

# Probabilistic Resources Allocation Based on a Large Deviation Principle

**Paulo Gonçalves**

DANTE - INRIA - ENS Lyon

Thomas Begin (Assist. Prof. at Univ. Lyon 1, DANTE)

Shubhabrata Roy (PhD DANTE, 2010-2013)

Patrick Loiseau (Assist. Prof. at EURECOM)

**Journées Scientifiques Inria – June 25-26, 2013**

# Workload Volatility

**Context** Applications that undergo **highly time-varying workloads** (e.g. Buzz demand in a VoD system)

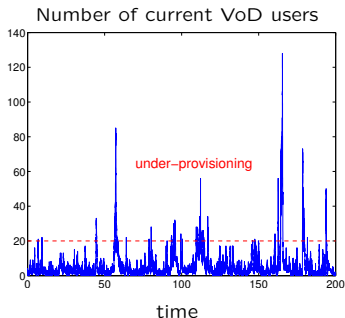
# Workload Volatility

- Context** Applications that undergo **highly time-varying workloads** (e.g. Buzz demand in a VoD system)
- Goal** **Dynamic resource allocation** yielding a good compromise between **capex** and **opex** costs

# Workload Volatility

**Context** Applications that undergo **highly time-varying workloads** (e.g. Buzz demand in a VoD system)

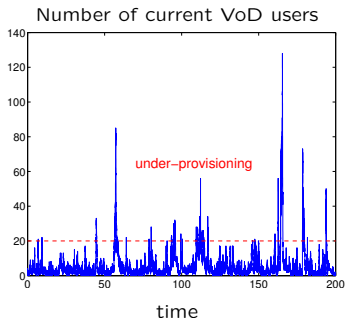
**Goal** **Dynamic resource allocation** yielding a good compromise between **capex** and **opex** costs



# Workload Volatility

**Context** Applications that undergo **highly time-varying workloads** (e.g. Buzz demand in a VoD system)

**Goal** **Dynamic resource allocation** yielding a good compromise between **capex** and **opex** costs

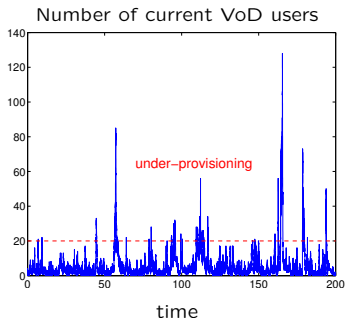


**Approach** Combine the three ingredients:

# Workload Volatility

**Context** Applications that undergo **highly time-varying workloads** (e.g. Buzz demand in a VoD system)

**Goal** **Dynamic resource allocation** yielding a good compromise between **capex** and **opex** costs



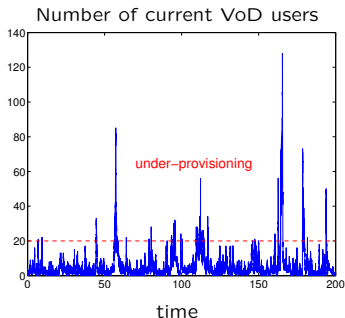
**Approach** Combine the three ingredients:

- A **sensible** (epidemic) model to catch the **burstiness** and the **dynamics** of the workload

# Workload Volatility

**Context** Applications that undergo **highly time-varying workloads** (e.g. Buzz demand in a VoD system)

**Goal** **Dynamic resource allocation** yielding a good compromise between **capex** and **opex** costs



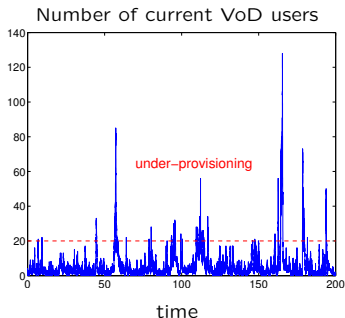
**Approach** Combine the three ingredients:

- A **sensible** (epidemic) model to catch the **burstiness** and the **dynamics** of the workload
- A (Markov) model that verifies a **large deviations principle**

# Workload Volatility

**Context** Applications that undergo **highly time-varying workloads** (e.g. Buzz demand in a VoD system)

**Goal** **Dynamic resource allocation** yielding a good compromise between **capex** and **opex** costs



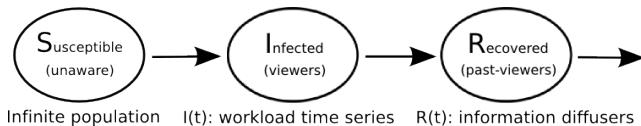
**Approach** Combine the three ingredients:

- A **sensible** (epidemic) model to catch the **burstiness** and the **dynamics** of the workload
- A (Markov) model that verifies a **large deviations principle**
- A **probabilistic** management policy based on the large deviation characterisation



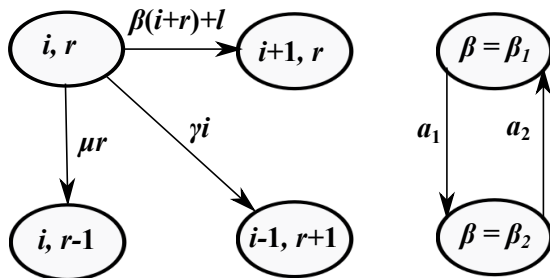
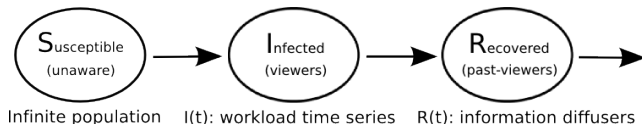
# An epidemic based model for volatile workload

A hidden state Markov process with memory effect [IEICE 2012, TRAC 2013]



# An epidemic based model for volatile workload

A hidden state Markov process with memory effect [IEICE 2012, TRAC 2013]

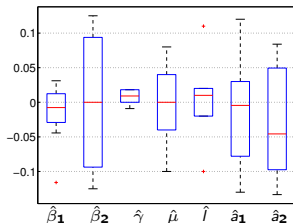


$i$ : current # of viewers /  $r$ : current # of infected

# Model identification and evaluation

A MCMC based estimation procedure for the model's parameters [Gretsi 2013]

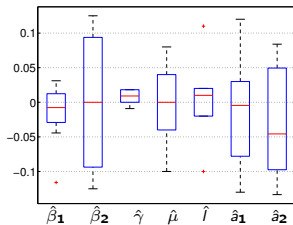
Param. estimation precision



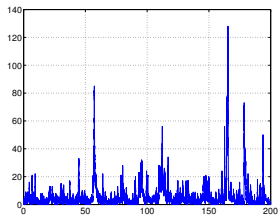
# Model identification and evaluation

A MCMC based estimation procedure for the model's parameters [Gretsi 2013]

Param. estimation precision



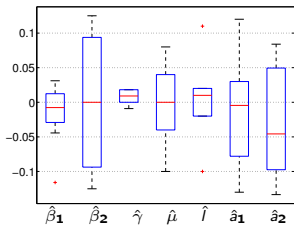
VoD workload - trace I



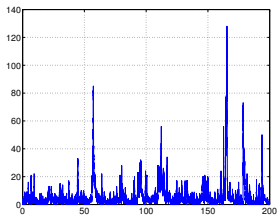
# Model identification and evaluation

A MCMC based estimation procedure for the model's parameters [Gretsi 2013]

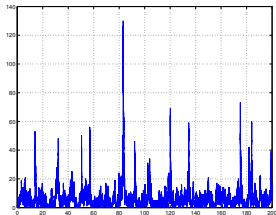
Param. estimation precision



VoD workload - trace I



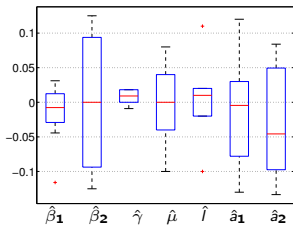
Proposed Model



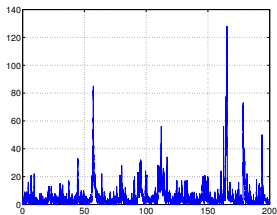
# Model identification and evaluation

A MCMC based estimation procedure for the model's parameters [Gretsi 2013]

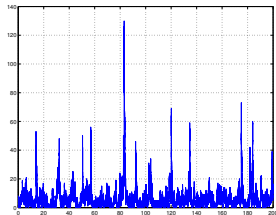
Param. estimation precision



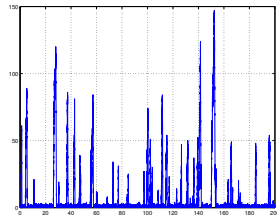
VoD workload - trace I



Proposed Model



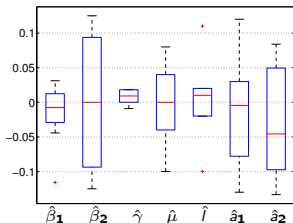
MMPP/M/1



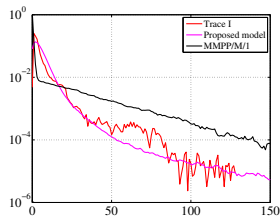
# Model identification and evaluation

A MCMC based estimation procedure for the model's parameters [Gretsi 2013]

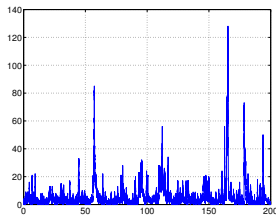
Param. estimation precision



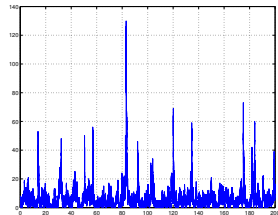
Steady state distribution



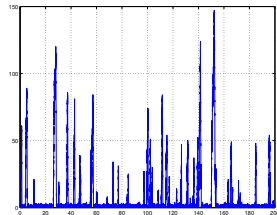
VoD workload - trace I



Proposed Model



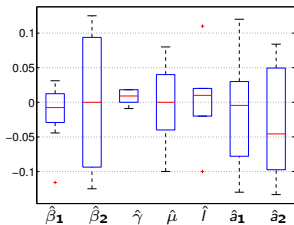
MMPP/M/1



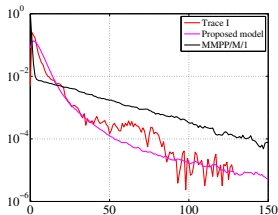
# Model identification and evaluation

A MCMC based estimation procedure for the model's parameters [Gretsi 2013]

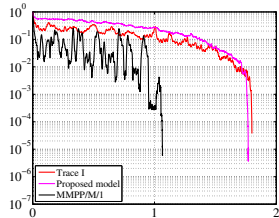
Param. estimation precision



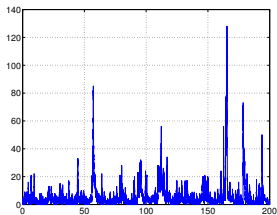
Steady state distribution



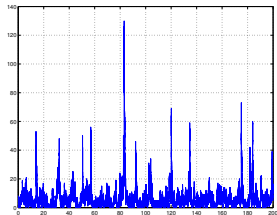
Autocorrelation function



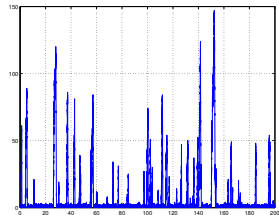
VoD workload - trace I



Proposed Model



MMPP/M/1





# Large Deviations Principle

A process  $I_t$  verifies a large deviations principle:

$$\mathbb{P}\{\langle I_t \rangle_\tau \in [\alpha - \varepsilon_\tau, \alpha + \varepsilon_\tau]\} \sim \exp(\tau \cdot f(\alpha)), \quad \tau \rightarrow \infty$$

$\tau$  : average time scale

$f(\alpha)$  : large deviations spectrum of  $I_t$

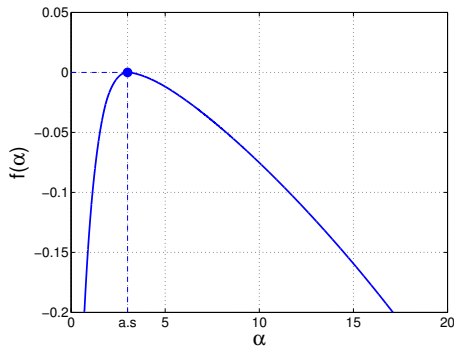
# Large Deviations Principle

A process  $I_t$  verifies a large deviations principle:

$$\mathbb{P}\{\langle I_t \rangle_\tau \in [\alpha - \varepsilon_\tau, \alpha + \varepsilon_\tau]\} \sim \exp(\tau \cdot f(\alpha)), \quad \tau \rightarrow \infty$$

$\tau$  : average time scale

$f(\alpha)$  : large deviations spectrum of  $I_t$



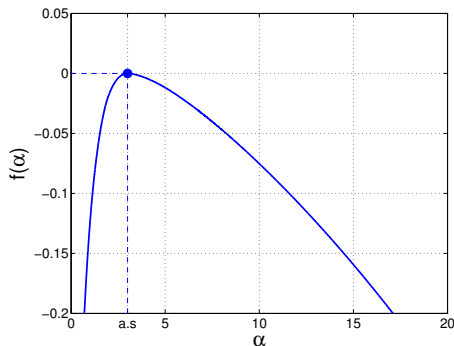
# Large Deviations Principle

A process  $I_t$  verifies a large deviations principle:

$$\mathbb{P}\{\langle I_t \rangle_\tau \in [\alpha - \varepsilon_\tau, \alpha + \varepsilon_\tau]\} \sim \exp(\tau \cdot f(\alpha)), \quad \tau \rightarrow \infty$$

$\tau$  : average time scale

$f(\alpha)$  : large deviations spectrum of  $I_t$



"Dynamic" implies time scale: a notion that is explicit in large deviations principle

# Two important assets for large deviations utility

General result ( "Large deviations for the local fluctuations of random walks", J. Barral, P. Loiseau, *Stochastic Processes and their Applications*, 2011)

*A wide class of processes (stationary & mixing) verifies an empirical large deviations principle. In particular, this results holds true for any time series that can reliably be modelled by an irreducible, aperiodic Markov process.*

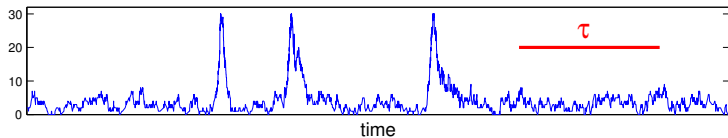
Theoretical spectrum from the Markov transition matrix (i.e. from the model parameters)

Theorem ( "On the estimation of the Large Deviations spectrum", J. Barral, P. G., *J. stat. Phys.*, 2011)

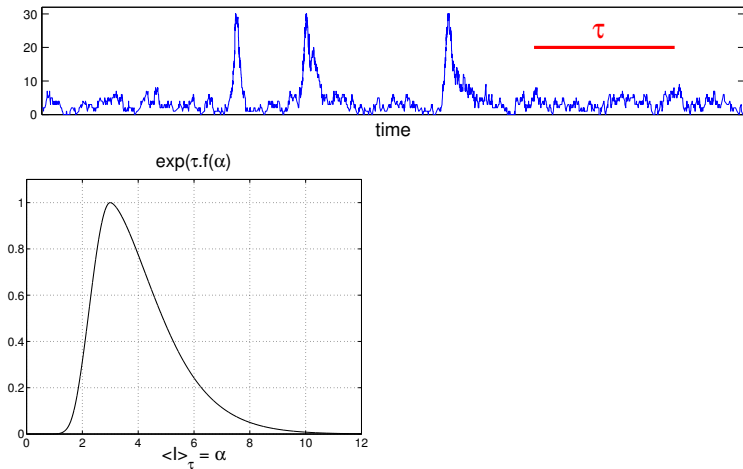
*We derived a consistent estimator of the large deviation spectrum from a finite size time series (observation samples). We proved convergence on mathematical objects with scale invariance properties (multifractal measures and processes).*

Empirical estimation from a finite length trace

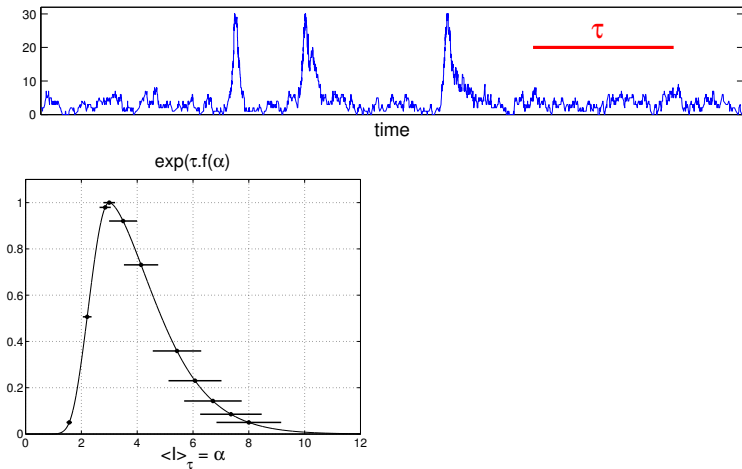
# Overflow propability



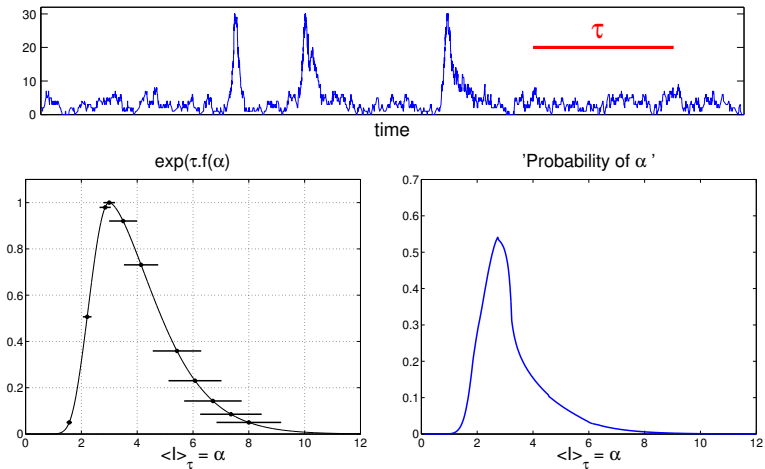
# Overflow probability



# Overflow probability

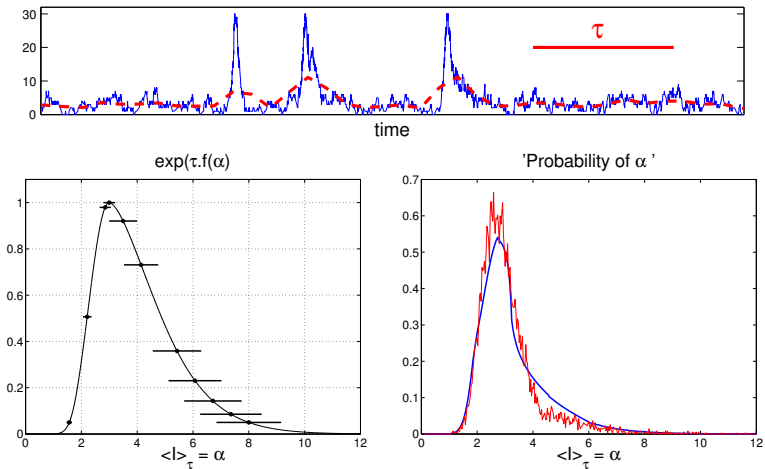


# Overflow probability

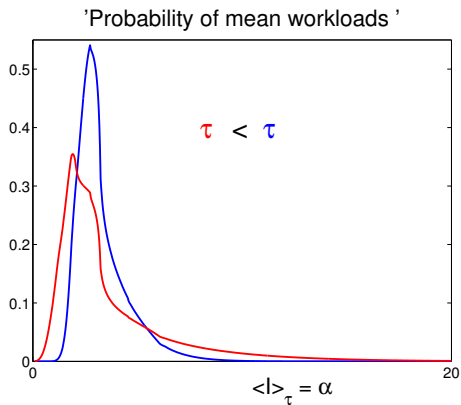




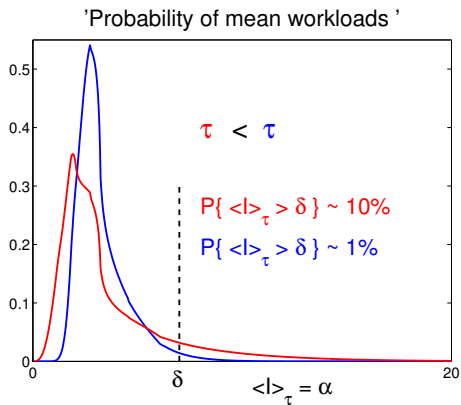
# Overflow probability



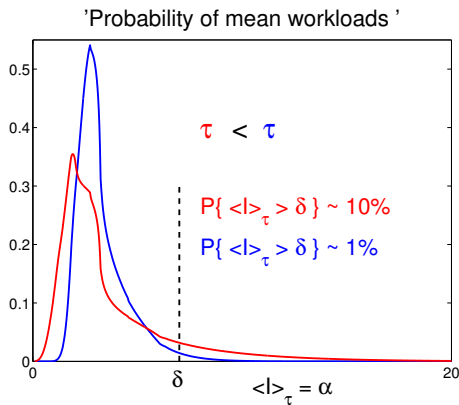
# Probabilistic resource provisioning



# Probabilistic resource provisioning

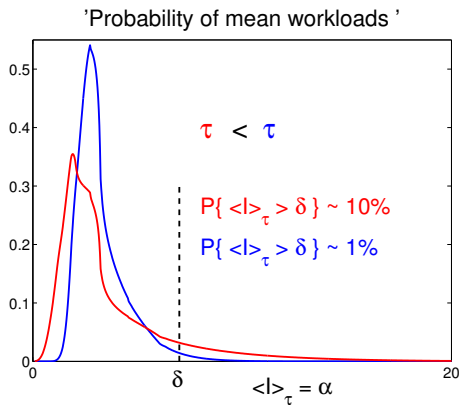


# Probabilistic resource provisioning



→ Resource provisioning based on a **time scale** dependent performance evaluation

# Probabilistic resource provisioning



→ Resource provisioning based on a **time scale** dependent performance evaluation

→ **Dynamic management**