Some Scaling Properties of Traffic in Communication Networks

Paulo Gonçalves
DANTE - INRIA - ENS Lyon

Patrick Loiseau (PhD, 2006-2009)
Shubhabrata Roy (PhD, 2010-2013)
M. Sokol (PhD, 2010-)
B. Girault (PhD, 2012-2015)

Seminars Complex Networks, LIP6, UPMC – march 7, 2013

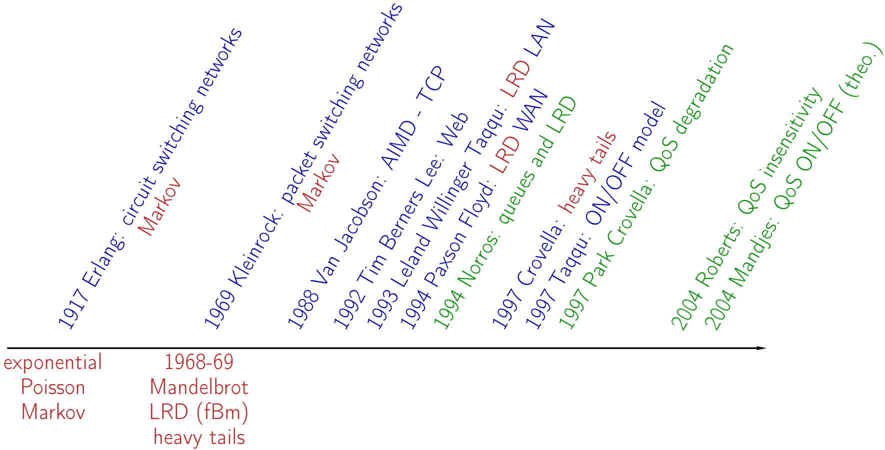
Historical perspective

- 1917 Erlang: circuit switching networks
  Markov
- 1969 Kleinrock: packet switching networks
  Markov
- 1988 Van Jacobson: AIMD - TCP
- 1992 Tim Berners Lee: Web
- 1993 Leland Willinger Taqqu: LRD LAN
- 1994 Paxson Floyd: LRD WAN
- 1997 Crovella: heavy tails
- 1997 Taqqu: ON/OFF model

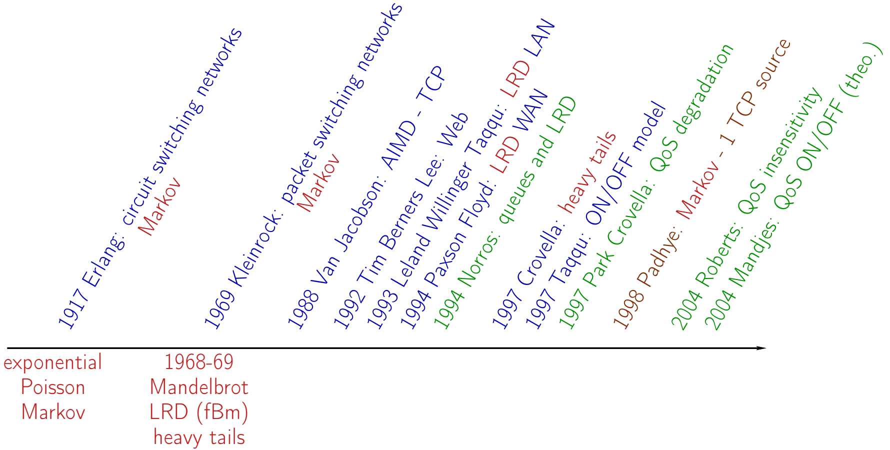
exponential
Poisson
Markov

1968-69
Mandelbrot
LRD (fBm)
heavy tails

Historical perspective



Historical perspective



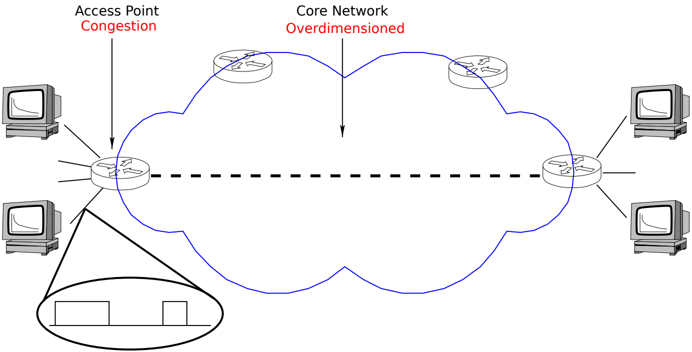
Some open questions:

- Long Range Dependence / Heavy Tailed distributions impact on QoS ?
- Existing models (e.g. Padhye) only predict mean metrics (e.g. throughput) : what about variability?

Our approach

To combine theoretical models with controlled experiments in realistic environments and real-world traffic traces

Simplified System



- Congestion essentially arises at the access points
  $\rightarrow$ Simplified System : single bottleneck
- Users' behavior : ON/OFF source model
- *MetroFlux*: a probe for traffic capture at packet level (O. Goga,. . .)

Long memory in aggregated traffic: the Taqqu model

- Heavy-tailed distributed ON periods: heavy tail index $\alpha_{ON} > 1$

Theorem (Taqqu, Willinger, Sherman, 1997)

In the limit of a large number of sources N_{src}, if:

- *flow throughput is constant,*
- *same throughput for all flows ;*

aggregated bandwidth $B^{(\Delta)}(t)$ is long range dependent, with parameter:

$$H = \max\left(\frac{3-\alpha_{ON}}{2}\ ,\ \frac{1}{2}\right)$$

Long memory: long range correlation $(H > 1/2)$

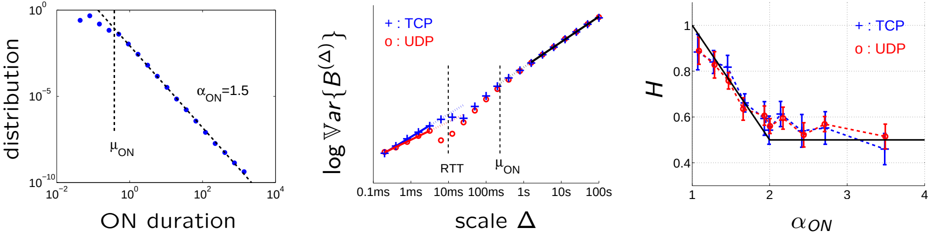
$$Cov_{B^{(\Delta)}}(\tau) = \mathbb{E}\left\{B^{(\Delta)}(t)B^{(\Delta)}(t+\tau)\right\} \underset{\tau\to\infty}{\sim} \tau^{(2H-2)}$$

Variance grows faster than Δ: $\mathbb{V}ar\left\{B^{(\Delta)}(t)\right\} \sim \Delta^{2H}$

Theorem validation on a realistic environment

- Controlled experiment: *MetroFlux* 1 Gbps, 100 sources, 8 hours traffic
- UDP/TCP: throughput limited to 5 Mbps (no congestion)

ON Distribution (source) | Log-diagram (aggregated traffic) | Taqqu Prediction

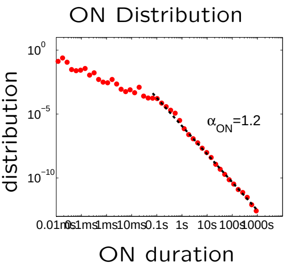


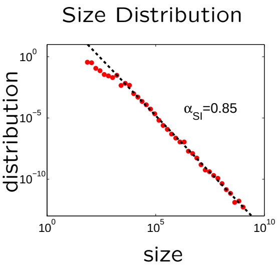
⇒ Protocol has no influence at large scales

⇒ Long memory shows up beyond scale $\Delta = \mu_{ON}$ (mean flow duration)

Influence of flow mean throughput / duration correlation

- Web traffic acquired at in2p3 (Lyon) with *MetroFlux* 10 Gbps

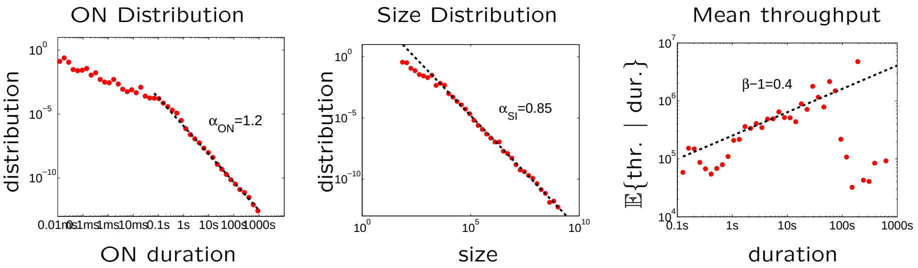




- Heavy-tailed ON periods, $\alpha_{ON} = 1.2$
- Heavy tailed flow sizes, $\alpha_{SI} = 0.85$

Influence of flow mean throughput / duration correlation

- Web traffic acquired at in2p3 (Lyon) with *MetroFlux* 10 Gbps



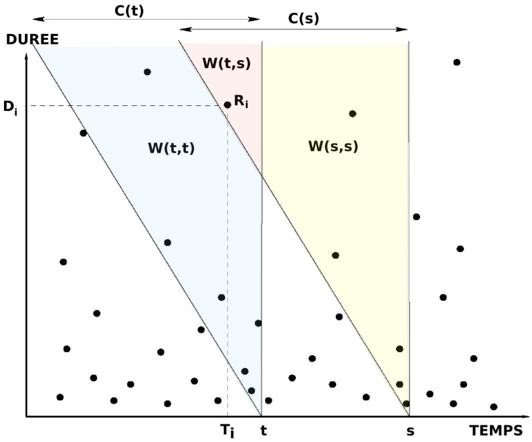
- Heavy-tailed ON periods, $\alpha_{ON} = 1.2$
- Heavy tailed flow sizes, $\alpha_{SI} = 0.85$
- Flow throughput and duration are correlated:

$$\mathbb{E}\{\text{thr.}|\text{dur.}\} \propto (\text{dur.})^{\beta-1}, \quad \beta = \alpha_{ON}/\alpha_{SI} \ \ (= 1.4)$$

$\Rightarrow$ Which heavy tail index does control LRD ? (α_{ON}, α_{SI}) ?

Taqqu model extension

- Planar Poisson process to describe arrival instant *vs* duration



Taqqu model extension

- Planar Poisson process to describe arrival instant *vs* duration

Proposition (LGVBP, 2009)

Model: $\mathbb{E}\{through.|dur.\} = M \cdot (dur.)^{\beta-1}$; $\mathbb{V}ar\{through.|dur.\} = V$

$$Cov_{B(\Delta)}(\tau) = CM^2\tau^{-(\alpha_{ON}-2(\beta-1))+1} + C'V\tau^{-\alpha_{ON}+1}$$

Taqqu model extension

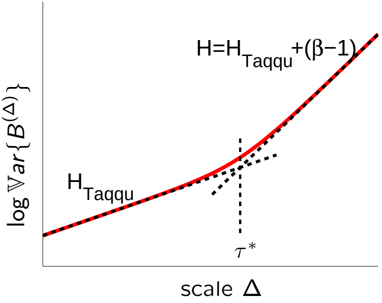
- Planar Poisson process to describe arrival instant *vs* duration

Proposition (LGVBP, 2009)

Model: $\mathbb{E}\{through.|dur.\} = M \cdot (dur.)^{\beta-1}$; $\mathbb{V}ar\{through.|dur.\} = V$

$$Cov_{B(\Delta)}(\tau) = CM^2 \tau^{-(\alpha_{ON}-2(\beta-1))+1} + C'V\tau^{-\alpha_{ON}+1}$$

Log-diagram, $\beta > 1$



- threshold $\tau^* = \left(\frac{C'V}{CM^2}\right)^{1/(2(\beta-1))}$

Taqqu model extension

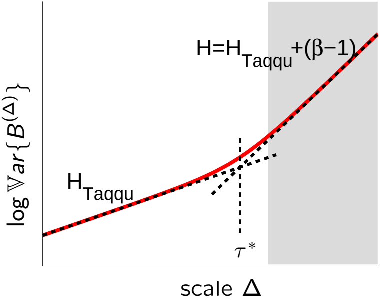
- Planar Poisson process to describe arrival instant *vs* duration

Proposition (LGVBP, 2009)

Model: $\mathbb{E}\{through.|dur.\} = M \cdot (dur.)^{\beta-1}$; $\mathbb{V}ar\{through.|dur.\} = V$

$$Cov_{B(\Delta)}(\tau) = CM^2\tau^{-(\alpha_{ON}-2(\beta-1))+1} + C'V\tau^{-\alpha_{ON}+1}$$

Log-diagram, $\beta > 1$



- threshold $\tau^* = \left(\frac{C'V}{CM^2}\right)^{1/(2(\beta-1))}$

  $\rightarrow$ if $\Delta \gg \tau^*$: $H = H_{\text{Taqqu}} + (\beta - 1)$

Taqqu model extension

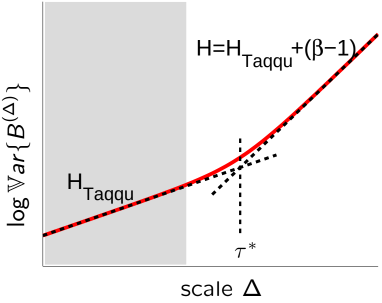
- Planar Poisson process to describe arrival instant *vs* duration

Proposition (LGVBP, 2009)

Model: $\mathbb{E}\{through.|dur.\} = M \cdot (dur.)^{\beta-1}$; $\mathbb{V}ar\{through.|dur.\} = V$

$$Cov_{B(\Delta)}(\tau) = CM^2\tau^{-(\alpha_{ON}-2(\beta-1))+1} + C'V\tau^{-\alpha_{ON}+1}$$

Log-diagram, $\beta > 1$



- threshold $\tau^* = \left(\frac{C'V}{CM^2}\right)^{1/(2(\beta-1))}$

  $\rightarrow$ if $\Delta \gg \tau^*$: $H = H_{\text{Taqqu}} + (\beta - 1)$

  $\rightarrow$ if $\Delta \ll \tau^*$: $H = H_{\text{Taqqu}}$

Taqqu model extension

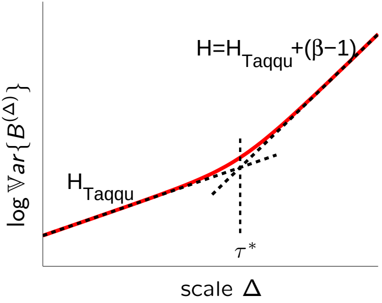
- Planar Poisson process to describe arrival instant *vs* duration

Proposition (LGVBP, 2009)

Model: $\mathbb{E}\{through.|dur.\} = M \cdot (dur.)^{\beta-1}$; $\mathbb{V}ar\{through.|dur.\} = V$

$$Cov_{B(\Delta)}(\tau) = CM^2\tau^{-(\alpha_{ON}-2(\beta-1))+1} + C'V\tau^{-\alpha_{ON}+1}$$

Log-diagram, $\beta > 1$



- Correlations intensify LRD ($\beta > 1$)
- Traffic evolution, future Internet: "flow-aware" control mechanisms, FTTH

LRD impact on QoS: a brief (experimental) outlook

The situation is complex. . .

LRD impact on QoS: a brief (experimental) outlook

The situation is complex. . .

- Negative on finite queues with UDP flows [cf. Mandjes, 2004 (infinite queues)]
  - LRD degrades QoS for large queue sizes (beyond some threshold)
  - but the threshold depends on the considered QoS metric (loss rate *vs* mean load)

LRD impact on QoS: a brief (experimental) outlook

The situation is complex...

- Negative on finite queues with UDP flows [cf. Mandjes, 2004 (infinite queues)]
  - LRD degrades QoS for large queue sizes (beyond some threshold)
  - but the threshold depends on the considered QoS metric (loss rate *vs* mean load)
- Questionable with TCP flows: [Park, 1997] against [Ben Fredj, 2001]
  - LRD has contradictory effects on QoS metrics depending on:

	with slow start	without slow start
Delay	$\searrow$	$\nearrow$
loss rate	$\searrow$	$\rightarrow$
mean throughput	$\rightarrow$	$\nearrow$

  - Heavy tailed distributions (i.e LRD) can favour QoS for large flows

LRD impact on QoS: a brief (experimental) outlook

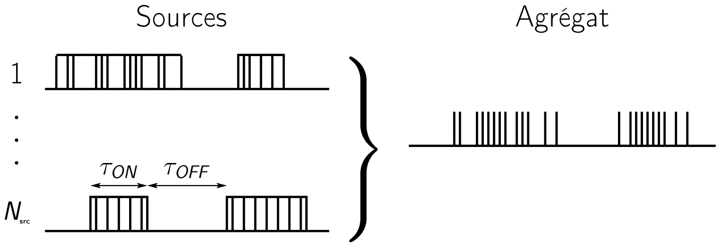
The situation is complex. . .

- Negative on finite queues with UDP flows [cf. Mandjes, 2004 (infinite queues)]
  - LRD degrades QoS for large queue sizes (beyond some threshold)
  - but the threshold depends on the considered QoS metric (loss rate *vs* mean load)
- Questionable with TCP flows: [Park, 1997] against [Ben Fredj, 2001]
  - LRD has contradictory effects on QoS metrics depending on:

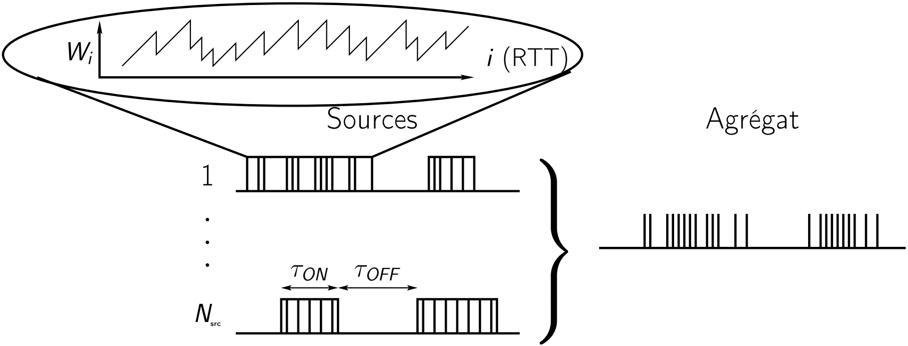
	with slow start	without slow start
Delay	$\searrow$	$\nearrow$
loss rate	$\searrow$	$\rightarrow$
mean throughput	$\rightarrow$	$\nearrow$

  - Heavy tailed distributions (i.e LRD) can favour QoS for large flows
- But in general, QOS is a complex function of multiple variables

Second level of description : single TCP source traffic



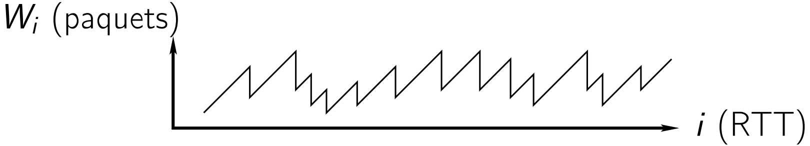
Second level of description : single TCP source traffic



- single TCP source traffic detail
- Long-lived flow → stationary regime

⇒ How to characterize the congestion window evolution?

Markov model

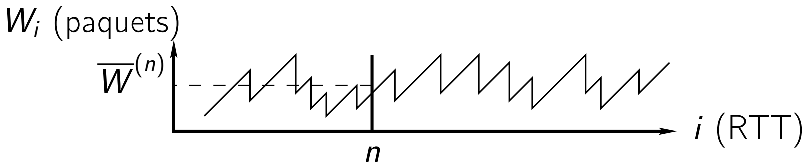


- long-lived flow stationary regime: AIMD
- model: $(W_i)_{i \geq 1}$ finite Markov chain (irreducible, aperiodic), transition matrix Q :

$$\begin{cases} Q_{w,\min(w+1,w_{\mathbf{max}})} & = & 1 - p(w), \\ Q_{w,\max(\lfloor w/2 \rfloor,1)} & = & p(w). \end{cases}$$

- $p(\cdot)$ loss probability of at least one packet, only depends on the current congestion window (hyp.)
- Example: [Padhye, 1998] Bernoulli loss: $p(w) = 1 - (1 - p_{pkt})^w$

Almost sure mean throughput



- mean throughput at scale n (RTT): $\overline{W}^{(n)} = \frac{\sum_{i=1}^{n} W_i}{n}$

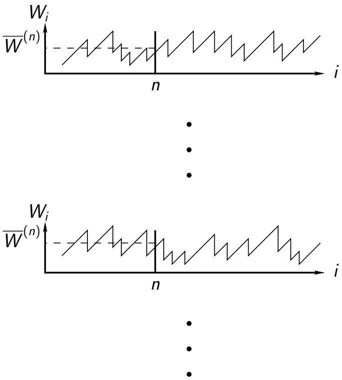
Ergodic Birkhoff theorem (1931): almost sure mean

For almost all realisation, the mean throughput at scale n converges towards a value corresponding to the expectation of the invariant distribution:

$$\overline{W}^{(n)} \xrightarrow[n\to\infty]{p.s.} \overline{W}^{(\infty)} = \mathbb{E}\{W_i\}$$

- Example: [Padhye, 1998], $\overline{W}^{(\infty)} \underset{p_{pkt}\to 0}{\sim} \sqrt{\frac{3}{2p_{pkt}}}$ (RTT=1, MSS=1)

Throughput variability: Large Deviations



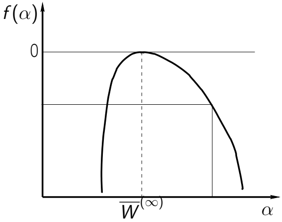
- $\overline{W}^{(n)} \simeq \alpha \neq \overline{W}^{(\infty)}$ Rare events

Large Deviations theorem (Ellis, 84)

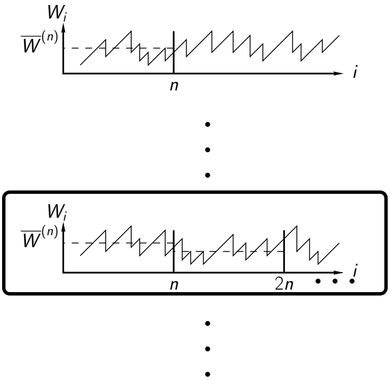
$$\mathbb{P}(\overline{W}^{(n)} \simeq \alpha) \underset{n \to \infty}{\sim} \exp(n \cdot f(\alpha))$$

- $f(\alpha)$ Large Deviation spectrum

→ Scale invariant quantity



Throughput variability: Large Deviations



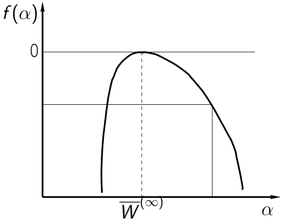
- $\overline{W}^{(n)} \simeq \alpha \neq \overline{W}^{(\infty)}$ Rare events

Large Deviations theorem (Ellis, 84)

$$\mathbb{P}(\overline{W}^{(n)} \simeq \alpha) \underset{n \to \infty}{\sim} \exp(n \cdot f(\alpha))$$

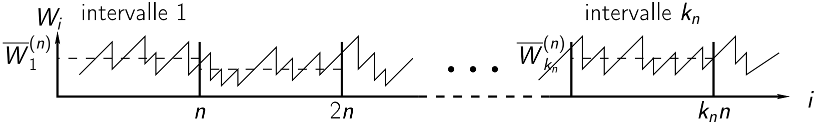
- $f(\alpha)$ Large Deviation spectrum

$\rightarrow$ Scale invariant quantity



$\Rightarrow$ Does a similar theorem exist for a single realization?

Large Deviation on almost all realizations



Large Deviation theorem on almost all realisations (Loiseau et al., 2010)

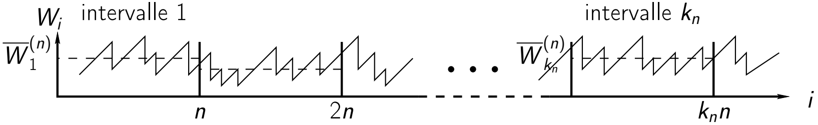
For a given α, if $k_n \geq e^{nR(\alpha)}$, then a.s.

$$\frac{\#\left\{j \in \{1, \cdots, k_n\} : \overline{W}_j^{(n)} \simeq \alpha\right\}}{k_n} \underset{n\to\infty}{\sim} \exp(n \cdot f(\alpha))$$

- "Price to pay": exponential increase of the number of intervals
- Finite realization (of size N): $nk_n = N$

$\Rightarrow$ $[\alpha_{\min}(n), \alpha_{\max}(n)]$ support of observable spectrum at scale n

Large Deviation on almost all realizations



Large Deviation theorem on almost all realisations (Loiseau et al., 2010)

For a given α, if $k_n \geq e^{nR(\alpha)}$, then a.s.

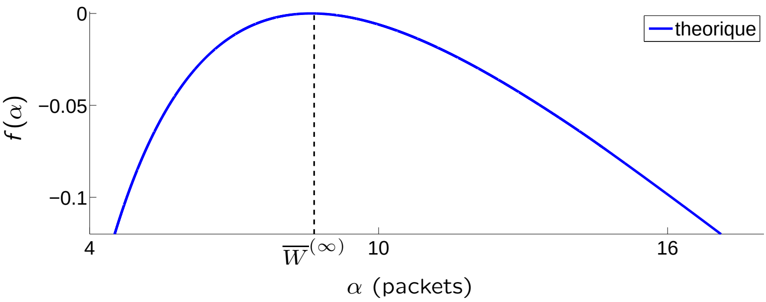
$$\frac{\#\left\{j \in \{1, \cdots, k_n\} : \overline{W}_j^{(n)} \simeq \alpha\right\}}{k_n} \underset{n\to\infty}{\sim} \exp(n \cdot f(\alpha))$$

- "Price to pay": exponential increase of the number of intervals
- Finite realization (of size N): $nk_n = N$

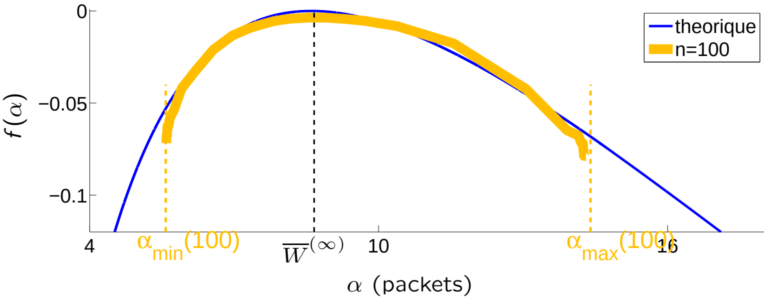
$\Rightarrow$ $[\alpha_{\min}(n), \alpha_{\max}(n)]$ support of observable spectrum at scale n

- Theory: $p(\cdot) \to Q \to f(\alpha), R(\alpha), \alpha_{\min}, \alpha_{\max}$
- Practice: $(W_i)_{i\leq N} \to$ observed distribution

Results: example of Bernoulli losses ($p_{\mathrm{pkt}} = 0.02$)

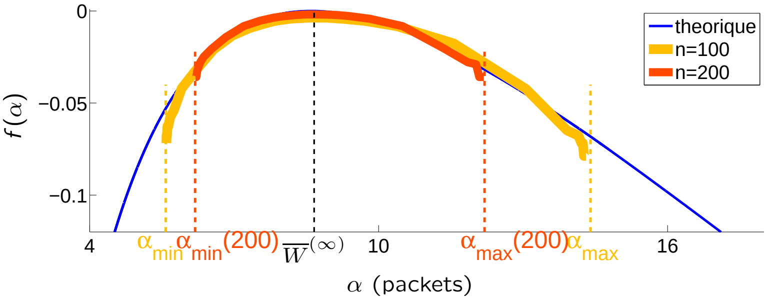


Results: example of Bernoulli losses ($p_{\text{pkt}} = 0.02$)



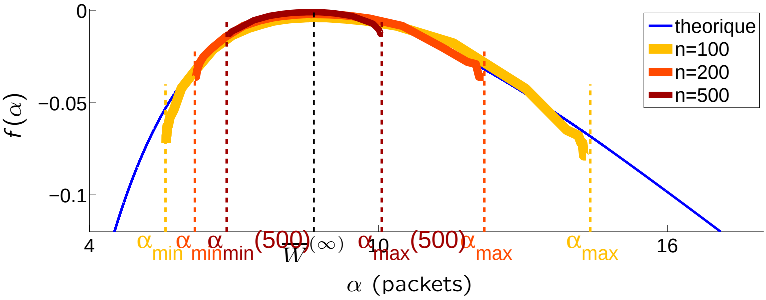
- Apex: almost sure mean: 8.6 packets (Padhye: $\sqrt{\frac{3}{2p_{pkt}}} = 8.66$)

Results: example of Bernoulli losses ($p_{\text{pkt}} = 0.02$)



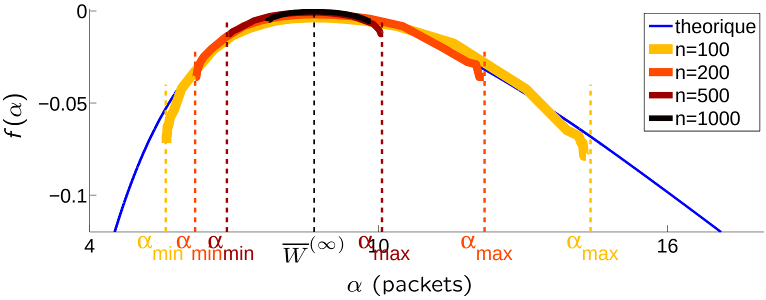
- Apex: almost sure mean: 8.6 packets (Padhye: $\sqrt{\frac{3}{2p_{pkt}}} = 8.66$)

Results: example of Bernoulli losses ($p_{\text{pkt}} = 0.02$)



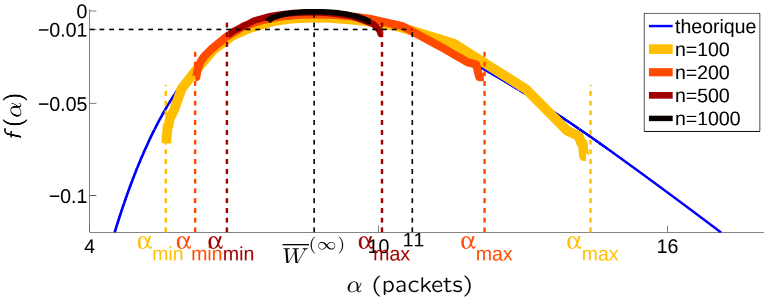
- Apex: almost sure mean: 8.6 packets (Padhye: $\sqrt{\frac{3}{2p_{pkt}}} = 8.66$)

Results: example of Bernoulli losses ($p_{\text{pkt}} = 0.02$)



- Apex: almost sure mean: 8.6 packets (Padhye: $\sqrt{\frac{3}{2p_{pkt}}} = 8.66$)
- Superimposition at different scales → scale invariance

Results: example of Bernoulli losses ($p_{\text{pkt}} = 0.02$)



- Apex: almost sure mean: 8.6 packets (Padhye: $\sqrt{\frac{3}{2p_{pkt}}} = 8.66$)
- Superimposition at different scales → scale invariance
- beyond $n = 100$: variability

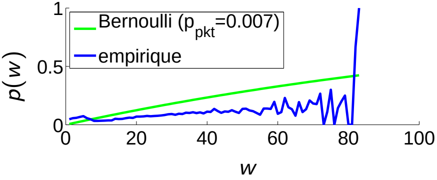
  $n = 100$, portion of intervals with mean ~ 11: $e^{-100\times 0.01} = 0.37$

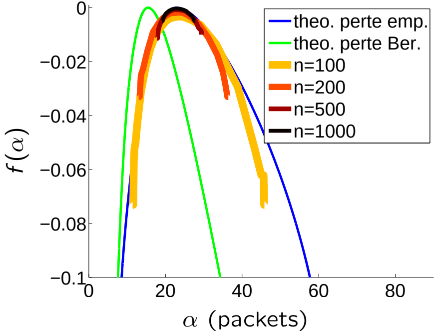
  $n = 200$, portion of intervals with mean ~ 11: $e^{-200\times 0.01} = 0.14$

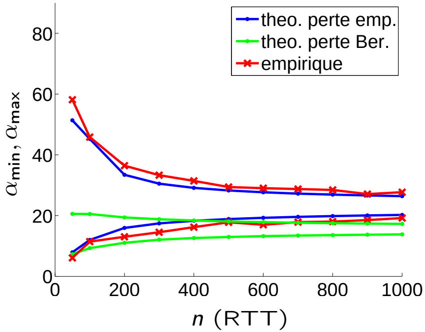
⇒ More accurate information than the almost sure mean

Results II: case of a long-lived flow

- losses: not Bernoulli
- empirical losses







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 deviation principle. In particular, this results holds true any time series that can reliably be modelled by an irreducible, aperiodic Markov process.

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).

An epidemic based model for volatile workload

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

An epidemic based model for volatile workload

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

Approach – Combine the three ingredients:

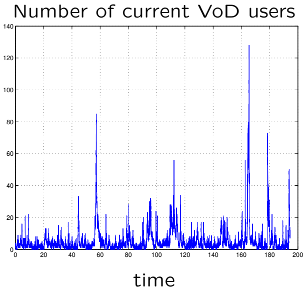
- A sensible (epidemic) model to catch the burstiness and the dynamics of the workload
- A (Markov) model that verifies a large deviation principle
- A probabilistic management policy based on the large deviation characterisation

An epidemic based model for volatile workload

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 deviation principle
- A probabilistic management policy based on the large deviation characterisation



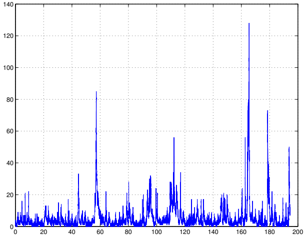
An epidemic based model for volatile workload

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

Approach – Combine the three ingredients:

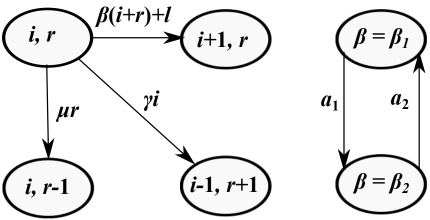
- A sensible (epidemic) model to catch the burstiness and the dynamics of the workload
- A (Markov) model that verifies a large deviation principle
- A probabilistic management policy based on the large deviation characterisation

Number of current VoD users



time

A hidden state Markov process with memory effect

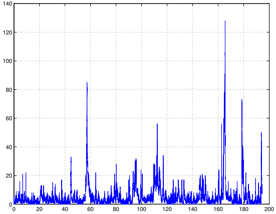


i : current # of viewers / r : current # of infected

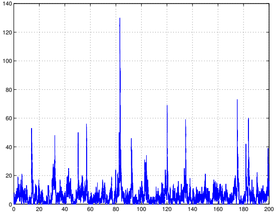
An epidemic based model for volatile workload

Calibration and evaluation

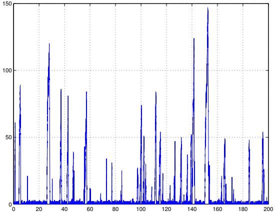
VoD workload trace



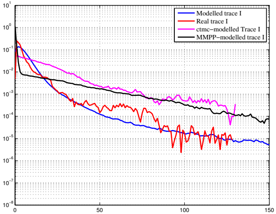
Memory Markov model



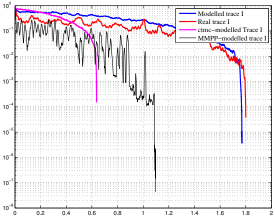
Modul. Markov Poisson



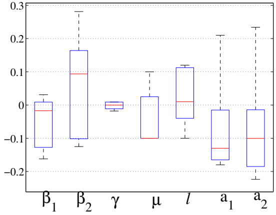
Steady state distribution



Autocorrelation function



Param. estimation precision



Markov processes

Under mild conditions, a Markov processes I_t verifies a large deviation principle:

$$\mathbb{P}\{\langle I_t \rangle_\tau \approx \alpha\} \equiv \exp\left(\tau \cdot f(\alpha)\right), \quad \tau \to \infty$$

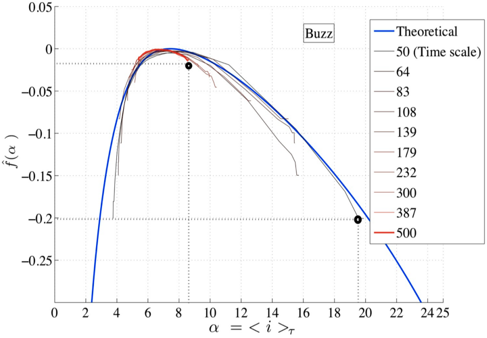
$f(\alpha)$: theoretically (from the transition matrix) or empirically (from a finite trace) identifiable

Markov processes

Under mild conditions, a Markov processes I_t verifies a large deviation principle:

$$\mathbb{P}\{\langle I_t \rangle_\tau \approx \alpha\} \equiv \exp\left(\tau \cdot f(\alpha)\right), \quad \tau \to \infty$$

$f(\alpha)$: theoretically (from the transition matrix) or empirically (from a finite trace) identifiable

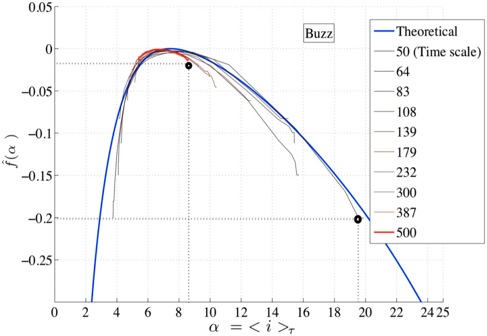


Markov processes

Under mild conditions, a Markov processes I_t verifies a large deviation principle:

$$\mathbb{P}\{\langle I_t \rangle_\tau \approx \alpha\} \equiv \exp\left(\tau \cdot f(\alpha)\right), \quad \tau \to \infty$$

$f(\alpha)$: theoretically (from the transition matrix) or empirically (from a finite trace) identifiable



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

Parametric generalisation of semi-supervised learning

Standard classification

Training set

$(X^{(t)}, Y^{(t)}) \longrightarrow \textit{classifier}$

Validation set

$X^{(v)} \xrightarrow{\textit{classifier}} Y^{(v')} \ : \ |Y^{(v)} - Y^{(v')}| \simeq 0$

Real data

$X \xrightarrow{\textit{classifier}} \textit{Answer}$

Semi-supervised classification

Validation set

$(X^{(v)}, Y^{(L)}) \xrightarrow{\textit{classifier}} Y^{(v')}$

$\textit{such that } |Y^{(v)} - Y^{(v')}| \simeq 0$

Real data

$(X, Y^{(L)}) \xrightarrow{\textit{classifier}} \textit{Answer}$

Allow to constantly update the classifier to match data evolution

Parametric generalisation of semi-supervised learning

Standard classification

Training set

$(X^{(t)}, Y^{(t)}) \longrightarrow$ *classifier*

Validation set

$X^{(v)} \overset{classifier}{\longrightarrow} Y^{(v')} \;:\; |Y^{(v)} - Y^{(v')}| \simeq 0$

Real data

$X \overset{classifier}{\longrightarrow} Answer$

Semi-supervised classification

Validation set

$(X^{(v)}, Y^{(L)}) \overset{classifier}{\longrightarrow} Y^{(v')}$

such that $|Y^{(v)} - Y^{(v')}| \simeq 0$

Real data

$(X, Y^{(L)}) \overset{classifier}{\longrightarrow} Answer$

Allow to constantly update the classifier to match data evolution

Dataset $\mathbf{X} = \underbrace{X_1, X_2, \ldots, X_p}_{\text{labeled points}}, X_{p+1}, \ldots, X_N$

Similarity matrix W and D (reap. D^*) the row-sum (reap. column)

Label matrix $Y = \{Y_{i,k} \in (0,1)$ for $i = 1, \ldots N$ and $k = 1, \ldots K\}$

Objective (classification) matrix $F_{N\times K}$: element i belongs to class $k^* = \underset{k}{\operatorname{argmax}} F_{i,k}$

Parametric generalisation of semi-supervised learning

Standard Laplacian solution

$$\underset{F}{\operatorname{argmax}} \left\{ \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} \parallel F_{i.} - F_{j.} \parallel^2 + \mu \sum_{i=1}^{N} d_i \parallel F_{i.} - Y_{i.} \parallel^2 \right\}$$

Parametric generalisation of semi-supervised learning

Standard Laplacian solution

$$\underset{F}{\text{argmax}} \left\{ \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} \parallel F_{i.} - F_{j.} \parallel^2 + \mu \sum_{i=1}^{N} d_i \parallel F_{i.} - Y_{i.} \parallel^2 \right\}$$

Generalised semi-supervised classification [M. Sokol, 2012]

$$\underset{F}{\text{argmax}} \left\{ \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} \parallel d_i^{\sigma-1} F_{i.} - d_j^{\sigma-1} F_{j.} \parallel^2 + \mu \sum_{i=1}^{N} d_i^{2\sigma-1} \parallel F_{i.} - Y_{i.} \parallel^2 \right\}$$

Parametric generalisation of semi-supervised learning

Standard Laplacian solution

$$\underset{F}{\text{argmax}} \left\{ \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} \parallel F_{i.} - F_{j.} \parallel^2 + \mu \sum_{i=1}^{N} d_i \parallel F_{i.} - Y_{i.} \parallel^2 \right\}$$

Generalised semi-supervised classification [M. Sokol, 2012]

$$\underset{F}{\text{argmax}} \left\{ \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} \parallel d_i^{\sigma-1} F_{i.} - d_j^{\sigma-1} F_{j.} \parallel^2 + \mu \sum_{i=1}^{N} d_i^{2\sigma-1} \parallel F_{i.} - Y_{i.} \parallel^2 \right\}$$

$\sigma = 1$	Standard Laplacian	*(Random walk from unlabelled to labelled points)*
$\sigma = 1/2$	Normalised Laplacian	
$\sigma = 0$	PageRank method	*(Random walk from labelled to unlabelled points)*

Parametric generalisation of semi-supervised learning

Standard Laplacian solution

$$\underset{F}{\text{argmax}} \left\{ \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} \parallel F_{i.} - F_{j.} \parallel^2 + \mu \sum_{i=1}^{N} d_i \parallel F_{i.} - Y_{i.} \parallel^2 \right\}$$

Generalised semi-supervised classification [M. Sokol, 2012]

$$\underset{F}{\text{argmax}} \left\{ \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} \parallel d_i^{\sigma-1} F_{i.} - d_j^{\sigma-1} F_{j.} \parallel^2 + \mu \sum_{i=1}^{N} d_i^{2\sigma-1} \parallel F_{i.} - Y_{i.} \parallel^2 \right\}$$

$\sigma = 1$	Standard Laplacian	*(Random walk from unlabelled to labelled points)*
$\sigma = 1/2$	Normalised Laplacian	
$\sigma = 0$	PageRank method	*(Random walk from labelled to unlabelled points)*

$$F_{.k} = \frac{\mu}{2+\mu} \left(I - \frac{2}{2+\mu} D^{-\sigma} W D^{\sigma-1} \right)^{-1} Y_{.k}, \text{ for } k = 1, \ldots, K$$

Tune the value of parameter σ to match the dataset

Duality and semi-supervised learning

graph (similarity) $\overset{ordination}{\longleftrightarrow}$ process (metric)

formulation (*multidimensional scaling*) : $\underset{F}{\operatorname{argmax}} \left\{ \sum_{i=1}^{N} \sum_{j=1}^{N} (\| F_{i.} - F_{j.} \| - w_{ij})^2 \right\}$

Duality and semi-supervised learning

$$\text{graph} \quad \text{(similarity)} \quad \overset{\textit{ordination}}{\longleftrightarrow} \quad \text{process} \quad \text{(metric)}$$

$$\text{formulation } (\textit{multidimensional scaling}) : \operatorname*{argmax}_F \left\{ \sum_{i=1}^{N} \sum_{j=1}^{N} (\| F_{i.} - F_{j.} \| - w_{ij})^2 \right\}$$

- bridge ordination (MDS) and generalised semi-supervised learning
  - ▷ leverage σ flexibility to vary duality principle

Duality and semi-supervised learning

$$\text{graph} \quad (\text{similarity}) \quad \overset{\text{ordination}}{\longleftrightarrow} \quad \text{process} \quad (\text{metric})$$

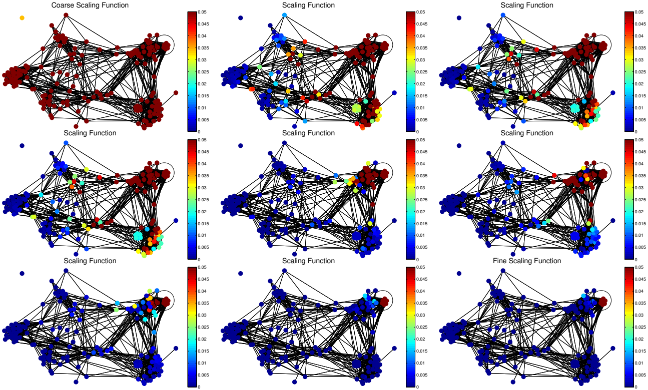
$$\text{formulation } (\textit{multidimensional scaling}) : \underset{F}{\text{argmax}} \left\{ \sum_{i=1}^{N} \sum_{j=1}^{N} (\| F_{i.} - F_{j.} \| - w_{ij})^2 \right\}$$

- bridge ordination (MDS) and generalised semi-supervised learning
  - ▷ leverage σ flexibility to vary duality principle
- data adaptivity of semi-supervised learning
  - ▷ use to update dynamic graph $\leftrightarrow$ non-stationary time series

Graph diffusion

Epidemic diffusion (MOSAR): Apply standard tools. . .

▷ Relationship between virus spreading and graph structure: Can diffusion wavelets help?



▷ How to take into account / reflect dynamicity of graphs

Context and collaborations

- Dante (B. Girault, E. Fleury,. . .)

Context and collaborations

- Dante (B. Girault, E. Fleury,. . .)
- Institut des Systèmes Complexes

Context and collaborations

- Dante (B. Girault, E. Fleury,. . .)
- Institut des Systèmes Complexes
- Sisyphe (ENS Lyon, P. Borgnat)

Context and collaborations

- Dante (B. Girault, E. Fleury,. . .)
- Institut des Systèmes Complexes
- Sisyphe (ENS Lyon, P. Borgnat)
- Other teams (e.g. Geodyn, Maestro. . .)

Context and collaborations

- Dante (B. Girault, E. Fleury,. . .)
- Institut des Systèmes Complexes
- Sisyphe (ENS Lyon, P. Borgnat)
- Other teams (e.g. Geodyn, Maestro. . .)
- International cooperations (e.g. EPFL)
- . . .