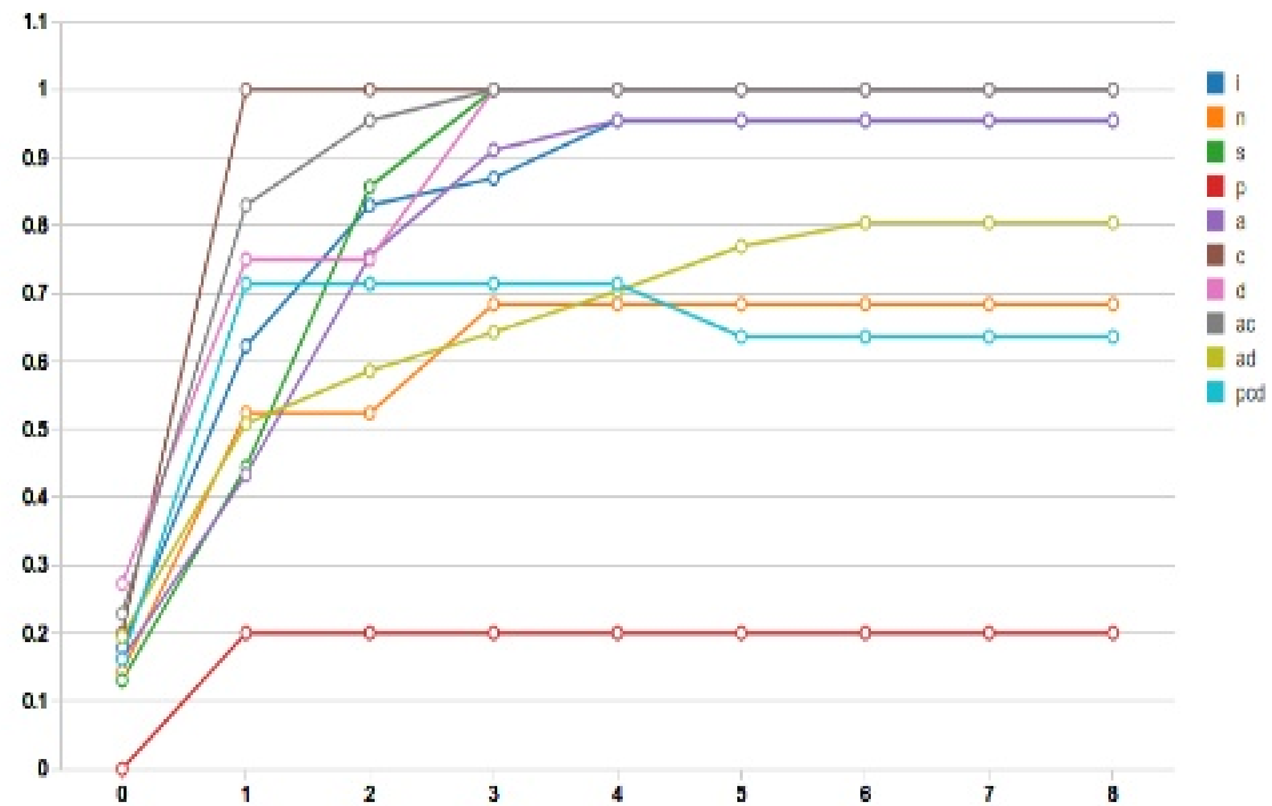
Slow improvements beyond fast mediocre accuracy.

Finding a "good" node

Effect of a node = Model(real features) - Model(features with good node)

In case of trainable embedding – iterative finding of good nodes:

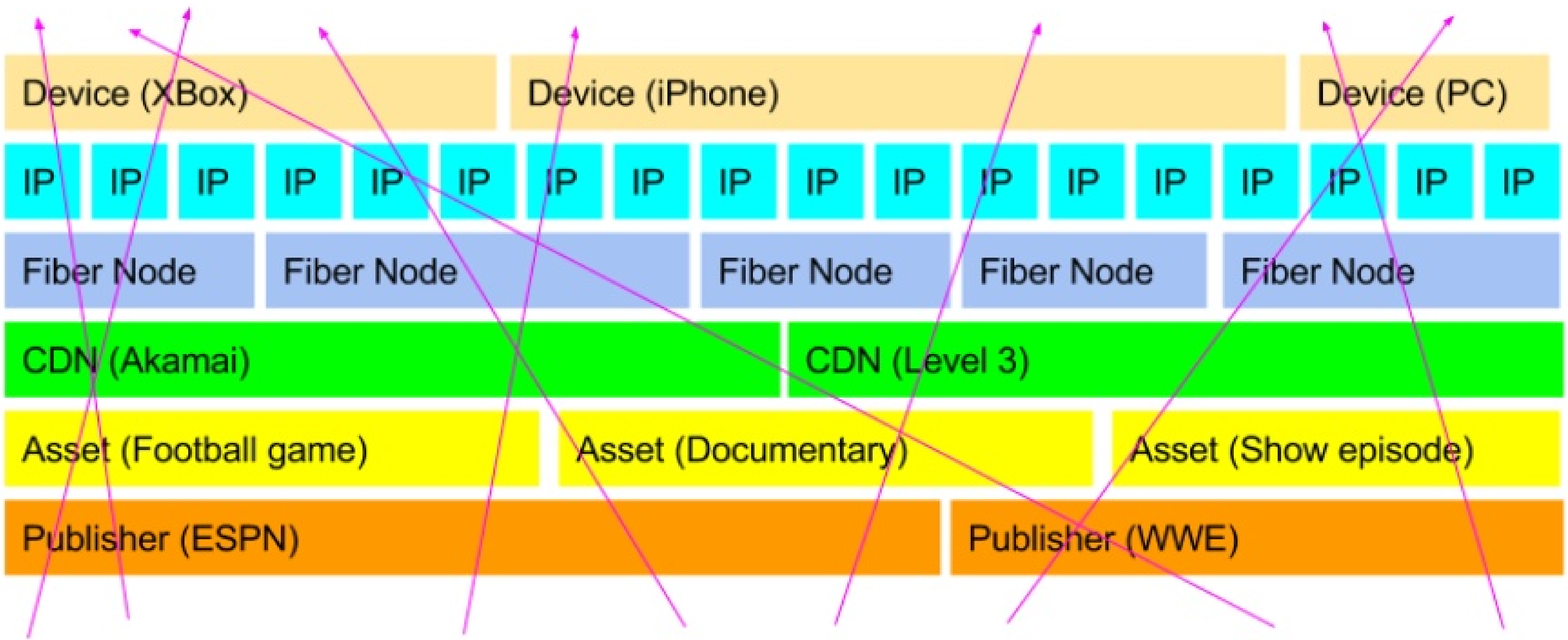
good = nodes with lowest avg label
while set of good nodes is changing:
  for each session:
    Effect = Model(real) - Model(good)
  good = nodes with lowest avg effect



Overlap of set of good nodes with the set from previous iteration - typically stabilizes in 2-3 iterations.

Sessions are affected by ...

Video sessions. Each has buffering >=0, from all involved nodes.



If effect is linear, this would be like ART in 3D tomography

Thanks to Spark and Databricks for making it easy for us

Thank You.

Contact information:

ovasilyev@conviva.com
hmilner@conviva.com

We Are Hiring!

www.conviva.com/our-team/careers

SPARK SUMMIT 2017