A COMPARATIVE STUDY OF DASH REPRESENTATION SETS USING REAL USER CHARACTERISTICS

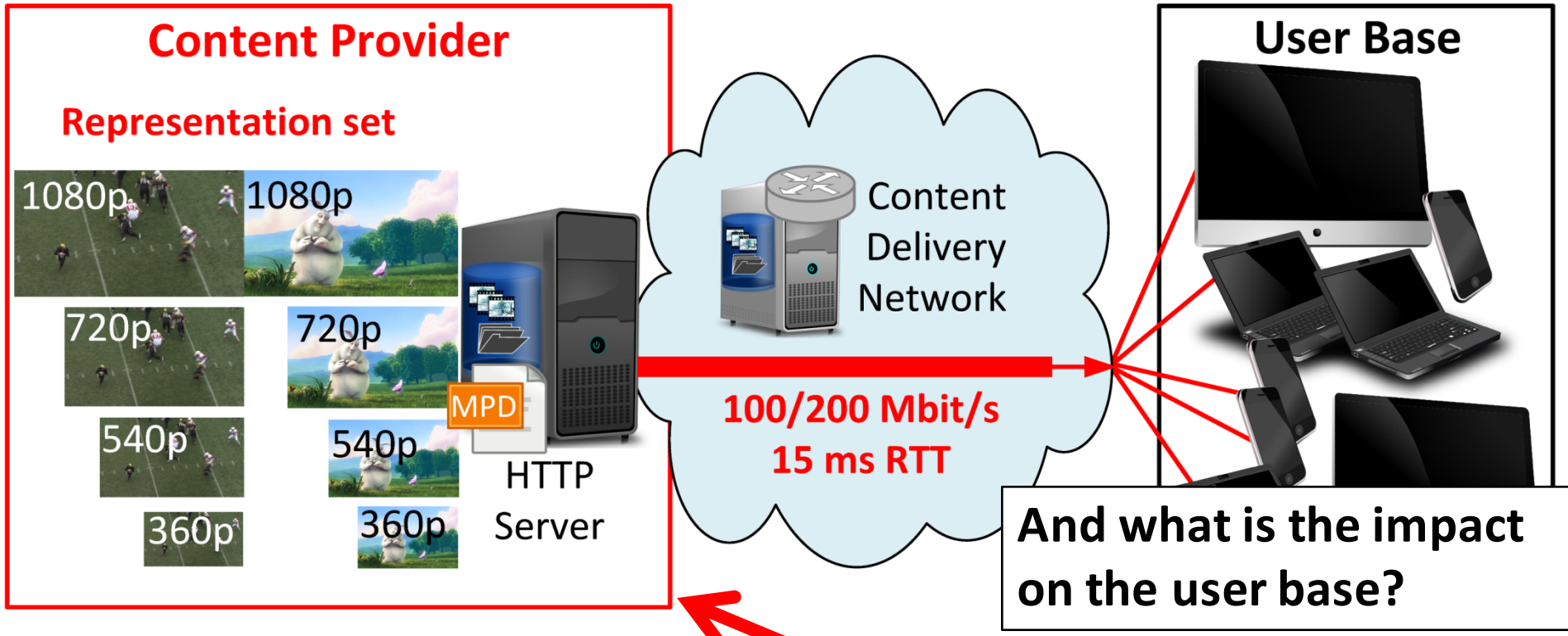
C. Kreuzberger, B. Rainer, H. Hellwagner
Institute of Information Technology (ITEC)
Faculty of Technical Sciences (TEWI)
Alpen-Adria-Universität Klagenfurt (AAU), Austria

L. Toni, P. Frossard
Signal Processing Laboratory (LTS4)
École polytechnique fédérale de Lausanne (EPFL), Switzerland

VIDEO STREAMING AND DASH

- Rapidly growing (> 60% of traffic at peak times)
- DASH: videos with various bitrates and resolutions (**representations**) on a per-segment basis
- Overall **goal** in video streaming (from the user‘s perspective): **maximize** Quality of Experience **(QoE) by chosing representations** based on restrictions of the **user** (e.g., bandwidth, spatial resolution, …) and network conditions

MULTIMEDIA STREAMING SCENARIO



And what is the impact on the user base?

How to decide which representation sets should be made available on the server side?

DASH Representations in the Wild

Name	Resolution	Bit Rates [kbit/s]
YouTube	1080p (1920x1080)	4,072
	720p (1280x720)	2,168
	540p (960x540)	1,109
	360p (640x360)	110 247 606
Netflix	1080p (1920x1080)	4,300 5,800
	720p (1280x720)	2,350 3,000
	540p (960x540)	1,050 1,750
	360p (640x360)	235 375 560 750
Apple	1080p (1920x1080)	11,000 24,000 39,000
	720p (1280x720)	2,500 4,500
	540p (960x540)	1,800
	360p (640x360)	110 200 400 600 1,200

Table 3: Summary of recommended representation sets from YouTube (experiment), Netflix [9], and Apple [3].

Parsing MPD of 51k videos

Netflix Blog

Tech. Docu.

OPTIMIZED DASH REPRESENTATIONS

$$\max_{\{\tau,\alpha,\beta,\gamma\}} \sum_{u\in\mathcal{U}}\sum_{v\in\mathcal{V}}\sum_{r\in\mathcal{R}}\sum_{s\in\mathcal{S}} f_{uvrs}\cdot\tau_{uvrs} \tag{1a}$$

$$\text{such that } \tau_{uvrs} \le \alpha_{uvrs}, \quad u\in\mathcal{U}, v\in\mathcal{V}, r\in\mathcal{R}, s\in\mathcal{S} \tag{1b}$$

$$\alpha_{uvrs} \le \beta_{vrs}, \quad u\in\mathcal{U}, v\in\mathcal{V}, r\in\mathcal{R}, s\in\mathcal{S} \tag{1c}$$

$$\beta_{vrs} \le \sum_{u\in\mathcal{U}} \alpha_{uvrs}, \quad v\in\mathcal{V}, r\in\mathcal{R}, s\in\mathcal{S} \tag{1d}$$

$$\sum_{v\in\mathcal{V}}\sum_{s\in\mathcal{S}}\sum_{\substack{r'\in\mathcal{R}\\ r'\ge r}} \tau_{uvr's} \le T_{ur}, \quad u\in\mathcal{U}, r\in\mathcal{R} \tag{1e}$$

$$\sum_{r\in\mathcal{R}}\sum_{s\in\mathcal{S}} \tau_{uvrs} \le \begin{cases} 1, & \text{if } v = v_u \\ & \& \; s\in\{s_u-1, s_u, s_u+1\} \\ 0, & \text{otherwise} \end{cases} \quad u\in\mathcal{U}, v\in\mathcal{V} \tag{1f}$$

$$(b_{vs}^{\min} - b_r)\cdot\tau_{uvrs} \le 0, \quad u\in\mathcal{U}, v\in\mathcal{V}, r\in\mathcal{R}, s\in\mathcal{S} \tag{1g}$$

$$(b_r - b_{vs}^{\max})\cdot\tau_{uvrs} \le 0, \quad u\in\mathcal{U}, v\in\mathcal{V}, r\in\mathcal{R}, s\in\mathcal{S} \tag{1h}$$

$$\sum_{u\in\mathcal{U}}\sum_{v\in\mathcal{V}}\sum_{r\in\mathcal{R}}\sum_{s\in\mathcal{S}} b_r\cdot\tau_{udvrs} \le C\cdot|\mathcal{U}|, \tag{1i}$$

$$\sum_{v\in\mathcal{V}}\sum_{r\in\mathcal{R}}\sum_{s\in\mathcal{S}} \beta_{vrs} \le K, \tag{1j}$$

$$\sum_{u\in\mathcal{U}} \gamma_u \ge P\cdot|\mathcal{U}|, \tag{1k}$$

$$\sum_{v\in\mathcal{V}}\sum_{r\in\mathcal{R}}\sum_{s\in\mathcal{S}} \tau_{uvrs} \ge T_{\min}\cdot\gamma_u, \quad u\in\mathcal{U} \tag{1l}$$

$$\tau_{uvrs}\in[0,1], \quad u\in\mathcal{U}, v\in\mathcal{V}, r\in\mathcal{R}, s\in\mathcal{S} \tag{1m}$$

$$\alpha_{uvrs}\in\{0,1\}, \quad u\in\mathcal{U}, v\in\mathcal{V}, r\in\mathcal{R}, s\in\mathcal{S} \tag{1n}$$

$$\beta_{vrs}\in\{0,1\}, \quad v\in\mathcal{V}, r\in\mathcal{R}, s\in\mathcal{S} \tag{1o}$$

- **Integer Linear Program (ILP)**
- Problems:
  - Mainly theoretical results
  - Static number of users

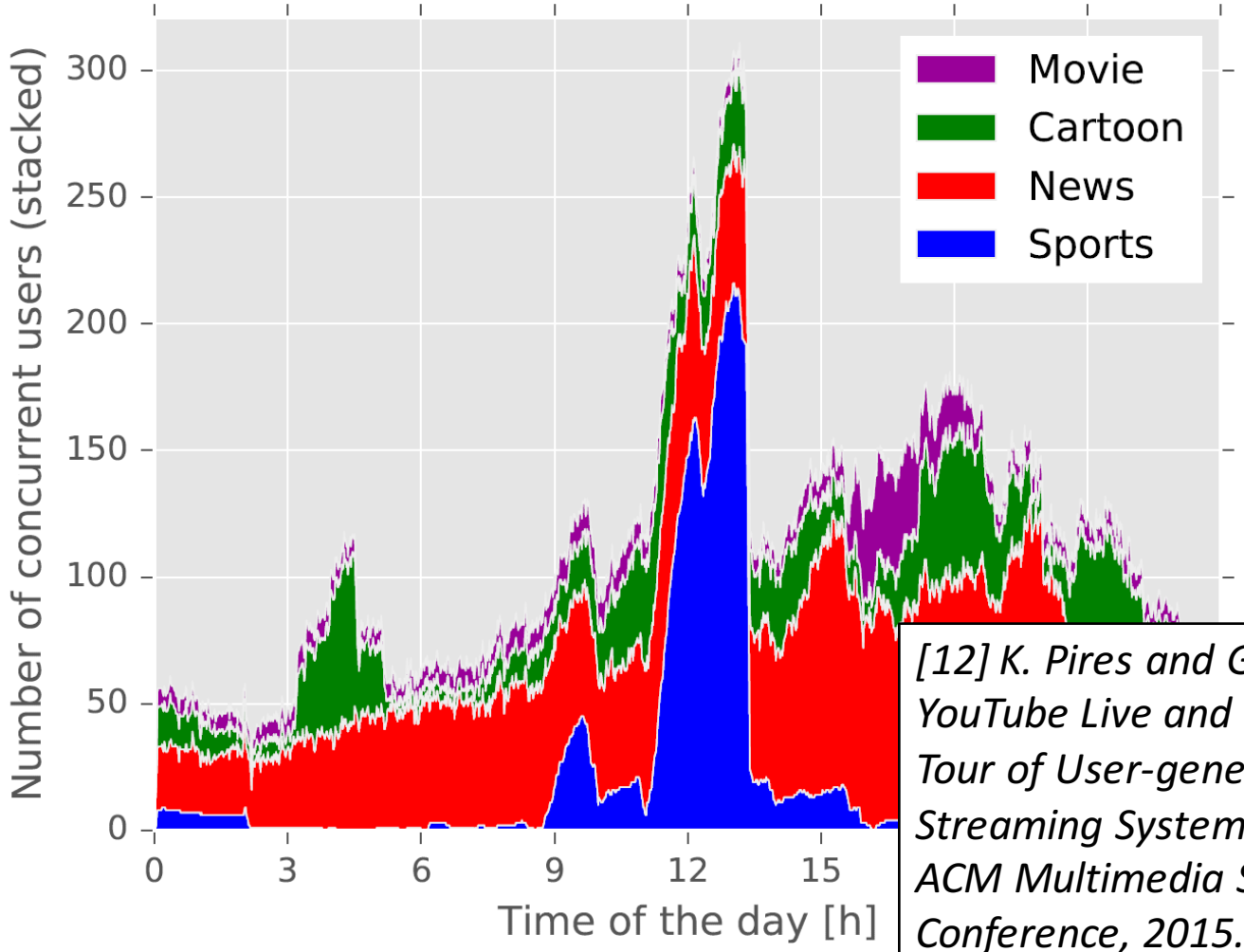
[16] Toni et al., „Optimal Selection of Adaptive Streaming Representations", ACM Transactions on Multimedia Computing Communications and Applications, 2015

Our Approach

- Evaluate optimized and existing DASH representations
  - by modelling **dynamic** user behaviour (**join/leave**),
  - with realistic **device** and **network characteristics**,
  - and conducting **extensive simulations** using NS-3.

- Metrics
  - Average User Satisfaction [SSIM per screen resolution]
  - Average Goodput [kbit/s]

DYNAMIC USER BEHAVIOUR



[12] K. Pires and G. Simon. YouTube Live and Twitch: A Tour of User-generated Live Streaming Systems. In Proc. 6th ACM Multimedia Systems Conference, 2015.

DEVICE AND NETWORK CHARACTERISTICS

Device Type (Connection)	Screen Res.	c_{min}	c_{max}	p
Smartphone (3G, WiFi)	360p, 540p	0.4	4	21.4%
Tablet (3G, WiFi)	540p, 720p	0.4	4	14.8%
Laptop (ADSL)	720p, 1080p	0.7	10	32.1%
HDTV (FTTH, Cable)	720p, 1080p	1.5	25	31.7%

Table 1: Devices with available screen resolutions and min/max link capacities (c_{min}/c_{max}) expressed in Mbit/s. p denotes the distribution of those devices

[10] Nielsen Research, "Binging" is the New Viewing for Over-the-top Streamers, 2013

[16] Toni et al., „Optimal Selection of Adaptive Streaming Representations", ACM Transactions on Multimedia Computing Communications and Applications, 2015

Optimized Representations

Video Id	Resol.	C100M-K24 kbit/s	C100M-K44 kbit/s
1	1080p	586	387 669
	720p	-	344 606
	540p	709	709
	360p	297 375	297 375 558
2	1080p	619 745 1190	526 619 745 1,042 1,380
	720p	297 534 676 1,093	297 370 534 676 777 1,093 1,361
	540p	173 407 529 747	329 529 620 747 1,242
	360p	315 568	220 315 568
3	1080p	-	819
	720p	761	533 761
	540p	553	320 553 785
	360p	245	245 595
4	1080p	-	-
	720p	-	1448
	540p	669 1,081	570 669 798 1,081
	360p	289 561	289 360 561

Table 5: Optimized representation sets [16], exemplified for $C = 100$ Mbit/s, $K = 24$ and $K = 44$ representations.

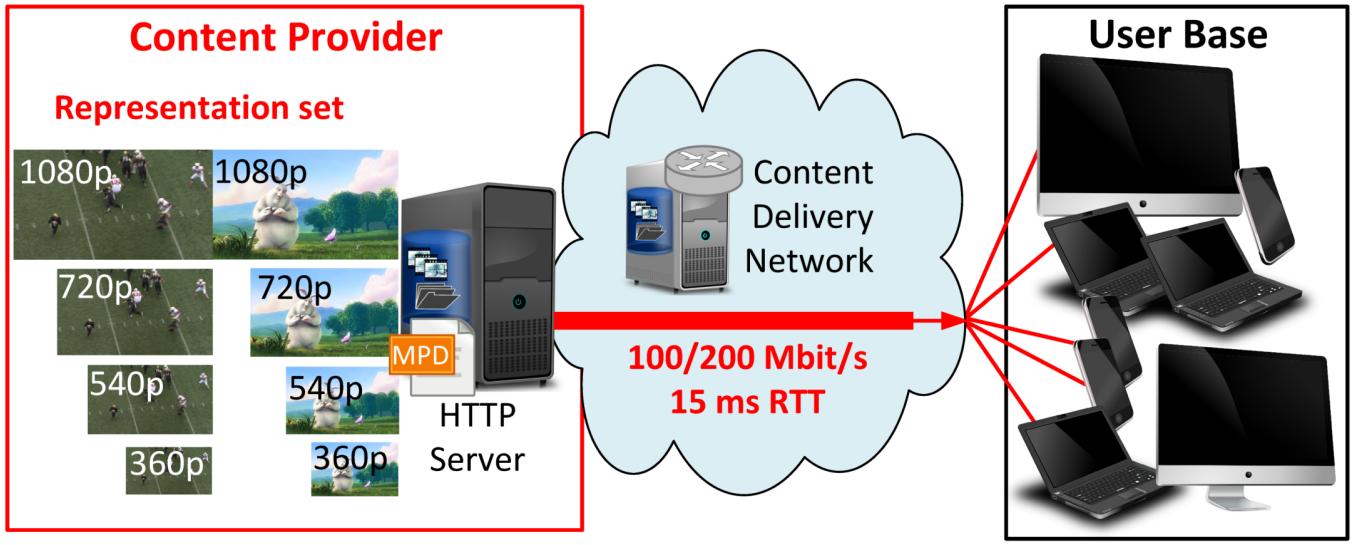
- Computed using ILP [16] with device/network characteristics and dynamic user behaviour (previous two slides)

Upper Bound: Max-Min Model

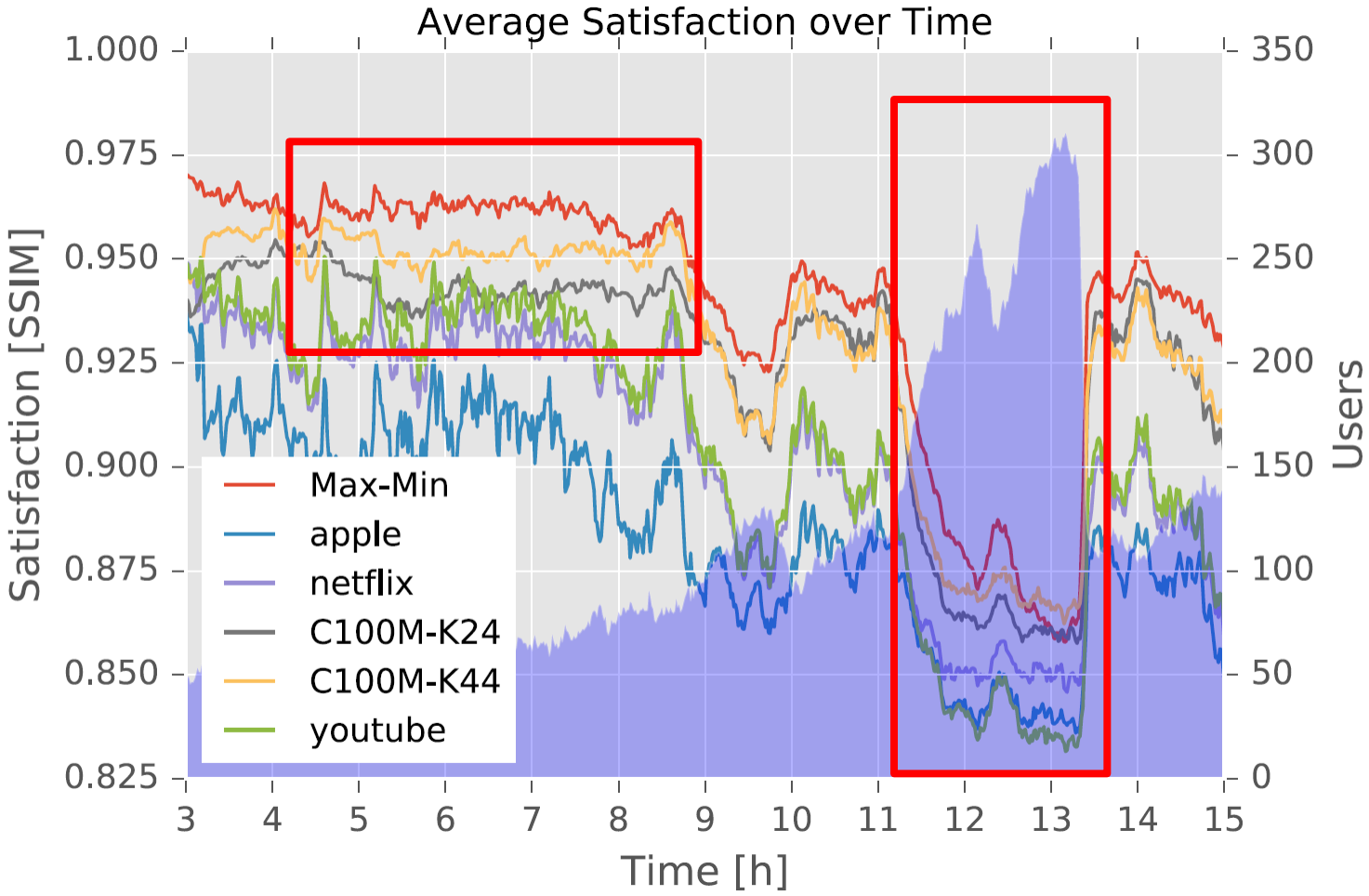
- In general:
  - Discrete bitrates (e.g., 100, 200, 500, … kbit/s)
  - Enables clients to maintain a local video playback buffer
- Upper bound: max-min model
  - Assumption: continous bitrates (0 – 20.000 kbit/s)
  - Distribute bandwidth among competitors on a best-effort principle (max-min fairness)
  - No local video playback buffer

NS-3 AND DASH

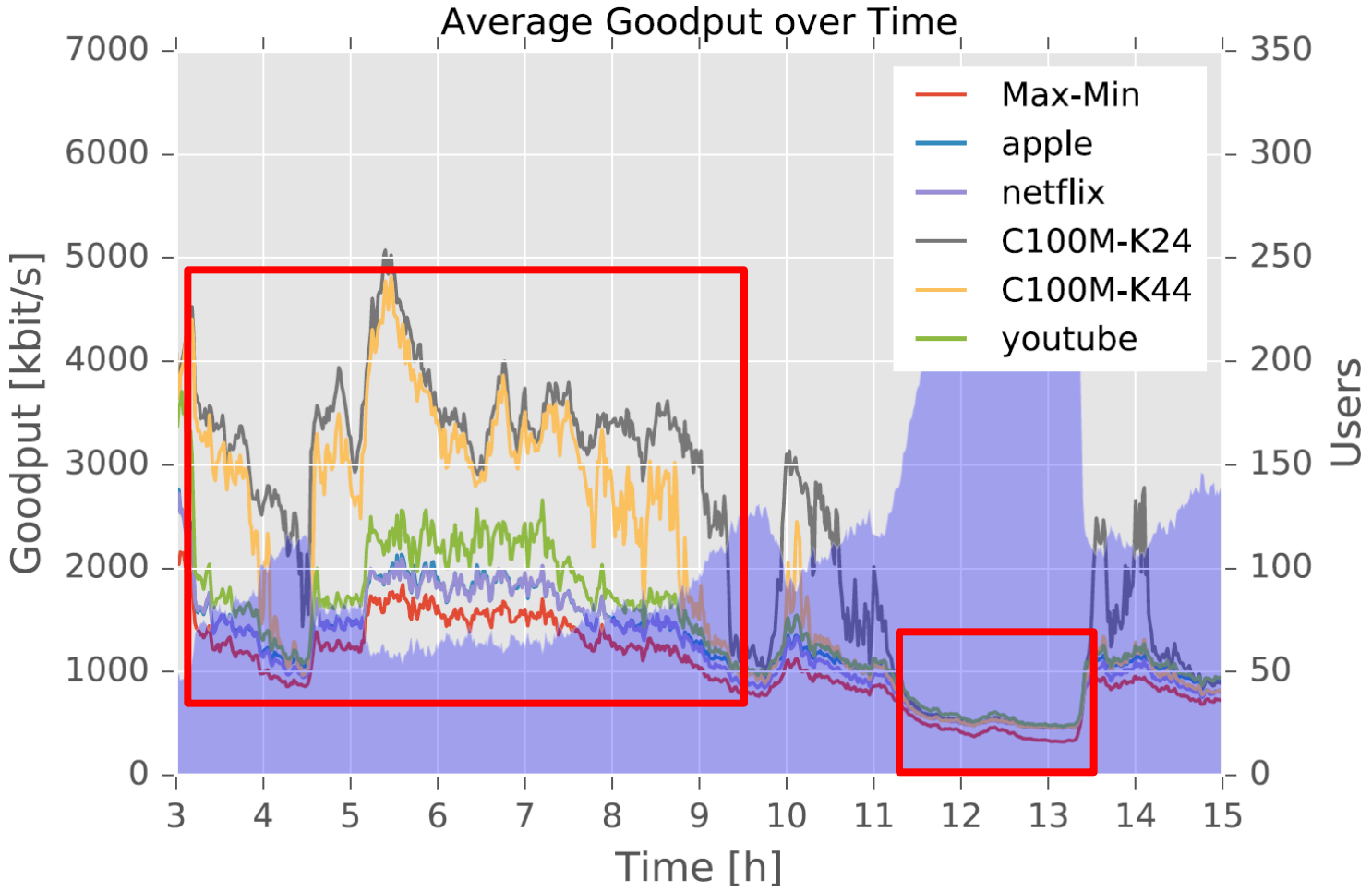
- NS-3 is a time-discrete packet-level simulator
- We
  - I
  - F
  - M
- Ne
  - MTU: 1500 bytes
  - TCP New Reno, w/ segment size (MSS) of 1430 bytes
  - Bottleneck link: 100, 200 Mbit/s and 15 ms RTT



RESULTS: AVG. USER SATISFACTION (BOTTLENECK 100 MBIT/S)



RESULTS: AVG. GOODPUT (BOTTLENECK 100 MBIT/S)



CONCLUSIONS

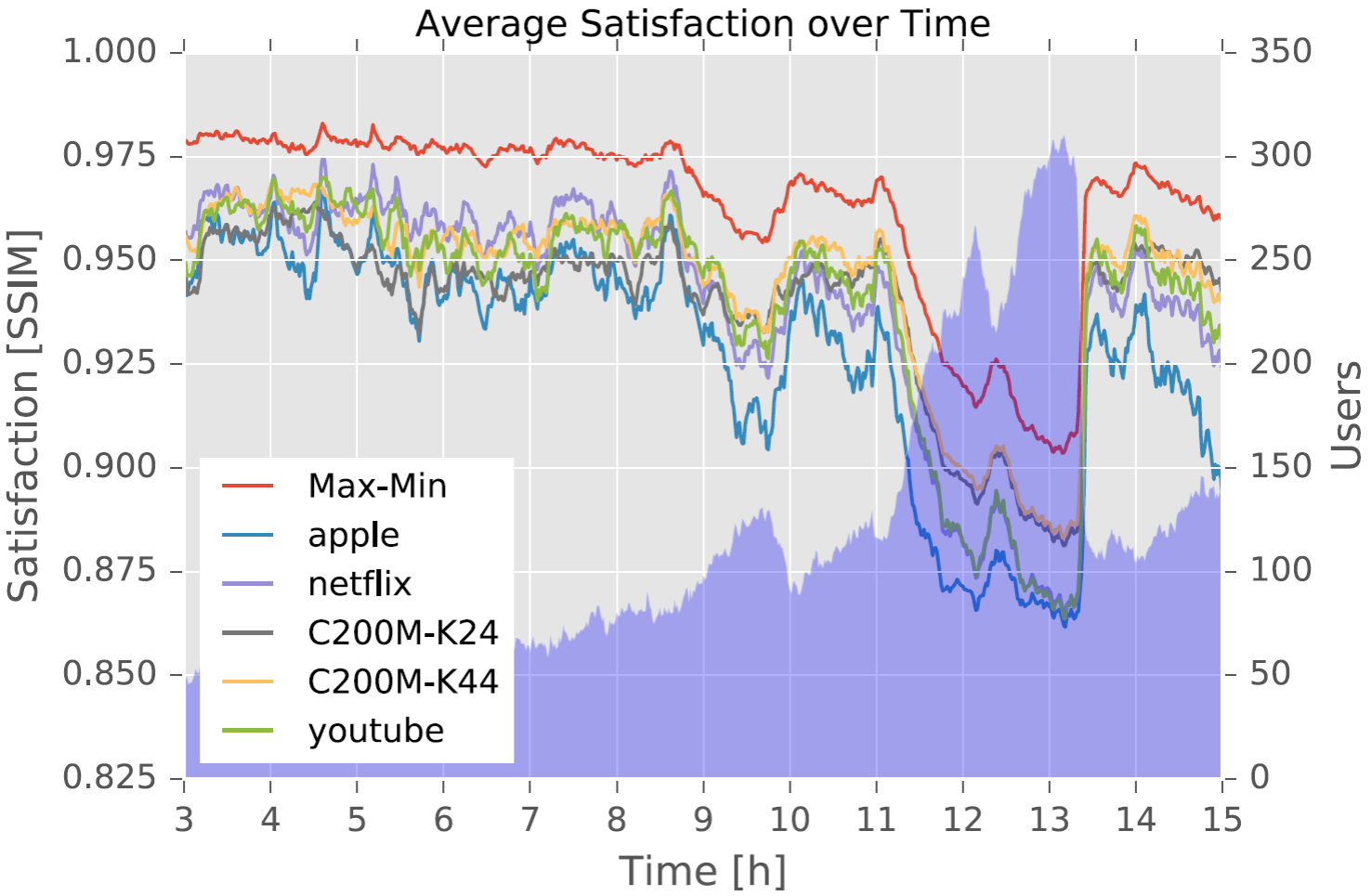
- The choice of representations has an impact on satisfaction
- Researchers need to consider this when evaluating their DASH adaptation strategies
- Optimized representations [16] are good, but don‘t work very well in all situations

- Data and simulation framework available at http://concert.itec.aau.at/NOSSDAV_2016/ and https://github.com/ChristianKreuzberger/AMuSt-Simulator/

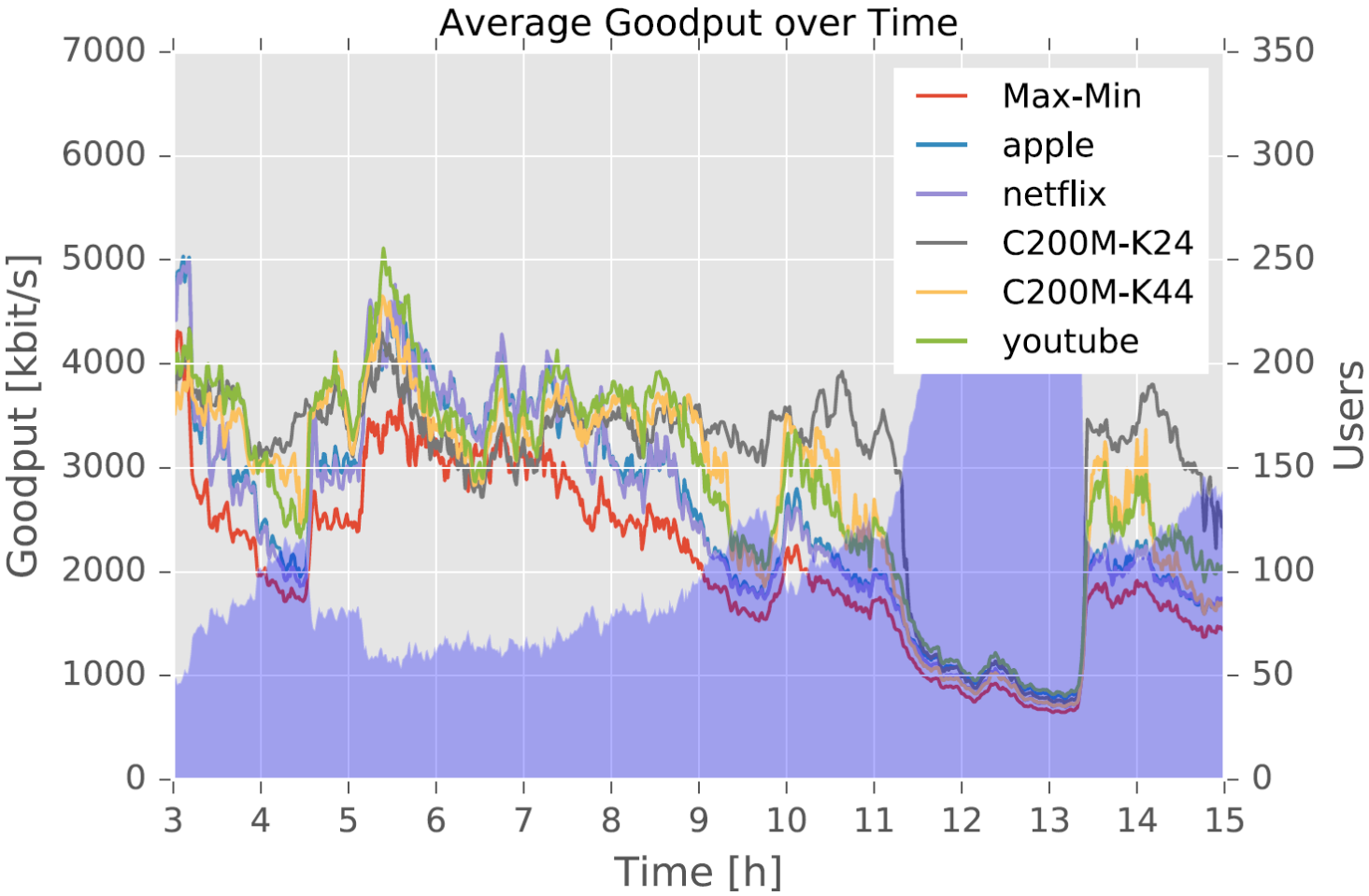
THANK YOU!

BACKUP SLIDES

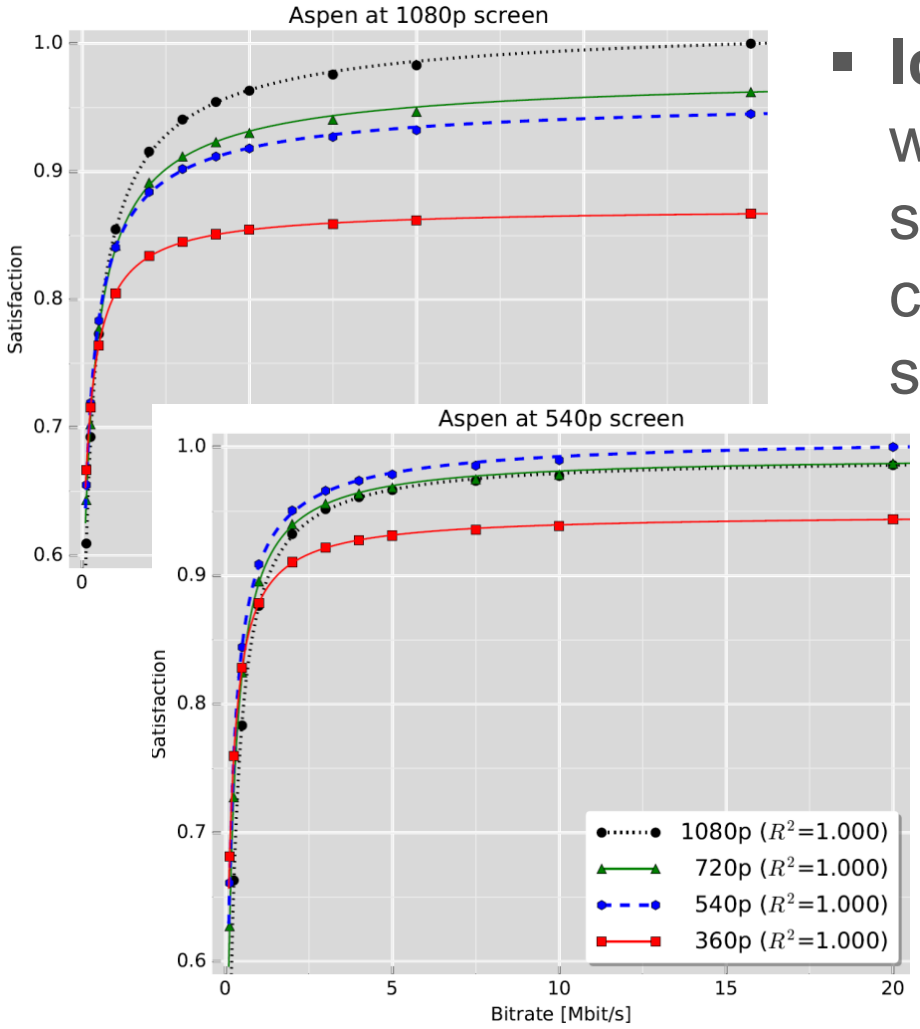
RESULTS: AVG. USER SATISFACTION (BOTTLENECK 200 MBIT/S)



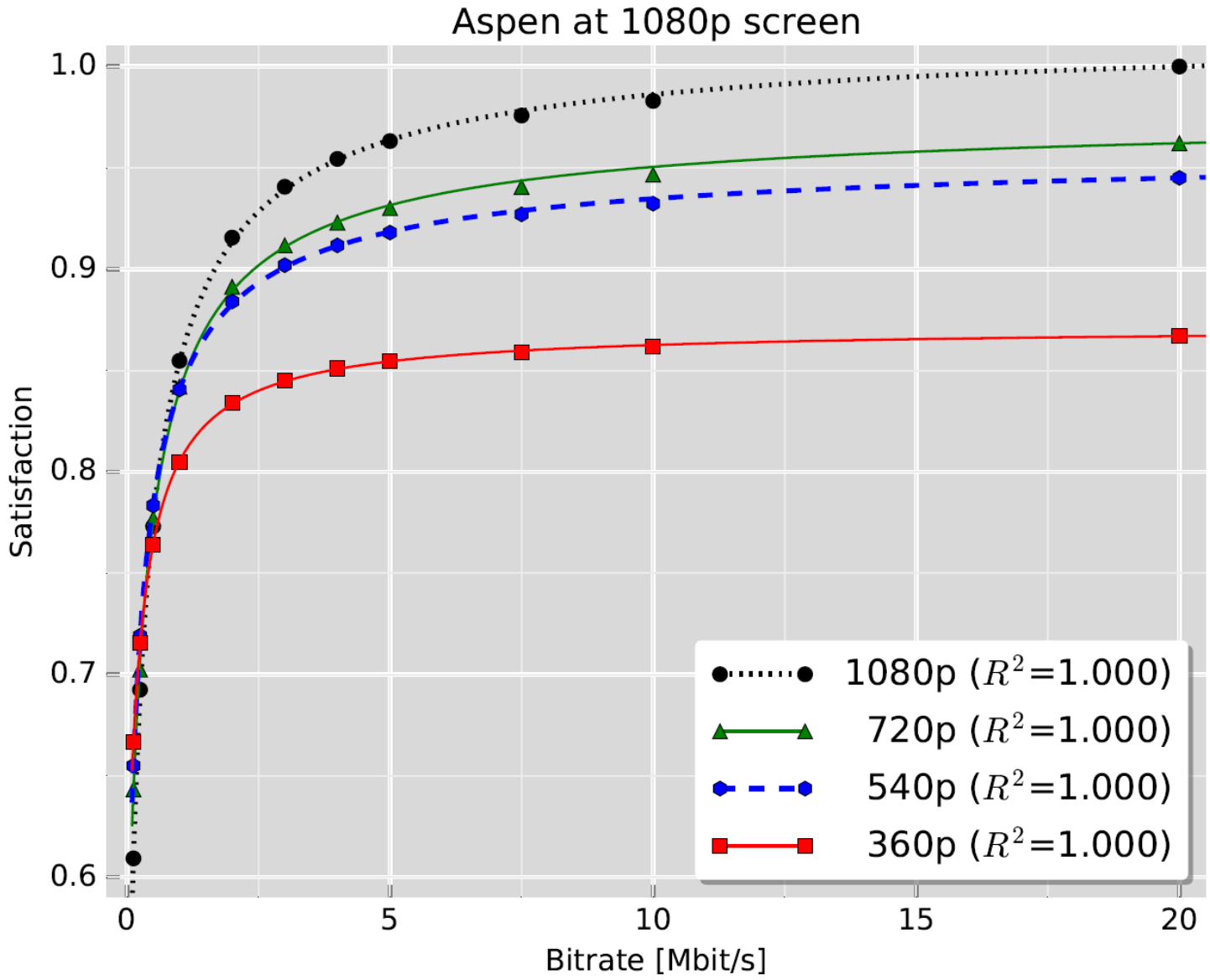
RESULTS: AVG. GOODPUT (BOTTLENECK 200 MBIT/S)

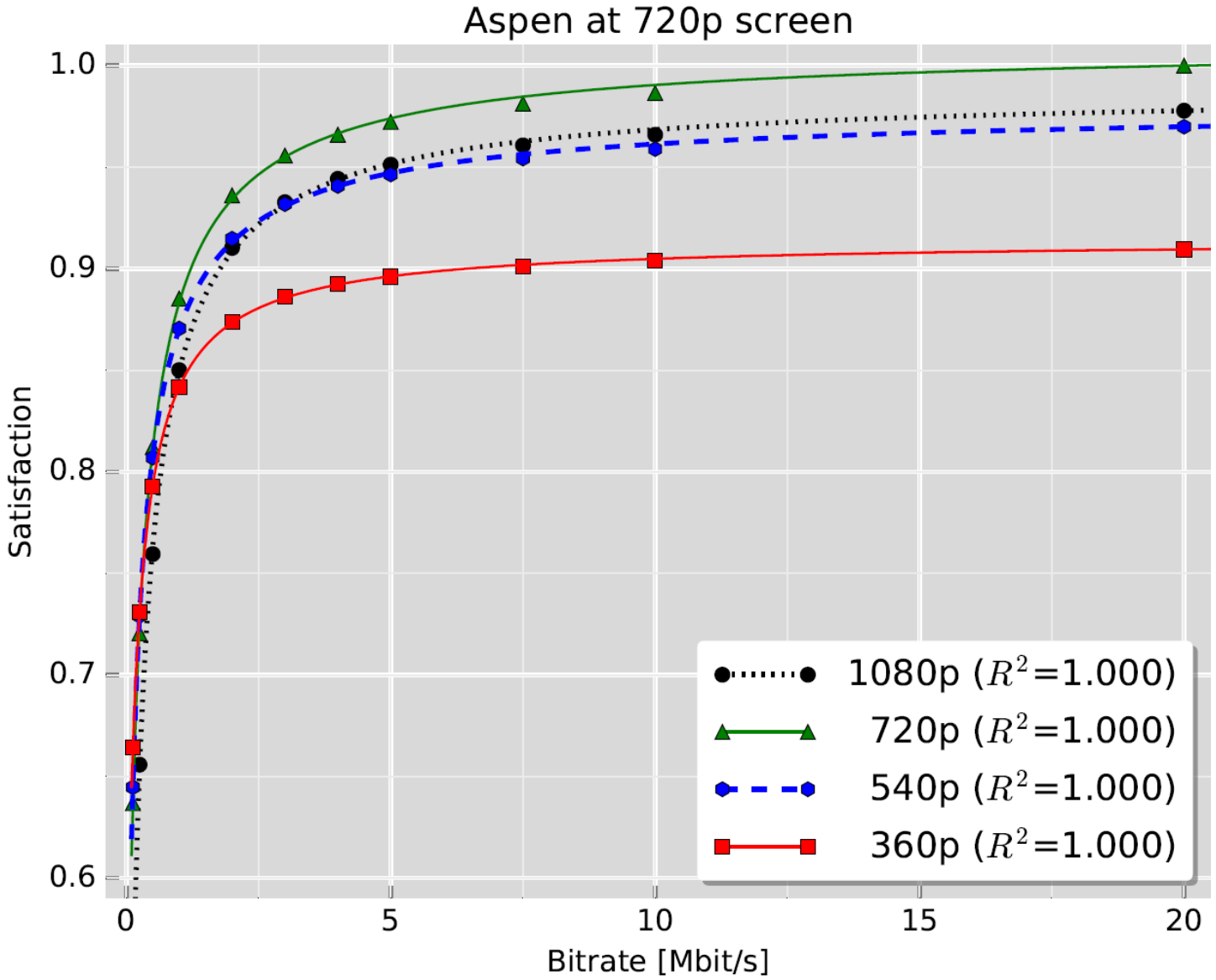


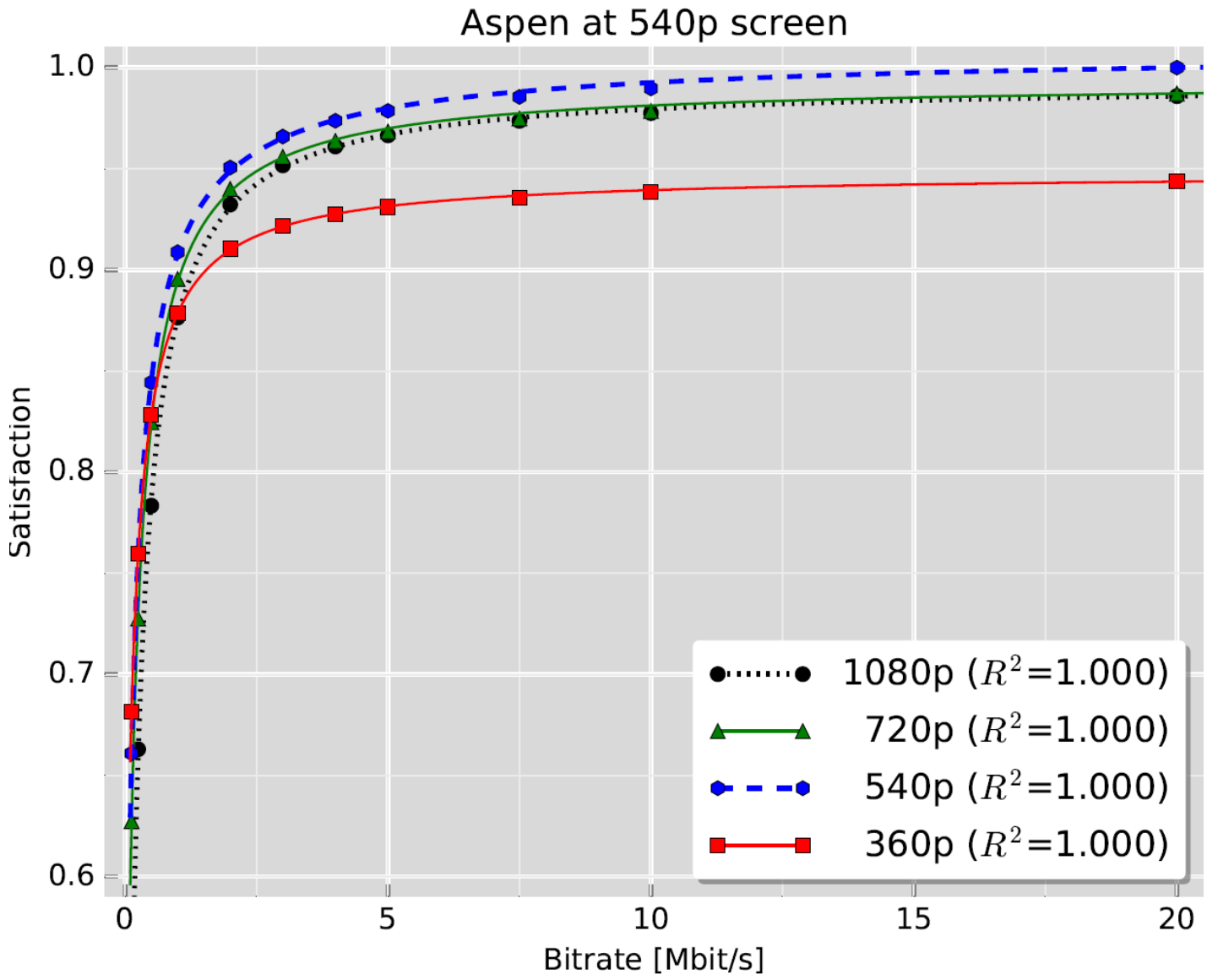
USER SATISFACTION



- **Idea:** lower bit rate required when satisfying a user with a low screen resolution (360p) compared to a user with a high screen resolution (540p–1080p)
  - Encoded 4 video sequences
  - 360p, 540p, 720p, 1080p
  - 100 kbit/s – 20 Mbit/s


Aspen at 1080p screen
Satisfaction
1.0
0.9
0.8
0.7
0.6
0
5
10
15
20
Bitrate [Mbit/s]
1080p (R^2=1.000)
720p (R^2=1.000)
540p (R^2=1.000)
360p (R^2=1.000)


Aspen at 720p screen
Satisfaction
1.0
0.9
0.8
0.7
0.6
0
5
10
15
20
Bitrate [Mbit/s]
1080p (R^2=1.000)
720p (R^2=1.000)
540p (R^2=1.000)
360p (R^2=1.000)


Aspen at 540p screen
Satisfaction
1.0
0.9
0.8
0.7
0.6
0
5
10
15
20
Bitrate [Mbit/s]
1080p (R^2=1.000)
720p (R^2=1.000)
540p (R^2=1.000)
360p (R^2=1.000)