Evaluation of DASH Rate Adaptation Algorithms in Operational Mobile Networks

MONROE

Cise Midoglu, Konstantinos Kousias, Carsten Griwodz, Ozgu Alay
Simula Research Laboratory

simula

Research problem: evaluation of different video rate adaptation algorithms over operational mobile networks
Comparison of basic rate adaptation, Segment Aware Rate Adaptation (SARA) and Buffer-Based Rate Adaptation (BBA)

Which DASH algorithm performs best over mobile networks?   **Which operator has the best support for video streaming?**

DASH: Dynamic Adaptive Streaming over HTTP

Basic rate adaptation is a throughput-based algorithm. BBA is a buffer-based algorithm proposed by [1], and currently in use by Netflix. SARA adds segment-awareness, it is proposed by [2].

Basic	SARA	BBA (Netflix)
throughput-based	hybrid	buffer-based

AStream, MONROE and OpenVQ

Astream [3] is a Python based emulated video player to evaluate the performance of DASH bitrate adaptation algorithms. We create a docker container with this code to run it on MONROE nodes.

MONROE [4] is a Europe-wide measurement platform with nodes deployed in 4 countries (Italy, Norway, Spain, Sweden). Each multihomed node pair supports 3 cellular network connections and can mimic common end user devices.

OpenVQ [5, 6] is a video quality assessment framework. We use OpenVQ to evaluate OPVQ, PSNR, and SSIM on a per-segment basis (transferred vs. original) using video segments pre-processed and up-scaled with FFMPEG [7].

Testbed and Measurement Campaign

We run 3 rate adaptation algorithms in the LTE networks of 3 Norwegian Mobile Network Operators (MNO): Telenor, Telia, and ICE with the testbed depicted in Figure 1.
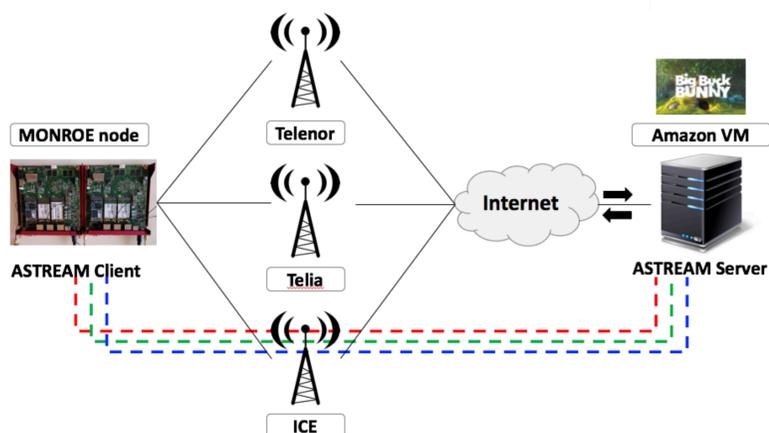

Figure 1. Testbed: MONROE nodes serve as DASH clients, Amazon Virtual Machine located in Ireland serves as DASH server.

References

[1] T.-Y. Huang et al., "A buffer-based approach to rate adaptation: evidence from a large video streaming service." ACM SIGCOMM, 2014.
[2] P. Juluri et al., "SARA: Segment aware rate adaptation algorithm for dynamic adaptive streaming over HTTP," ICCW, 2015.
[3] https://github.com/pari685/astream.
[4] O. Alay et al., "Measuring and assessing mobile broadband networks with MONROE," IEEE WoWMoM, 2016.
[5] K. Skarseth, H. Bjorlo, P. Holversen, M. Riegler, and C. Griwodz, "OpenVQ - a video quality assessment toolkit," ACM MM, 2016.
[6] https://bitbucket.org/mpg code/openvq.
[7] https://ffmpeg.org.

Comparison of Playback Types and Operators

We observe that for all operators, Netflix is providing the worst performance. Average scores for basic, SARA, and Netflix are given in the top section of Table 1. We observe that, albeit slightly, Telia outperforms the other operators for all playback types, followed by Telenor, and then ICE. Average scores for Telenor, Telia, and ICE are given in the bottom section of Table 1.
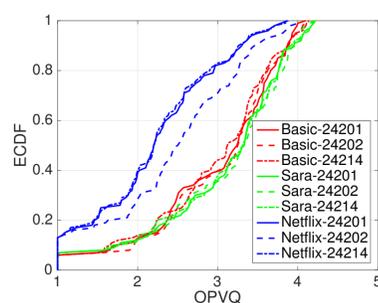

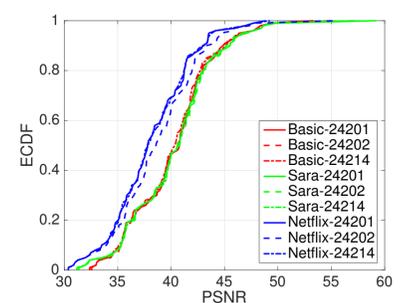**Figure 2. ECDF of the OPVQ value per playback type and operator.**

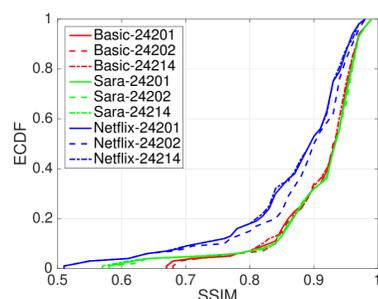**Figure 3. ECDF of the PSNR value per playback type and operator.**

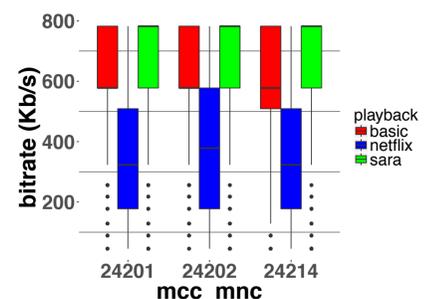**Figure 4. ECDF of the SSIM value per playback type and operator.**

Figure 5. Boxplot of bitrate per operator for all segments.

Playback / Operator	OPVQ	PSNR	SSIM
Basic	2.9859	40.2677	0.9140
SARA	3.0702	40.3730	0.9124
BBA (Netflix)	2.2963	38.3963	0.8730
Telenor	2.7619	39.6257	0.8987
Telia	2.8515	39.8265	0.9027
ICE	2.7390	39.5848	0.8981

Table 1. Average OPVQ, PSNR, and SSIM (top) over all operators, (bottom) over all playback types.

Conclusions and Future Work

- On average, SARA is performing best among playback types while Netflix is worst.
- There is no significant difference between operators, Telia seems to slightly outperform Telenor and ICE.
- Difference between roaming and no roaming cases as well as between different modems are not visible due to low bitrate requirements.
- Larger number of runs required for higher accuracy.
- Playback types might be optimized for different scenarios (handheld/large screen, focus on bitrate/buffering time, etc.), current comparison is too crude.
- Selected metrics do not consider other factors possibly effecting user experience (e.g. number and duration of stalls).
- Future work includes the integration of further metrics considering stalls, as well as a more diverse and longitudinal analysis using the MONROE platform.