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
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 buffer starvation and restart of playback, given by rebuffering delay and flow throughput


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:

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:**

\[ \lambda_i \ [s^{-1}] \quad \mu_i = \frac{C_i}{\Omega} \ [s^{-1}] \]

\[ \lambda_j \ [s^{-1}] \quad \mu_j = \frac{C_j}{\Omega} \ [s^{-1}] \quad \text{with BS/cell indices } i, j \]

**mutual interference → service rates are coupled**

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

\[
c_i(u, y) := aB \min \left\{ \log_2 \left( 1 + b \frac{p_i(u)}{\sum_{j \in N_1(y) \setminus \{i\}} p_j(u) + N_0} \right), c_{\text{max}} \right\}
\]

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

\[
C_i(y) := \left( \int_{L_i} \delta_i(u)c_i(u, y)^{-1} du \right)^{-1}
\]

\[
\delta_i(u) := \frac{\delta(u)}{\int_{L_i} \delta(u) \ du}
\]
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 $i$, $X_i \in \{0, ..., K_i\}$

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

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

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

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

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

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

Def. Random vector $Y := \text{sgn}(X)$ with realizations $y$. $Y \in \mathcal{Y} := \{0, 1\}^N$. 
3. Startup Delay Distribution (1/4)

Prefetching phase:

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

- Given by states $x$ and probs. $\pi(x)$, and the corresponding, instantaneous flow throughput $r_i(x, y) / R_{\text{CBR}}$

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

Delay can be directly mapped to Quality of Experience
1st step: Entry state distribution $\pi_z$ and probability of interference $\zeta_i(y)$

2nd step: Evolution of the buffer fill status for static (fixed) interference $\gamma$

- 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.

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

\[ \psi_i(z) = \frac{z}{z + 1} \frac{C_i(y)}{\Omega} \]

3. Startup Delay Distribution (3/4)

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 \(U_i\):

\[
U_i(u, y, t; z_i, q) = (1 - \lambda_i \Delta t - \varphi_i(y; z_i)\Delta t) \cdot U_i(u, y, t - \Delta t; z_i, q - v_i(u, y; z_i)\Delta t)
\]

\[
+ \lambda_i \Delta t \cdot U_i(u, y, t - \Delta t; z_i + 1, q - v_i(u, y; z_i)\Delta t)
\]

\[
+ \varphi_i(y; z_i)\Delta t \cdot U_i(u, y, t - \Delta t; z_i - 1, q - v_i(u, y; z_i)\Delta t)
\]

\[
+ o(\Delta t)
\]

Let \(\Delta t \to 0\):

\[
\frac{\partial U_i(u, y, t; q)}{\partial t} = -M_i(y)U_i(u, y, t; q)
\]

with

\[
M_i(y) = \begin{pmatrix}
\lambda_i + \varphi_i(y; 0) & -\lambda_i & 0 & \ldots & 0 \\
-\varphi_i(y; 1) & \lambda_i + \varphi_i(y; 1) & -\lambda_i & \ldots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & \varphi_i(y; K_i - 1)
\end{pmatrix}
\]

with

\[
U_i(u, y, t; q) := \left( U_i(u, y, t; 0, q), \ldots, U_i(u, y, t; K_i - 1, q) \right)^T.
\]
3. Startup Delay Distribution (4/4)

Efficient solution:

\[ U_i(u, y, t; q) = \left( D_i \exp(-\Lambda_i t) D_i^{-1} G_i \right)(u, y, t; q) \]

where \( M_i(y) = \left( D_i \Lambda_i D_i^{-1} \right)(y) \), \( D_i \) is invertible, and \( \Lambda_i \) is diagonal containing the eigenvalues of \( M_i \)

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

\[ G_i(u, y, t; q) := \left( G_i(q - v_i(u, y; 0)t), \ldots, G_i(q - v_i(u, y; K_i - 1)t) \right)^T \]

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

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

Startup delay distribution independent of \( z_i \) and \( y \): Compound startup delay distribution in cell \( i \):

\[ U_i(u, t; q_a) = \sum_{y \in \mathcal{Y}_i} \pi_{z_i} U_i(u, y, t; q_a) \zeta_i(y) \]

\[ U_i(t; q_a) = \int_{\mathcal{L}_i} U_i(u, t; q_a) \delta_i(u) du \]
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: $q_a = 3$ s
- Arrival intensity: $\lambda_i = 0.0375$ s$^{-1}$
- Mean video length: 480 s
- Video bitrate: $R_{CBR} = 2$ Mbps
4. Results (2/2)

- 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:
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