ITG Workshop Sound, Vision and Games 2015

Modeling Buffered Video Streaming Startup Delays in Multi-Cellular Wireless Networks

Henrik Klessig and Gerhard Fettweis

Vodafone Chair Mobile Communications Systems,
University of Technology Dresden, Germany

Hanover, September 22nd, 2015

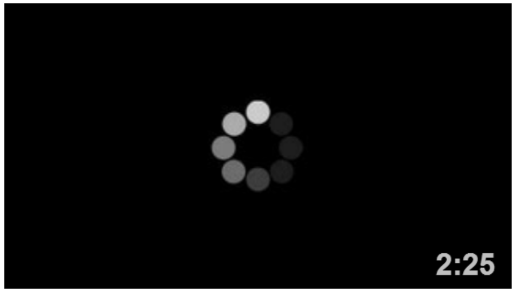
Agenda

1. Motivation and KPIs
2. Multi-Cellular Flow Level Model
3. Startup Delay Distribution
4. Results
5. Take-Aways

1. Motivation and KPIs

What are the most annoying „features“ of video streaming?

- Long initial buffering phase – users start abandoning the video if prefetch phase exceeds two seconds [1]
- Many (>1) and long rebuffering phases



QoS metrics such as
- Data rates
- User throughput

are not appropriate anymore!

Buffered streaming = elastic traffic:
- Download of a file with variable rate
- Simultaneous playback with constant rate

QoS/QoE metrics [2]:

Startup delay:	interval between begin of the streaming session and start of playback, given by startup threshold q_a [s] and flow throughput
Starvation probability:	probability that the playout buffer becomes empty
Rebuffering delay:	interval between playout butter starvation and restart of playback, given by rebuffering delay and flow throughput

---

[1] S. S. Krishnan and R. K. Sitaraman, "Video stream quality impacts viewer behavior: Inferring causality using quasi-experimental designs," in *Proceedings of the 2012 ACM Conference on Internet Measurement Conference*, ser. IMC '12. New York, NY, USA: ACM, 2012, pp. 211– 224.

[2] Yuedong Xu; Elayoubi, S.E.; Altman, E.; El-Azouzi, R., "Impact of flow-level dynamics on QoE of video streaming in wireless networks," INFOCOM, 2013 Proceedings IEEE , vol., no., pp.2715,2723, 14-19 April 2013.

2. Multi-Cellular Flow Level Model (1/3)

Goals:

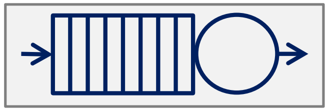
- Compute user QoS/QoE *but avoid* extensive Monte-Carlo simulations
- Build a framework that models the user QoS/QoE based on common traffic characteristics
- Use the framework to develop self-organizing network algorithms

Procedure:

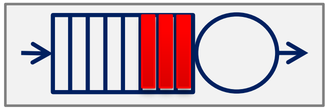
rate adapts to available channel capacity / bandwidth

Packets (IP layer)
Data flows
User sessions
Connections

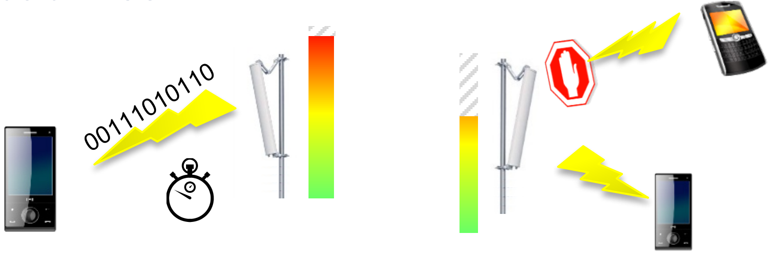
1. Model so-called *elastic* data traffic
2. Interpret a BS as a server in a queuing system



3. Characterize the BS's state

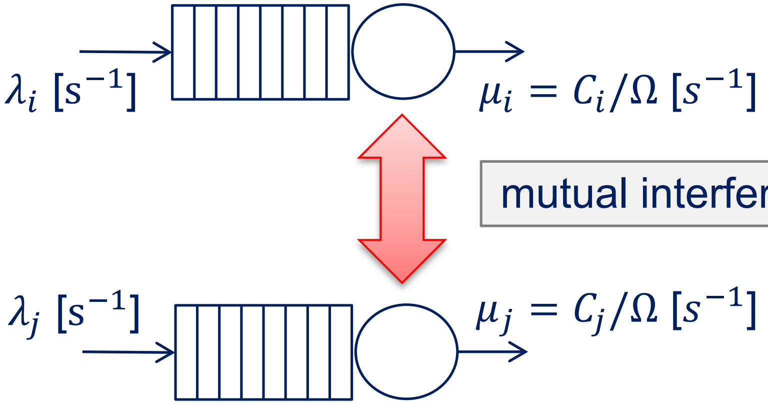


4. Derive QoS/QoE metrics from BS state probabilities
5. Optimize user QoS/QoE



2. Multi-Cellular Flow Level Model (2/3)

Coupling:



mutual interference → service rates are coupled

with BS/cell indices i, j

Achievable rate at location u in bps, if flow is connected to BS i, interference scenario y:

$$c_i(u,y) \coloneqq aB \min\left\{ \log_2\left(1 + \mathrm{b}\frac{p_i(u)}{\sum_{j\in\mathcal{N}_1(y)\setminus\{i\}} p_j(u) + N_0}\right), c_{\max} \right\}$$

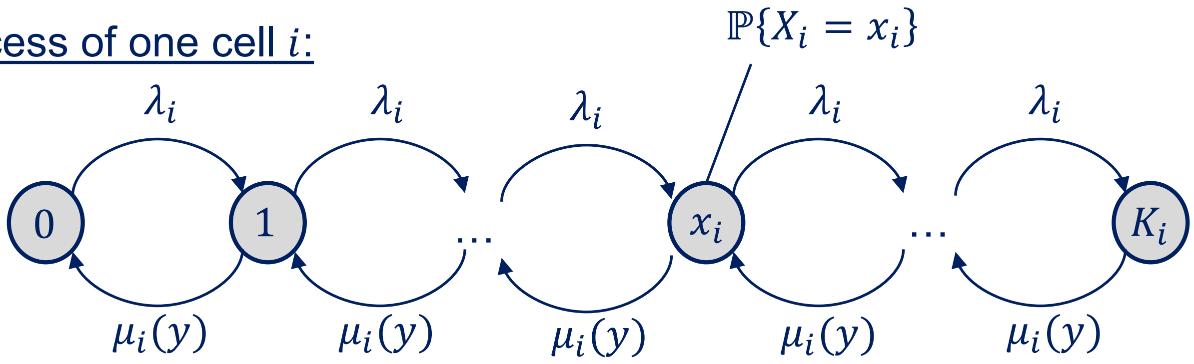
Definition: Interference-dependent cell capacity (*harmonic mean* of achievable rates):

$$C_i(y) \coloneqq \left(\int_{\mathcal{L}_i} \delta_i(u) c_i(u,y)^{-1} \mathrm{d}u \right)^{-1}$$

$$\delta_i(u) \coloneqq \frac{\delta(u)}{\int_{\mathcal{L}_i} \delta(u)\,\mathrm{d}u}$$

2. Multi-Cellular Flow Level Model (3/3)

Continuous time Markov process of one cell i:



Def. Random variable X_i describing the number of active flows in cell i, $X_i \in \{0, \dots, K_i\}$

$$\boxed{\rho_i(y) = \frac{\lambda_i}{\mu_i(y)}}$$

- y varies on the same time scale of flow dynamics!
- No M/M/1/K model applicable!
- No closed (product) form!

$$\xcancel{\mathbb{P}[X_i = x_i] = \frac{(1-\rho_i)\rho_i^{x_i}}{1-\rho_i^{K_i+1}}}$$

- Build a multi-dimensional Markov process to
- Account for the variation of the interference y

(2.2) Compute state probabilities $\boxed{\pi(x) \coloneqq \mathbb{P}\{X = x\}}$

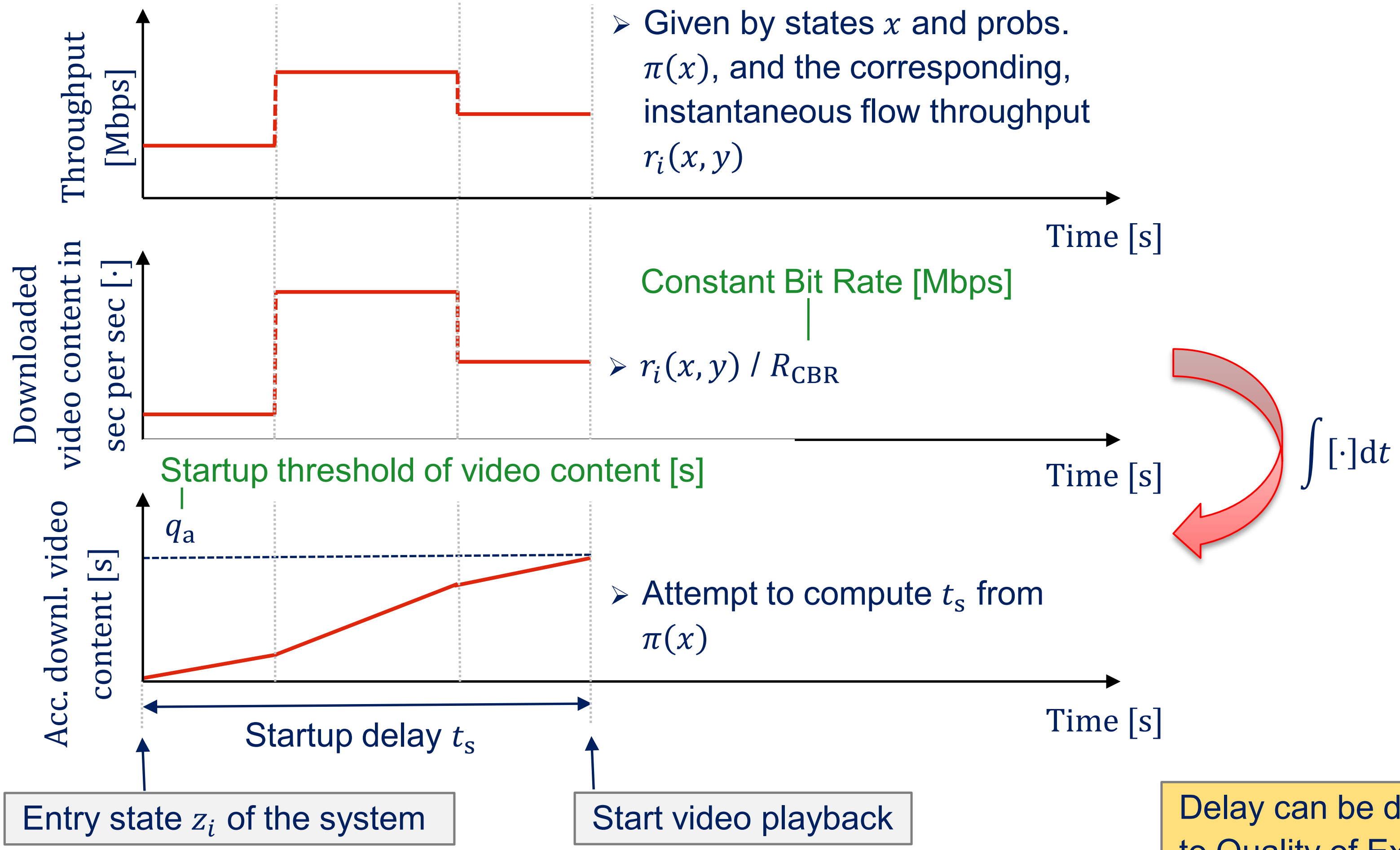
(2.3) Compute performance metrics from $\pi(x)$

Def. Random vector X with elements X_i.
$X \in \mathcal{X} \coloneqq \{0, \dots, K_1\} \times \cdots \times \{0, \dots, K_N\}$.

Def. Random vector $Y \coloneqq \operatorname{sgn}(X)$ with realizations y. $Y \in \mathcal{Y} \coloneqq \{0,1\}^N$.

3. Startup Delay Distribution (1/4)

Prefetching phase:



- Given by states x and probs. $\pi(x)$, and the corresponding, instantaneous flow throughput $r_i(x,y)$

Constant Bit Rate [Mbps]

- $r_i(x,y)$ / R_CBR

Startup threshold of video content [s]

- Attempt to compute t_s from $\pi(x)$

Entry state z_i of the system

Start video playback

Delay can be directly mapped to Quality of Experience

3. Startup Delay Distribution (2/4)

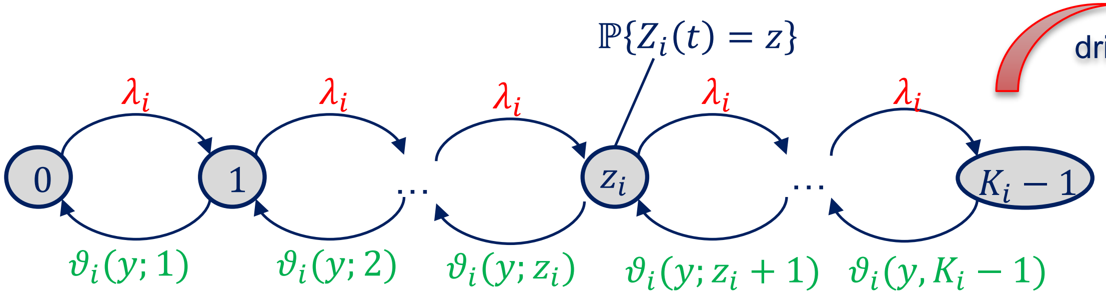
<u>1st step:</u> Entry state distribution π_z and probability of interference $\zeta_i(y)$

<u>2nd step:</u> Evolution of the buffer fill status for <u>static (fixed)</u> interference y

- We assume that the „tagged“ flow observes the external interference process in the quasi-stationary (QS) regime.
- Well supported if startup delay is much smaller than the average flow sojourn time.

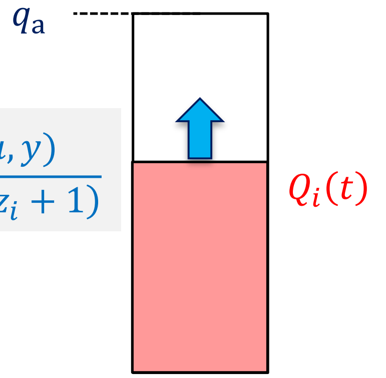
Approach (from [2]): We model the system as two queues „in tandem“.

1. Markov chain describing the process $Z_i(t)$, i.e. the number of other concurrent data flows in the cell.
2. A queue describing the buffer fill status $Q_i(t)$ in seconds of video content.



drives

$$v_i(u, y; z_i) = \frac{c_i(u, y)}{R_{\mathrm{CBR}}(z_i + 1)}$$



Video buffer fill status

- The service rate is determined without the tagged flow and, therefore, is state-dependent.

$$\vartheta_i(y; z_i) = \frac{z_i}{z_i + 1} \frac{C_i(y)}{\Omega}$$

[2] Yuedong Xu; Elayoubi, S.E.; Altman, E.; El-Azouzi, R., "Impact of flow-level dynamics on QoE of video streaming in wireless networks," INFOCOM, 2013 Proceedings IEEE , vol., no., pp.2715,2723, 14-19 April 2013.

3. Startup Delay Distribution (3/4)

$U_i(u, y, t; z_i, q)$ … probability that the prefetching process finishes until time t

- Flow at location $u \in \mathcal{L}_i$
- Under interference y
- With entry state z_i
- With prefetch content q

In the interval $[t, t + \Delta t]$ four possible events can occur:

- No change of number of flows,
- Arrival of one flow,
- Departure of one flow (not the „tagged“), or
- More than one event.

Dynamics of the probability U_i:

$$\begin{aligned} U_i(u,y,t;z_i,q) = \; & (1 - \lambda_i \Delta t - \vartheta_i(y; z_i)\Delta t) \cdot U_i(u, y, t - \Delta t; z_i, \quad q - v_i(u, y; z_i)\Delta t) && \text{(no change)} \\ & + \lambda_i \Delta t \cdot U_i(u, y, t - \Delta t; z_i + 1, q - v_i(u, y; z_i)\Delta t) && \text{(arrival)} \\ & + \vartheta_i(y; z_i)\Delta t \cdot U_i(u, y, t - \Delta t; z_i - 1, q - v_i(u, y; z_i)\Delta t) && \text{(departure)} \\ & + \; o(\Delta t) && \text{(multiple events)} \end{aligned}$$

Let $\Delta t \to 0$:

$$\frac{\partial \boldsymbol{U}_i(u, y, t; q)}{\partial t} = -\boldsymbol{M}_i(y) \boldsymbol{U}_i(u, y, t; q)$$

with $$\boldsymbol{M}_i(y) = \begin{pmatrix} \lambda_i + \vartheta_i(y; 0) & -\lambda_i & 0 & \dots & 0 \\ -\vartheta_i(y; 1) & \lambda_i + \vartheta_i(y; 1) & -\lambda_i & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \vartheta_i(y; K_i - 1) \end{pmatrix}$$

with $\boldsymbol{U}_i(u, y, t; q) \coloneqq \left(U_i(u, y, t; 0, q), \dots, U_i(u, y, t; K_i - 1, q)\right)^{\mathrm{T}}$.

3. Startup Delay Distribution (4/4)

Efficient solution:

$$\boxed{\boldsymbol{U}_i(u,y,t;q) = \left(\boldsymbol{D}_i \exp(-\boldsymbol{\Lambda}_i t)\boldsymbol{D}_i^{-1}\boldsymbol{G}_i\right)(u,y,t;q)}$$

where $\boldsymbol{M}_i(y) = \left(\boldsymbol{D}_i\boldsymbol{\Lambda}_i\boldsymbol{D}_i^{-1}\right)(y)$, $\boldsymbol{D}_i$ is invertible, and $\boldsymbol{\Lambda}_i$ is diagonal containing the eigenvalues of $\boldsymbol{M}_i$

with $G_i(q - v_i(u,y;z_i)t) = \begin{cases} 0 \text{ for } \; q - v_i(u,y;z_i)t \geq 0 \\ 1 \text{ for } \; q - v_i(u,y;z_i)t < 0 \end{cases}$ and

$$\boldsymbol{G}_i(u,y,t;q) \coloneqq \left(G_i(q - v_i(u,y;0)t), \dots, G_i(q - v_i(u,y;K_i - 1)t)\right)^{\mathrm{T}}$$

So far, we have the startup delay distribution $U_i(u,y,t;z_i,q)|_{q=q_{\mathrm{a}}}$ for

- a specific location $u \in \mathcal{L}_i$ (we know $\delta_i(u)$),
- a specific entry state z_i (we know π_{z_i}), and
- a specific interference scenario y (we know $\zeta_i(y)$).

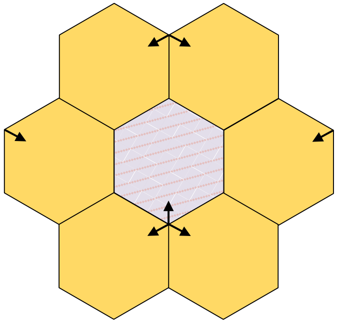
Startup delay distribution independent of z_i and y:

$$\boxed{U_i(u,t;q_{\mathrm{a}}) = \sum_{y \in \mathcal{Y}} \pi_{z_i} \boldsymbol{U}_i(u,y,t;q_{\mathrm{a}})\zeta_i(y)}$$

Compound startup delay distribution in cell i:

$$\boxed{U_i(t;q_{\mathrm{a}}) = \int_{\mathcal{L}_i} U_i(u,t;q_{\mathrm{a}})\delta_i(u)\mathrm{d}u}$$

4. Results (1/2)



Scenario:

- 7 cells in a hexagonal layout
- 3GPP-compliant configuration
- Central cell under consideration

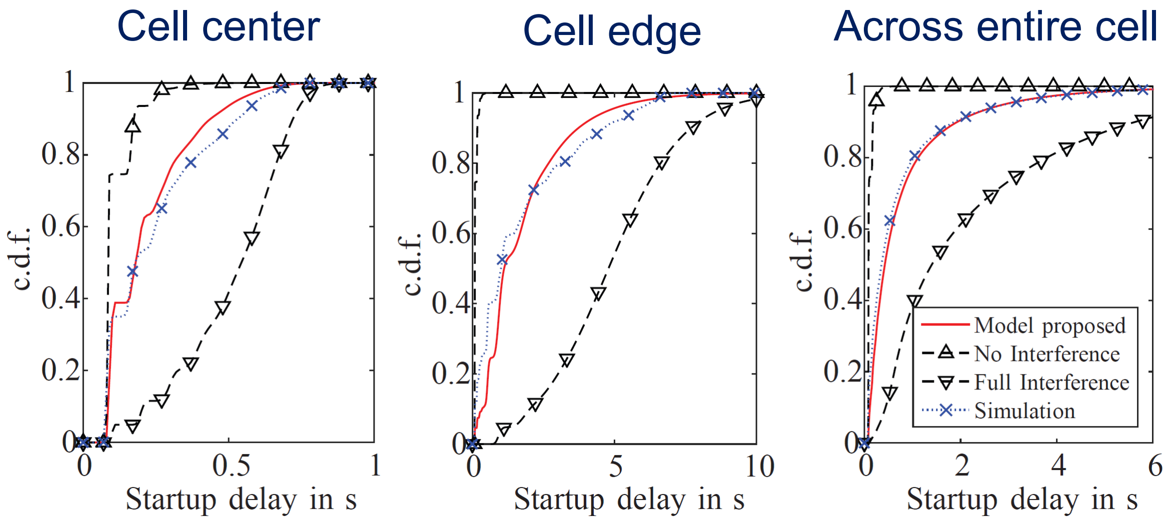
Model:

- State aggregation
- Full and no interference (lower and upper performance bounds)

Experiment:

- Compare model results with simulations + full and no interference
- Admission control: $K_i = 7, \forall i$
- Startup threshold: $q_\mathrm{a} = 3\ \mathrm{s}$
- Arrival intensity: $\lambda_i = 0.0375\ \mathrm{s}^{-1}$
- Mean video length: $480\ \mathrm{s}$
- Video bitrate: $R_\mathrm{CBR} = 2\ \mathrm{Mbps}$

4. Results (2/2)


Cell center

Cell edge

Across entire cell

- Remarkable accuracy
- Much higher startup delays at cell edge due to strong inter-cell interference (~ factor 10)
- Interference affects cell center users, since low performance data flows „steel“ radio resources.

Published in:

H. Klessig and G. Fettweis, “Short Paper: Impact of Inter-Cell Interference on Buffered Video Streaming Startup Delays” in Proceedings of the 82nd IEEE Vehicular Technology Conference (VTC Fall'15), Boston, USA, 6.9. - 9.9.2015

5. Main Take-Aways

1. Buffered streaming = elastic traffic
2. Mathematical model with a few assumptions:
   a) Constant bitrate
   b) Poisson arrivals and exponentially distr. flow sizes
   c) Resource fair scheduler
3. Predominant effects:
   a) Concurrent flows
   b) Inter-cell interference
   c) Heterogeneous rate distribution
4. Variable bitrates and fast fading have minor effects

THANK YOU