Research team RESO

Optimized Protocols and Softwares for High-Performance Networks

Networks and Telecommunications

Evaluation Seminar – Rungis, march 21–23, 2012

Facts and Numbers

- Created in 2003 by Pascale Vicat-Blanc (DR Inria)
- Team-project from Inria Grenoble Rhône-Alpes (ENS Lyon, LIP)
- Scientific lead: P. Vicat-Blanc $<$ 2010 $<$ Paulo Gonçalves (CR Inria)

		2008	in/out	2012
Permanent	Researchers (Inria)	3	0/-1	2
	Professors (U. Lyon 1)	3	+1/0	4
	Engineers	1,5	0/-0,5	1
Temporary	PhD Students	8	+10/-11	7
	Post-docs	1	+9/-7	3
	Engineers	4	+12/-13	3
		20,5		**20**

- 232 publications (46 journals or chapters)
- 5 best papers, demonstration or project awards
- 5 patents: creation of Inria spinoff Lyatiss (awarded Oseo emergence 2009 & French Tech Tour in Silicon Valley prizes)
- 24 projects (4 Inria, 9 national, 9 european, 2 ass. teams, 4 ind.)
  Inria / Alcatel Lucent Bell Labs Common Lab

Scientific context

Resource management and communication protocols in high speed networks (e.g. grids), virtualized infrastructures (e.g. clouds) and wireless networks.

Motivated by high demanding applications and the Future Internet

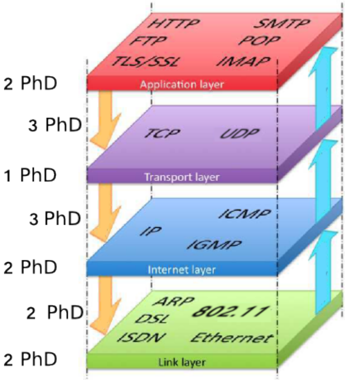
Network virtualization – Integrate the network into the cloud landscape. Develop a **unified framework** to address network resource virtualization

Adaptive resource management – Self-organized and **autonomic networks** to face scalability issues. Rationalize resources usage to cope with workload volatility and energy consumption efficiency

Semantic networking – Investigate the notion of **Knowledge Plane** to steer traffic aware networking policies

Network metrology – Measure, analyze, model **traffic dynamics** and **energy consumption patterns**

Scientific strengths



Main expertise

- Virtualization
- Protocol design
- Software development
- Large scale testbeds and experiments
- Performance evaluation

Additional competence

- Multi-scale signal processing
- Stochastic processes
  (modeling, statistical estimation)

Methodology

Experimentation & Simulation
$\Downarrow$ Theory $\Uparrow$
Software solutions

Scientific achievements

- Network virtualization
  - Virtual Infrastructure Description Language (*VXDL*)
  - Allocating resources to Virtual Infrastructures (*VxAlloc*)
  - Virtualized switching fabric (*VxSwitch*)
  - Virtual network service (*HiperNet*)
- Adaptive resource management
  - **Autonomic configuration fight ossification of Internet infrastructures and allow human free network management.** Interoperable and robust solutions for router assisted based protocols.
  - Dynamic provisioning of network infrastructure & bandwidth scheduling service (*BDTS & FLOC*)
  - Wireless networks: fair bandwidth sharing (*Profiterole*) + available bandwidth estimation (*RABE*)

Scientific achievements ⇒ successful outcome

- Network virtualization
  - Virtual Infrastructure Description Language (*VXDL*)
  - Allocating resources to Virtual Infrastructures (*VxAlloc*)
  - Virtualized switching fabric (*VxSwitch*)
  - Virtual network service (*HiperNet*)
- Adaptive resource management
  - **Autonomic configuration fight ossification of Internet infrastructures and allow human free network management.** Interoperable and robust solutions for router assisted based protocols.
  - Dynamic provisioning of network infrastructure & bandwidth scheduling service (*BDTS & FLOC*)
  - Wireless networks: fair bandwidth sharing (*Profiterole*) + available bandwidth estimation (*RABE*)

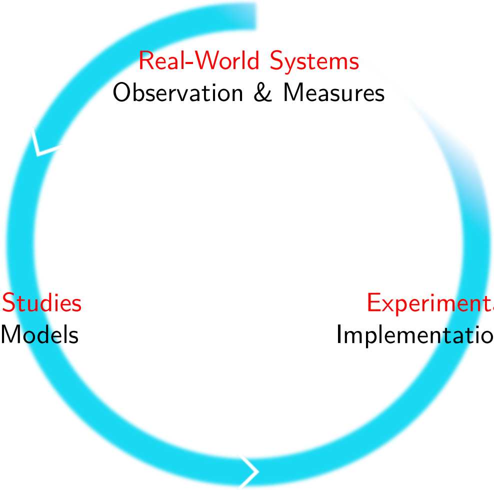
Scientific achievements (cont'd)

- Semantic networking
  - Size-based flow scheduling architecture
  - Advocate use of *jumbo frames* for large flow transmission protocols
  - **Traffic-aware flow admission based on a dynamic knowledge plane (link modeled as a on-line calibrated *M/G/1* queue)**
  - Assessment framework relying on the user-satisfaction vis-à-vis the bandwidth allocation fairness for long-lived service sessions
- Network metrology
  - **Design of energy aware and energy efficient reservation infrastructures frameworks**
  - Design of green exascale services
  - Numerical solutions to queueing systems
  - Semi-supervised machine learning for traffic classification
  - Statistical characterization of traffic and QoS impact

Scientific achievements (cont'd)

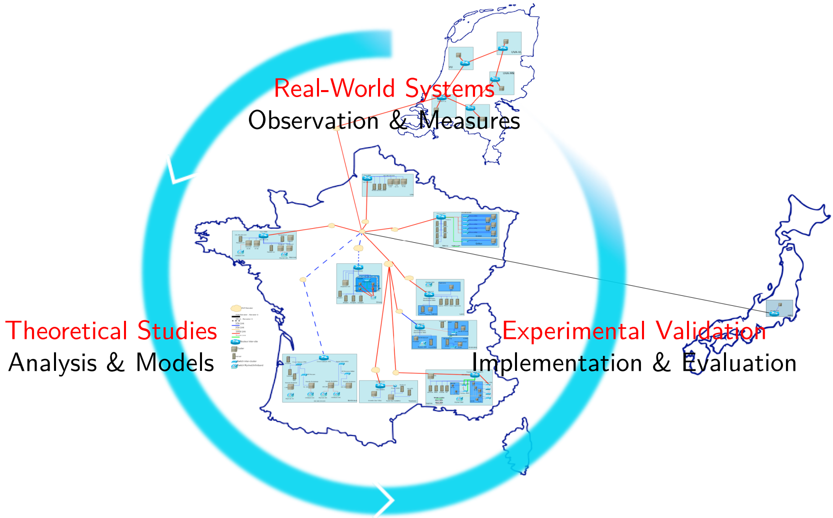
- Semantic networking
  - Size-based flow scheduling architecture
  - Advocate use of *jumbo frames* for large flow transmission protocols
  - **Traffic-aware flow admission based on a dynamic knowledge plane (link modeled as a on-line calibrated *M/G/1* queue)**
  - Assessment framework relying on the user-satisfaction vis-à-vis the bandwidth allocation fairness for long-lived service sessions
- Network metrology
  - **Design of energy aware and energy efficient reservation infrastructures frameworks**
  - Design of green exascale services
  - Numerical solutions to queueing systems
  - Semi-supervised machine learning for traffic classification
  - Statistical characterization of traffic and QoS impact

Network metrology



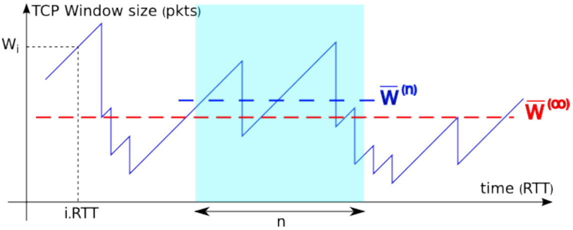
Sensible and *successful* approach to deal with systems' complexity

Network metrology



GRID'5000: Large scale experimental testbed (MetroFlux + ShowWATTS)

Focus: Mean throughput of a long lived TCP flow



Mean throughput at scale n (RTT):

$$\overline{W}^{(n)} = \frac{1}{n}\sum_{i=1}^{n} W_i$$

In stationary regime (AIMD)

$(W_i)_{i\geq 1}$ commonly modeled as a Markov chain : $\begin{cases} Q_{w,\min(w+1,w_{\max})} = 1 - p(w) \\ Q_{w,\max(\lfloor w/2 \rfloor,1)} = p(w) \end{cases}$

For almost all realizations (Ergodic Birkhoff Theorem, 1931)

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

Padhye relation (98) – Assuming Bernoulli loss probability $p(w) = p_0(1-p_0)^w$

$$\overline{W}^{(\infty)} \underset{p_0\to 0}{\sim} \sqrt{\frac{3}{2p_0}}$$

Characterize the throughput variability: How often $\overline{W}^{(n)} \simeq \alpha \neq \overline{W}^{(\infty)}$?

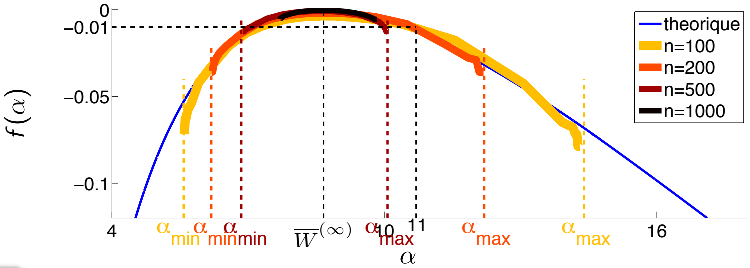
A large deviation principle holds true for Markov processes (Ellis, 84)

$$\mathbb{P}(\overline{W}^{(n)} \simeq \alpha) \underset{n\to\infty}{\sim} \exp(n \cdot f(\alpha)) \quad \textit{with } f(\alpha)\textit{: scale invariant spectrum}$$

Ergodic large deviation theorem (Loiseau et al., 2009)

For almost all realizations, if the number k_n of intervals of size n is large enough, *then almost surely*

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



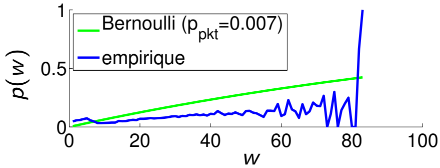
In practice...

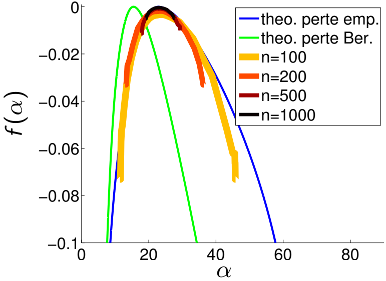
Parametric estimate (theoretical) –

(known or estimated loss) $p(w)$ $\rightarrow$ Q (transition matrix) $\rightarrow$ $f(\alpha)$

Empirical estimate (sample) –

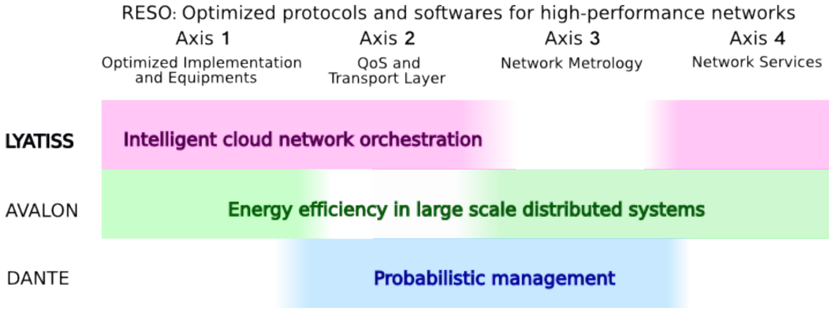
(observations) $(W_i)_{i \leq N}$ $\rightarrow$ $f_n(\alpha)$





Ongoing work – Similar LDP applied to probabilistic resource allocation to face volatile workload environments in cloud networking

RESO – Follow-up



Lyatiss – Inria spinoff started in 2010: P. Vicat-Blanc and S. Soudan

Avalon – Inria team-project (expected in *Distributed calculs and high-performance applications* theme) leaded by *C. Perez*: L. Lefèvre, J.-P. Gelas and O. Glück

Dante – Inria team-project (expected in *Networks and telecommunications* theme) leaded by *E. Fleury* (D-Net): P. Gonçalves, T. Begin and I. Guérin-Lassous

Avalon Algorithms and Software Architectures for Service Oriented Platforms

Transverse axis: Energy efficiency in large scale distributed systems

Challenge

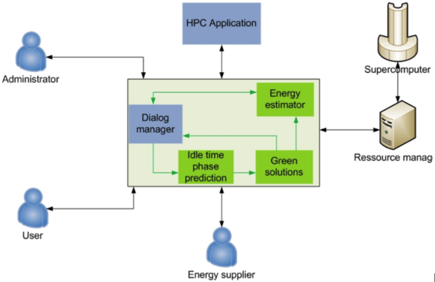
- Energy usage and efficiency is not taken into account since the design of systems
- Energy is becoming the major constraint for large scale systems (i.e. exascale platforms)
- How to express QoS w.r.t performance and energy efficiency?

Methodology

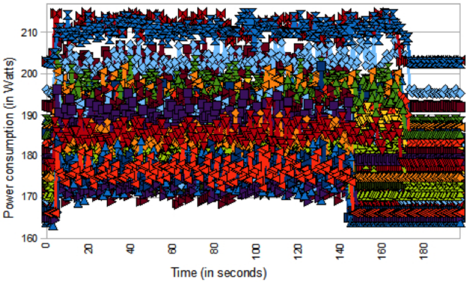
- Profile and predict energy usage of applications
- Investigate energy usage of (yet to come) virtualized / high density and high performance infrastructures : Clouds, exascale, next generation networks
- Promote and inject energy usage in measurement decision systems
- Apply adapted green leverages at multiple layers (network, cpu)

Energy efficiency in exascale infrastructures (scientific challenge)

All actors have to cooperate!



Energy estimator



Power consumption of 10 Gbytes message logging on HDD for 62 nodes of Grid5000 Lyon site

Prediction : profiling energy usage of applications and services

Green solutions: Shutdown / Slowdown / Optimizing / Coordinating

Dante Dynamic Networks: Temporal and Structural Capture Approach

Context – Inter-related and interacting entities in different networks

(A dynamic network)

- technological: phone (churn), computers, routers (metrology), blogs, web
- scientometrics: articles, bibliography coupling, profiling
- human: social netwoks, contacts, epidemiology. . .

Objectives – Propose a unified methodology to formalize the properties of the time and space dynamics of interacting entities

Approach – Combine graph theory and signal processing

Analogy – Switching from an image to a video sequence: requires specific tools to leverage time-space coherence ($\neq$ ensemble of independent pixel time series)

Graph-Oriented Signal Processing (scientific challenge)

Longitudinal evolution of graph features

- Duality principle: **network** $\leftrightarrow$ **time series** mapping (Campanharo et al., 2011)
- **Non stationary** statistical tests via surrogate data (Borgnat et al., 2010)
- **Relative stationarity** & explicit observation scale

Harmonic analysis on graphs

- Select graph features with an **adapted level of granularity** in space
- **Graph laplacians** and spectral interpretation of wavelets in a multiresolution setting (Hammond et al. 2011)
- **Diffusion maps** and diffusion wavelets (Coifman et al., 2005)

$\Rightarrow$ Harmonic analysis of dynamic graphs in the spirit of a time-frequency representation

External support

Inria programs Green-Net (08-09) – Mission (10-11) – Aladdin (07-12) – Netools (10-12)

National initiatives Igtmd (06-08) – DslLab (06-08) – HipCal (07-10) – Dmasc (08-12) – Petaflow (09-12) – Rescue (10-13) – Compatible One (10-12) – Carriocas (06-09) – Interfaces (10-11)

European projects Sail (09-12) – Geysers (09-12) – Autonomic Internet (08-10) – Aeolus (05-10) – EC-Gin (06-09) – OGF Europe (08-10) – Cost (09-13) – PrimeeEnergyIT (10-12) – EuroNF (08-10)

Associated teams Grid-Net (07-09) – Negst (06-09)

Industrial contracts **Common Lab Inria Alcatel-Lucent Bell Labs** (08-12) – France Telecom R&D (05-08, 09-12) – Anagram (08)

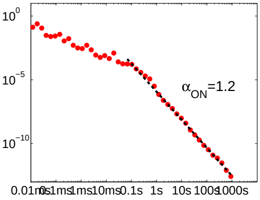
Focus: Aggregated traffic properties

ON/OFF model and LRD properties (Taqqu et al. 1997)

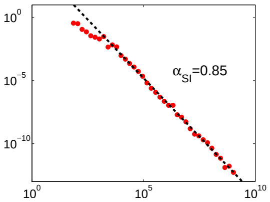
- *$N_{src} \to \infty$ i.i.d. ON/OFF sources with heavy tailed exponent α_{ON}*
- *Flow throughput constant and the same for all flows*
- *Aggregate traffic bandwidth $B^{(\Delta)}(t)$ at scale Δ is LRD*

$$\mathbb{C}ov_{B^{(\Delta)}}(\tau) \underset{\tau\to\infty}{\sim} \tau^{2H-2} \quad \textit{with} \quad H = \max\left(\frac{3-\alpha_{ON}}{2}, \frac{1}{2}\right)$$

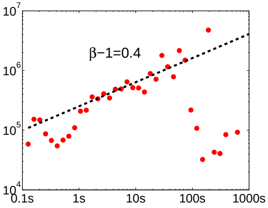
ON distribution



Size distribution

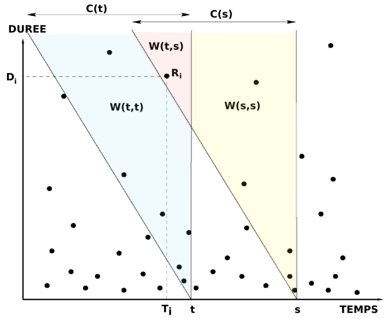


$\mathbb{E}\{\text{thr.}|\text{dur.}\}$



Correlated flow throughput and duration: $\mathbb{E}\{\text{thr.}|\text{dur.}\} \propto (\text{dur.})^{\beta-1}, \ \beta = \frac{\alpha_{ON}}{\alpha_{SI}}$

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



Flow model:

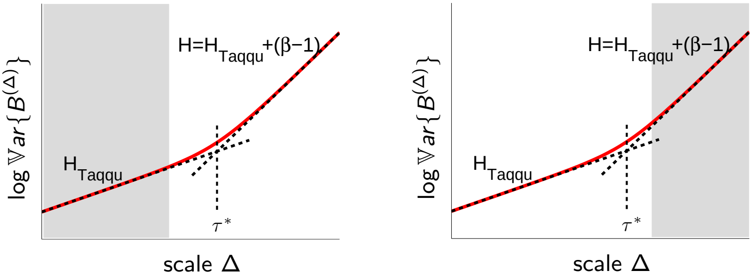
- T_i Arrival time
- D_i Duration
- R_i Throughput (reward)
- Q_i Size $(= R_i \cdot D_i)$

$$\mathbb{E}\{R|D\} = M \cdot D^{\beta-1} \quad \& \quad \mathbb{V}ar\{R|D\} = V$$

Proposition (Loiseau et al., 2009)

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

Log-diagram, $\beta > 1 \Rightarrow \tau^* = \left(\frac{C'V}{CM^2}\right)^{1/(2(\beta-1))}$



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

- Correlations intensify LRD $(\beta > 1)$
- Traffic evolution, future Internet: “flow-aware” control mechanisms
- **Series of experiments to evaluate the impact (not necessarily bad) of LRD on QoS. . .**

Heavy tail index impact on QoS