Stochastic Analysis of Self-Sustainability in Peer-Assisted VoD Systems

Delia Ciullo[1], Valentina Martina[1], Michele Garetto[2], Emilio Leonardi[1], Gianluca Torrisi[3]

[1]Politecnico di Torino

[2]Università di Torino

[3]CNR - Instituto per le Applicazioni di Calcolo

March 29-th, 2012

Introduction

In **peer-assisted Video-on-Demand (VoD) systems:**

Introduction

In **peer-assisted Video-on-Demand (VoD) systems:**

- users browse a catalog of available videos and asynchronously issue requests to watch a given content

Introduction

In **peer-assisted Video-on-Demand (VoD) systems:**

- users browse a catalog of available videos and asynchronously issue requests to watch a given content
- content is divided into chunks that can be retrieved either from

Introduction

In **peer-assisted Video-on-Demand (VoD) systems:**

- users browse a catalog of available videos and asynchronously issue requests to watch a given content
- content is divided into chunks that can be retrieved either from
    - the servers

Introduction

In **peer-assisted Video-on-Demand (VoD) systems:**

- users browse a catalog of available videos and asynchronously issue requests to watch a given content
- content is divided into chunks that can be retrieved either from
  - the servers
  - other peers currently retrieving the same video (leechers)

Introduction

In **peer-assisted Video-on-Demand (VoD) systems:**

- users browse a catalog of available videos and asynchronously issue requests to watch a given content
- content is divided into chunks that can be retrieved either from
  - the servers
  - other peers currently retrieving the same video (leechers)
  - peers storing the whole video (seeds)

Introduction

In **peer-assisted Video-on-Demand (VoD) systems:**

- users browse a catalog of available videos and asynchronously issue requests to watch a given content
- content is divided into chunks that can be retrieved either from
  - the servers
  - other peers currently retrieving the same video (leechers)
  - peers storing the whole video (seeds)
- chunks must be retrieved by peers almost in sequence to guarantee small play-out delays

Introduction

In **peer-assisted Video-on-Demand (VoD) systems:**

- users browse a catalog of available videos and asynchronously issue requests to watch a given content
- content is divided into chunks that can be retrieved either from
  - the servers
  - other peers currently retrieving the same video (leechers)
  - peers storing the whole video (seeds)
- chunks must be retrieved by peers almost in sequence to guarantee small play-out delays
- a minimum average download rate equal to the video playback rate must be sustained to guarantee service continuity; the system (exploiting servers bandwidth when needed) is able to steadily meet this constraint.

Assumptions

Assumptions

- Video is downloaded by each user at constant rate d, equal to the playback rate.

Assumptions

- Video is downloaded by each user at constant rate d, equal to the playback rate.
- Upload available bandwidth U_i of peer i is a random variable with a assigned distribution (U_i are i.i.d.).

Assumptions

- Video is downloaded by each user at constant rate d, equal to the playback rate.
- Upload available bandwidth U_i of peer i is a random variable with a assigned distribution (U_i are i.i.d.).
- Users contribute their upload bandwidth to the video distribution as long as they are in the system.

Assumptions

- Video is downloaded by each user at constant rate d, equal to the playback rate.
- Upload available bandwidth U_i of peer i is a random variable with a assigned distribution (U_i are i.i.d.).
- Users contribute their upload bandwidth to the video distribution as long as they are in the system.
- The arrival process of requests (and users) for a video follows a Poisson process of intensity λ.

Assumptions

- Video is downloaded by each user at constant rate d, equal to the playback rate.
- Upload available bandwidth U_i of peer i is a random variable with a assigned distribution (U_i are i.i.d.).
- Users contribute their upload bandwidth to the video distribution as long as they are in the system.
- The arrival process of requests (and users) for a video follows a Poisson process of intensity λ.
- User's sojourn time is described by an arbitrary random variable T with finite mean $\overline{T}$.

Preliminary Observations

- The number of active users N follows a Poisson distribution with mean $\lambda\overline{T}$.

Preliminary Observations

- The number of active users N follows a Poisson distribution with mean $\lambda\overline{T}$.
- Two classes of active users:

Preliminary Observations

- The number of active users N follows a Poisson distribution with mean $\lambda\overline{T}$.
- Two classes of active users:
  - downloading users N_d with mean $\overline{N}_d = \lambda\overline{T}_d$, where $\overline{T}_d$ is the mean time spent downloading the video

Preliminary Observations

- The number of active users N follows a Poisson distribution with mean $\lambda\overline{T}$.
- Two classes of active users:
  - downloading users N_d with mean $\overline{N}_d = \lambda\overline{T}_d$, where $\overline{T}_d$ is the mean time spent downloading the video
  - seeds (users that have completed the download) N_{seed} with mean $\overline{N}_{\text{seed}} = \lambda\overline{T}_{\text{seed}}$, where $\overline{T}_{\text{seed}} = \overline{T} - \overline{T}_d$

Preliminary Observations

- The number of active users N follows a Poisson distribution with mean $\lambda\overline{T}$.
- Two classes of active users:
  - downloading users N_d with mean $\overline{N}_d = \lambda\overline{T}_d$, where $\overline{T}_d$ is the mean time spent downloading the video
  - seeds (users that have completed the download) N_{seed} with mean $\overline{N}_{\text{seed}} = \lambda\overline{T}_{\text{seed}}$, where $\overline{T}_{\text{seed}} = \overline{T} - \overline{T}_d$
- We define the system load as:

$$\gamma = \frac{\overline{N}_d\overline{T}_d}{\overline{N}\ \overline{T}}.$$

Goal

Our goal is:

Goal

Our goal is:

- to characterize the bandwidth requested from the servers S (and its average $\overline{S}$);

Goal

Our goal is:

- to characterize the bandwidth requested from the servers S (and its average $\overline{S}$);
  - we derive analytical upper (and lower) bounds on S,

Goal

Our goal is:

- to characterize the bandwidth requested from the servers S (and its average $\overline{S}$);
  - we derive analytical upper (and lower) bounds on S,
  - we derive exact expressions for S under the assumption that peer upload bandwidth is phase-type distributed,

Goal

Our goal is:

- to characterize the bandwidth requested from the servers S (and its average $\overline{S}$);
  - we derive analytical upper (and lower) bounds on S,
  - we derive exact expressions for S under the assumption that peer upload bandwidth is phase-type distributed,
  - we show that our bounds are tight as number of users increases.

Goal

Our goal is:

- to characterize the bandwidth requested from the servers S (and its average $\overline{S}$);
  - we derive analytical upper (and lower) bounds on S,
  - we derive exact expressions for S under the assumption that peer upload bandwidth is phase-type distributed,
  - we show that our bounds are tight as number of users increases.

We prove the asymptotic optimality of the simple sequential delivery scheme, and conditions for **self-sustainability** when $\lambda \to \infty$

A peer-assisted VoD system in steady state conditions is **self-sustainable** when no server bandwidth is needed to sustain the video distribution.

Analysis

Let S_d be the aggregate bandwidth requested by the downloading users. The aggregate upload bandwidth offered by the seeds is

$$S_{\text{seed}} = \sum_{i=1}^{N_{\text{seed}}} U_i.$$

Analysis

Let S_d be the aggregate bandwidth requested by the downloading users. The aggregate upload bandwidth offered by the seeds is

$$S_{\text{seed}} = \sum_{i=1}^{N_{\text{seed}}} U_i .$$

The bandwidth requested from the servers is:

$$S \triangleq \max\{0, S_d - S_{\text{seed}}\}$$

where S_d is the bandwidth demanded by downloading peers.

Universal Lower Bound

A universal lower bound[1] to $\overline{S}$ for any chunk distribution scheme is

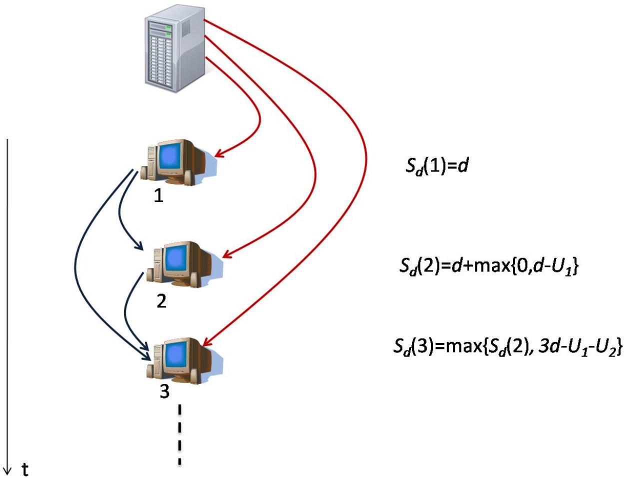
$$\overline{S} \leq \max\{0, d\overline{N}_d - \overline{U}\ \overline{N}\}.$$

Note that the above lower bound is trivially zero for $\gamma < 1$.

Intuition: The additional server bandwidth is given by users requested bandwidth minus their total upload bandwidth.

[1] Note that this trivial lower bound was already shown in: C. Huang, J. Li, and K. W. Ross, Can Internet Video-on-Demand Be Profitable? in ACM SIGCOMM, 2007.

Graphical example



Derivation of Upper Bounds

We define an auxiliary variable as:

$$z_d(k) = \max_{1 \le j \le k} \{\sum_{i=1}^{j} (d - U_i)\}.$$

Derivation of Upper Bounds

We define an auxiliary variable as:

$$z_d(k) = \max_{1 \le j \le k} \{\sum_{i=1}^{j} (d - U_i)\}.$$

Then $S_d(k)$ can be expressed in terms of $z_d(k-1)$ according to:

$$S_d(k) = d + \max\{0, z_d(k-1)\}.$$

Derivation of Upper Bounds

We define an auxiliary variable as:

$$z_d(k) = \max_{1 \le j \le k} \{ \sum_{i=1}^{j} (d - U_i) \}.$$

Then $S_d(k)$ can be expressed in terms of $z_d(k-1)$ according to:

$$S_d(k) = d + \max\{0, z_d(k-1)\}.$$

$z_d(k)$ can be regarded as the maximum value (up to time k) reached by a unidimensional random walk with increments $x_i = d - U_i$.

Derivation of Upper Bounds

We define an auxiliary variable as:

$$z_d(k) = \max_{1 \le j \le k} \{ \sum_{i=1}^{j} (d - U_i) \}.$$

Then $S_d(k)$ can be expressed in terms of $z_d(k-1)$ according to:

$$S_d(k) = d + \max\{0, z_d(k-1)\}.$$

$z_d(k)$ can be regarded as the maximum value (up to time k) reached by a unidimensional random walk with increments $x_i = d - U_i$.

We exploit martingale/random walks theory to upper bound $\mathbb{P}\{z_d(k) \ge w\}$. From which we can derive an upper bound on $\overline{S}_d$ and $\overline{S}$.

Derivation of Upper Bounds (2)

Lemma

*(**Lundberg inequality**) Consider a sequence of i.i.d. variables $(X_i)_{i\geq 1}$, satisfying the following three properties:*

- $\mathbb{E}[X_1] < 0$;
- $\mathbb{P}(X_1 > 0) > 0$,
- $\mathbb{E}[e^{tX_1}]$ *is finite in a neighborhood of the origin.*

Define the r.v. $Q(k) \triangleq \sum_{i=1}^{k} X_i$, $k \geq 1$, $Q(0) \triangleq 0$.
Then, denoted θ^* *the strictly positive solution of* $\mathbb{E}[\mathrm{e}^{\theta^* X_1}] = 1$, *which exists unique under i), ii), and iii), we have, for all* $k \geq 1$:

$$\mathbb{P}\{\max_{1\leq j\leq k} Q(j) > w\} \leq \mathrm{e}^{-\theta^* w}$$

Derivation of Upper Bounds (3)

Lundberg inequality can be generalized and adapted to our context, to obtain an upper bound to $\mathbb{P}(z_d(k) > w)$.

Derivation of Upper Bounds (3)

Lundberg inequality can be generalized and adapted to our context, to obtain an upper bound to $\mathbb{P}(z_d(k) > w)$.
In our case $X_1 = d - U_1$.

Derivation of Upper Bounds (3)

Lundberg inequality can be generalized and adapted to our context, to obtain an upper bound to $\mathbb{P}(z_d(k) > w)$.
In our case $X_1 = d - U_1$.
Observe, however, that:

Derivation of Upper Bounds (3)

Lundberg inequality can be generalized and adapted to our context, to obtain an upper bound to $\mathbb{P}(z_d(k) > w)$.
In our case $X_1 = d - U_1$.
Observe, however, that:

- when $d < \overline{U}$, we can directly apply Lundberg inequality

Derivation of Upper Bounds (3)

Lundberg inequality can be generalized and adapted to our context, to obtain an upper bound to $\mathbb{P}(z_d(k) > w)$.
In our case $X_1 = d - U_1$.
Observe, however, that:

- when $d < \overline{U}$, we can directly apply Lundberg inequality
- $d > \overline{U}$, since $\mathbb{E}[d - U_i] > 0$, we can not apply Lundberg inequality directly. Therefore we need to define an auxiliary sequence of random variables, tightly related to $z_d(k)$, on which Lundberg bound can be applied.

Derivation of Upper Bounds (3)

Lundberg inequality can be generalized and adapted to our context, to obtain an upper bound to $\mathbb{P}(z_d(k) > w)$.
In our case $X_1 = d - U_1$.
Observe, however, that:

- when $d < \overline{U}$, we can directly apply Lundberg inequality
- $d > \overline{U}$, since $\mathbb{E}[d - U_i] > 0$, we can not apply Lundberg inequality directly. Therefore we need to define an auxiliary sequence of random variables, tightly related to $z_d(k)$, on which Lundberg bound can be applied.

The approach of the auxiliary sequence of variables is generalized also to the case $d < \overline{U}$, to obtain a possibly tighter upper bound.

Exact solution

- We can also show that $F_Z(w \mid k) = \mathbb{P}\{z_d(k) \geq x\}$ satisfies the following recursion equation:

$$F_Z(w \mid k) = \int_{-\infty}^{w} F_Z(w - \alpha \mid k-1)\, \mathrm{d}F_Z(\alpha \mid 1). \tag{1}$$

Exact solution

- We can also show that $F_Z(w \mid k) = \mathbb{P}\{z_d(k) \geq x\}$ satisfies the following recursion equation:

$$F_Z(w \mid k) = \int_{-\infty}^{w} F_Z(w - \alpha \mid k-1)\,\mathrm{d}F_Z(\alpha \mid 1). \tag{1}$$

- An analytical solution of $F_Z(w \mid k)$ is obtained when U is phase-type distributed.

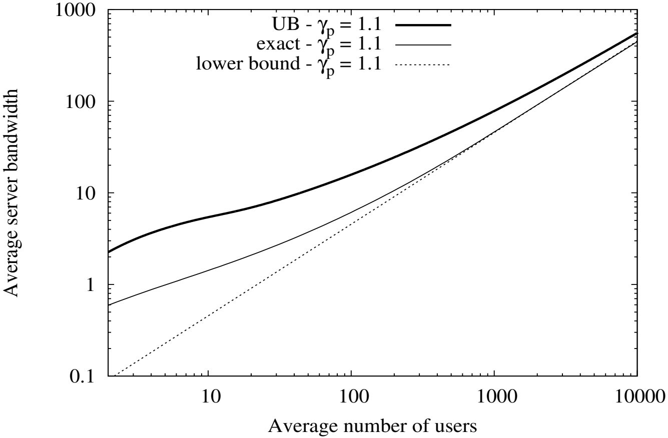
Exact solution

- We can also show that $F_Z(w \mid k) = \mathbb{P}\{z_d(k) \geq x\}$ satisfies the following recursion equation:

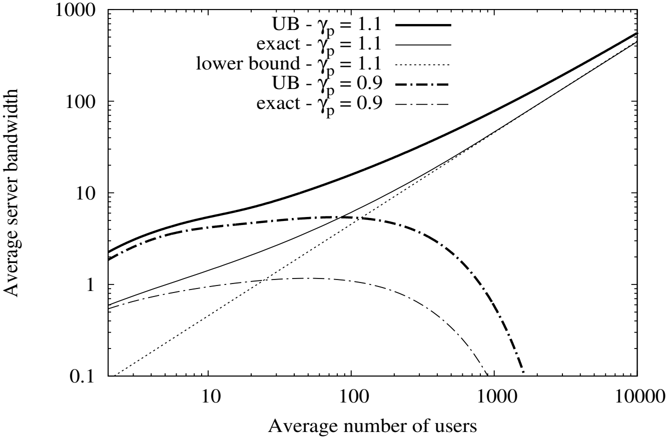
$$F_Z(w \mid k) = \int_{-\infty}^{w} F_Z(w - \alpha \mid k - 1)\, \mathrm{d}F_Z(\alpha \mid 1). \qquad (1)$$

- An analytical solution of $F_Z(w \mid k)$ is obtained when U is phase-type distributed.
- Then unconditioning over k (we recall the number of downloaders is Poisson distributed) we obtain an exact expression for the distribution of S_d.

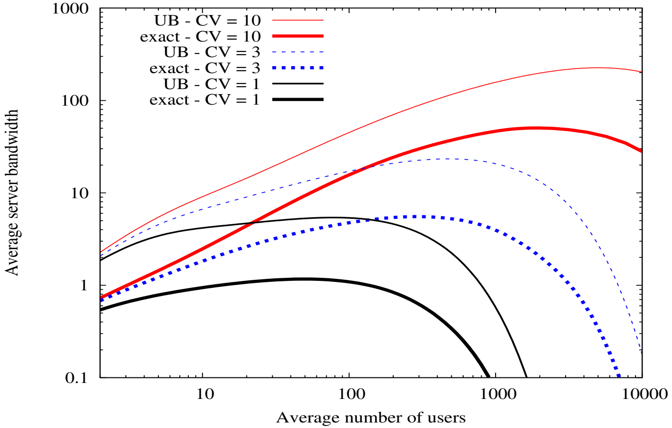
Results


UB - γp = 1.1
exact - γp = 1.1
lower bound - γp = 1.1
1000
100
10
1
0.1
10
100
1000
10000
Average server bandwidth
Average number of users

Results


UB - γ_p = 1.1
exact - γ_p = 1.1
lower bound - γ_p = 1.1
UB - γ_p = 0.9
exact - γ_p = 0.9
1000
100
10
1
0.1
10
100
1000
10000
Average server bandwidth
Average number of users

Results ($\gamma = 0.9$)



Asymptotic Results

Theorem

When $\lambda \to \infty$ the following asymptotic regimes hold for any chunk distribution scheme: For $\gamma < 1$ and $\overline{T} > \overline{T}_d$, the system is **self-sustainable** ($\lim_{\lambda\to\infty} \overline{S} = 0$); For $\gamma > 1$, $\overline{S}$ grows linearly with the number of users. In particular, $\lim_{\lambda\to\infty} \overline{S}/(d\overline{N}_d - \overline{U}\ \overline{N}) = 1$.

Conclusions and Future Work

- We derived general upper bounds to the bandwidth requested from the servers in the case of sequential chunk delivery.

Conclusions and Future Work

- We derived general upper bounds to the bandwidth requested from the servers in the case of sequential chunk delivery.
- Our bounds tightly characterize the behavior of peer-assisted VoD systems at steady-state as the number of users increases, proving the self-sustainability of systems in the bandwidth surplus mode.

Conclusions and Future Work

- We derived general upper bounds to the bandwidth requested from the servers in the case of sequential chunk delivery.
- Our bounds tightly characterize the behavior of peer-assisted VoD systems at steady-state as the number of users increases, proving the self-sustainability of systems in the bandwidth surplus mode.
- We showed that the simple sequential delivery scheme is asymptotically optimal, under weak assumptions on the user behavior.

Conclusions and Future Work

- We derived general upper bounds to the bandwidth requested from the servers in the case of sequential chunk delivery.
- Our bounds tightly characterize the behavior of peer-assisted VoD systems at steady-state as the number of users increases, proving the self-sustainability of systems in the bandwidth surplus mode.
- We showed that the simple sequential delivery scheme is asymptotically optimal, under weak assumptions on the user behavior.
- We derived exact estimate of the bandwidth requested from the servers under the assumption that the peer upload bandwidth is phase-type distributed.

Conclusions and Future Work

- We derived general upper bounds to the bandwidth requested from the servers in the case of sequential chunk delivery.
- Our bounds tightly characterize the behavior of peer-assisted VoD systems at steady-state as the number of users increases, proving the self-sustainability of systems in the bandwidth surplus mode.
- We showed that the simple sequential delivery scheme is asymptotically optimal, under weak assumptions on the user behavior.
- We derived exact estimate of the bandwidth requested from the servers under the assumption that the peer upload bandwidth is phase-type distributed.
- We are extending our analysis to non stationary scenarios where file popularity changes with time, i.e., $\lambda(t)$.

Thank you!