Non Gaussian Long Memory Internet Traffic Statistical Modeling Application to Anomaly Detection.

P. Abry [1], P. Borgnat [1], P. Owezarski [2],
METROlogy for SECurity Research Project [3]

[1]Physics Lab, CNRS, ENS Lyon, France,
[2]LAAS, CNRS, Toulouse, France,
[3]`www.metrosec.fr`

Intimate 2006, July 6th-7th, Paris

ECOLE NORMALE SUPERIEURE DE LYON

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Motivation and Goals: General Framework

Statistical Modeling

- Non Gaussian
- Short vs Long Range Dependence
- $\Downarrow$ Detection

Anomaly Detection

- Detection Proc.
- Perf. Evaluation
- Need for a Database
- Classification

Regular Data

- Major Trace Repositories
- Self Collected
- a large variety of Traffic !

Data with Anomalies

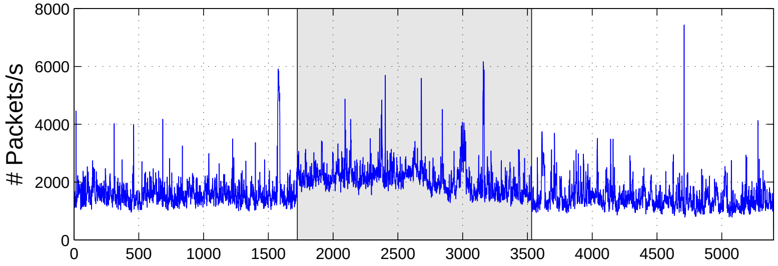
- Documented Anomalies
- Reproducible, Controlled
- DDoS Attacks, Flash Crowds
- Low Level Intensities
- Real Network, Real Traffic

Outline

Outline

Outline

Aggregated Time Series



- Aggregation level : Δ,
- Packet Count,
- Byte Count,

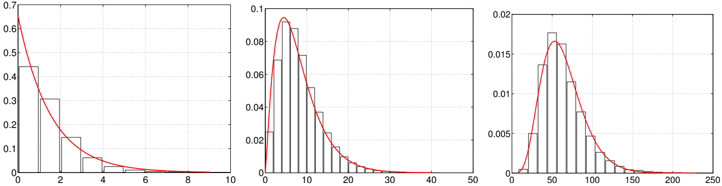
Intuitions and Issues

1. How should we choose a model ?
   - Based on significant data characteristics,
   - Parsimony,
   - Detection Goal in mind: parameters suited for detection.
2. What should we model ?
   - Difficult: The full statistics (high order statistics) ?
   - Simple: Marginal Distributions ? Covariances ?
3. What Aggregation level should we choose ?
   - Small ? Large ? Compared to which scale ?
   - Depends on data ? on goals ?
4. Proposed Solutions
   1. $\Rightarrow$ Long Range vs Short range dependencies ?
      Gaussian vs non Gaussian ?
   2. $\Rightarrow$ Trade-off: Marginals (1st stat order) and
      covariances (2nd stat order) **jointly**
   3. $\Rightarrow$ Modeling covariant with a change of aggregation level ?

Outline

Marginals

- **Empirical PDFs** LBL-TCP-3

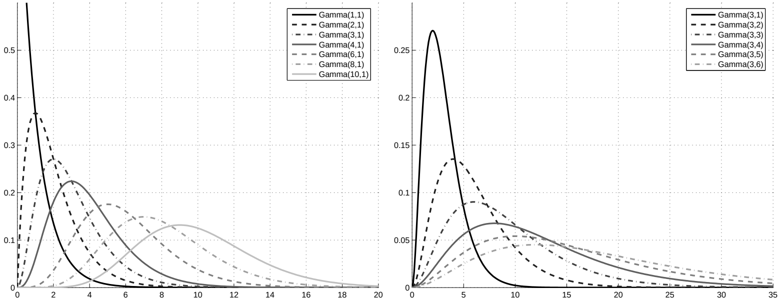


$\Delta = 4$ms $\Delta = 32$ms $\Delta = 256$ms

- Poisson ? Exponential ? Gaussian ?
- Aggregation level ?

Gamma Distributions

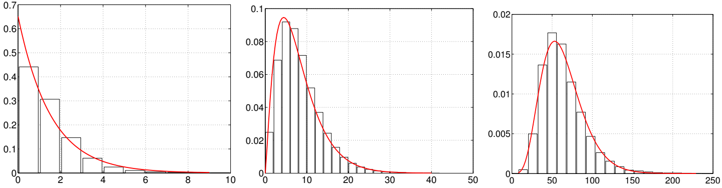
$$\Gamma_{\alpha,\beta}(x) = \frac{1}{\beta\Gamma(\alpha)}\left(\frac{x}{\beta}\right)^{\alpha-1}\exp\left(-\frac{x}{\beta}\right).$$



- Shape parameter α : From Gaussian to exponential, $1/\alpha \simeq$ distance from Gaussian,
- Scale parameter β : Multiplicative factor.

Gamma Fits

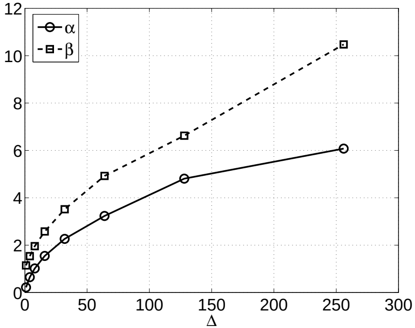
- **Empirical PDFs and Gamma Fits** LBL-TCP-3



$\Delta = 4$ms $\Delta = 32$ms $\Delta = 256$ms

- Accurately Fits data for all aggregation levels Δ,
- Stability under addition :
  $X_1 : \Gamma_{\alpha_1,\beta}, X_2 : \Gamma_{\alpha_2,\beta}, (X_1, X_2)$ Indep. $\Longrightarrow X_1 + X_2 : \Gamma_{\alpha_1+\alpha_2,\beta}$,
- Aggregation : $X_{2\Delta}(k) = X_\Delta(k) + X_\Delta(k+1)$.

Parameter Estimation: $\hat{\alpha}_\Delta, \hat{\beta}_\Delta$



- Stability under addition and Independence

$$\Rightarrow \begin{cases} \alpha(\Delta) = \alpha_0 \Delta \\ \beta(\Delta) = \beta_0 \end{cases}$$

- $\hat{\alpha}_\Delta, \hat{\beta}_\Delta$ accommodate correlations !

Outline

Covariance : the wavelet point of view

- X_Δ stationary stochastic process, with spectrum $f_{X_\Delta}(\nu)$,
- Wavelet Coefficients: $d_X(j,k)$,

WaveletTransform

- Wavelet Spectrum: $S(j) = \frac{1}{n_j}\sum_{k=1}^{n_j} |d_{X_\Delta}(j,k)|^2$,

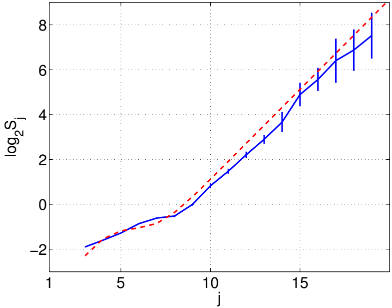
$$\mathbb{E}S(j) = \int f_X(\nu) 2^j |\Psi_0(2^j \nu)|^2 du \simeq \hat{f}_X(\nu = 2^{-j}\nu_0).$$

- Log-scale Diagram: $\log_2 S_2(j)$ vs. $\log_2 2^j = j$.

Both Short and Long Range Dependencies

- **Log-scale Diagram**: $\log_2 S_2(j)$ vs. $\log_2 2^j = j$.

X_Δ, LBL-TCP-3, Δ =1ms



- Power law at coarse scales (low frequencies):
  $\Rightarrow$ Long range dependence,
- Short dependence at fine scales (low frequencies),
- $\Rightarrow$ Use a FARIMA(P, d, Q) covariance form.

FARIMA(P, d, Q) covariance

farima = fractionally Intregrated ARMA.

1. fractional integration with parameter d,
2. short-range correlations as an ARMA(1,1) $\rightarrow$ params. θ, ϕ.

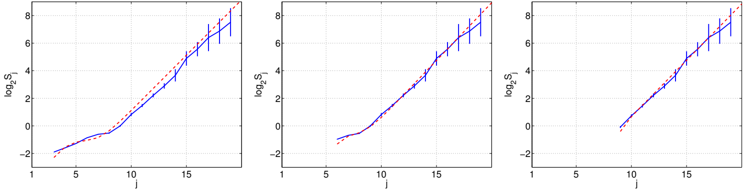
$$f_{X_\Delta}(\nu) = \sigma_\epsilon^2 \left|1 - e^{-i2\pi\nu}\right|^{-2d} \frac{|1 - \theta e^{-i2\pi\nu}|^2}{|1 - \phi e^{-i2\pi\nu}|^2},$$

- d controls Long Range Dep., with $\gamma = 2d$,

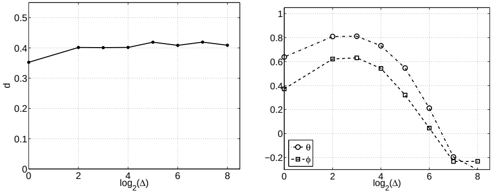
LRD

- P, Q control Short Range Dep.

Empirical LDs and FARIMA(P,d,Q) Fits LBL-TCP-3



$\Delta = 4$ms $\Delta = 32$ms $\Delta = 256$ms



- Accurately Fits data for all aggregation levels Δ,
- LRD is persistent, SRD are cancelled out.

Outline

Non Gaussian Long Range Dependent Models

- Jointly 1st and 2nd order statistics,
- Parsimony,
- Covariance with respect to the Aggregation level Δ,

ShowResults

- For various data, various traffics, various links, various networks,

TableData

- Suboptimal but robust and low cost parameter estimation procedures,
- Numerical synthesis procedures (with A. Scherrer, LIP6, ENS Lyon),

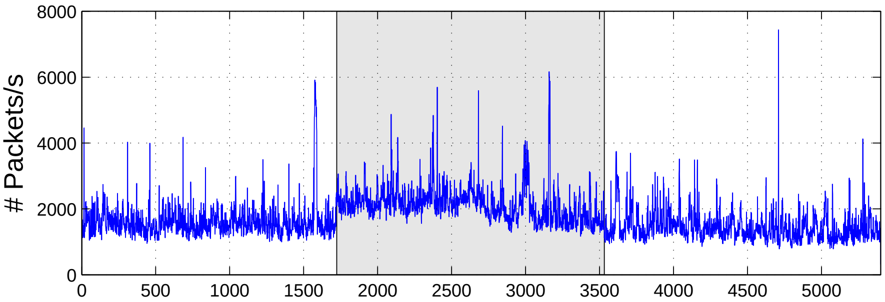
NumericalSynthesis

- Detection ?

Outline

Outline

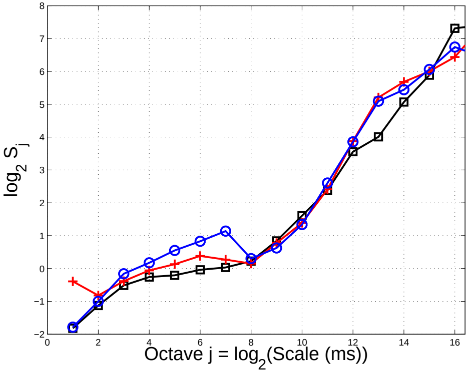
Aggregated Time Series



- IPerf, UDP Flooding.

DDoS Attack (UDP Flooding)

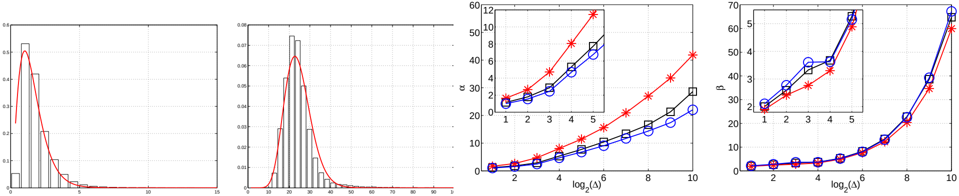
LogScale Diagrams



- Black: before, Red: during, blue: After Attack.
- LRD not caused by nor altered by DDoS Attacks.

DDoS Attack (UDP Flooding)

Gamma Fits (during attack)



- Black: before, Red: during, blue: After Attack,
- Model fits data with anomaly equally satisfactorily,
- Goes faster to Gaussian → DDoS changes the SRD,
- Multiresolution nature (multi Δ) of the model.

Principles

- Choose a Reference time windows,
- Split data into sliding time windows of length T,
- For each time window I:
  - Aggregate data at levels $\Delta = 2^j$, $j = 1, \ldots, J$
  - Estimate the chosen statistics: $\hat{\alpha}_\Delta(I)$, $\hat{\beta}_\Delta(I)$
  - Compute a distance between I and R

$$D_\alpha(I) = \frac{1}{J}\sum_{j=1}^{J}\left(\hat{\alpha}_{2^j}(I) - \hat{\alpha}_{2^j}(ref)\right)^2, \quad (1)$$

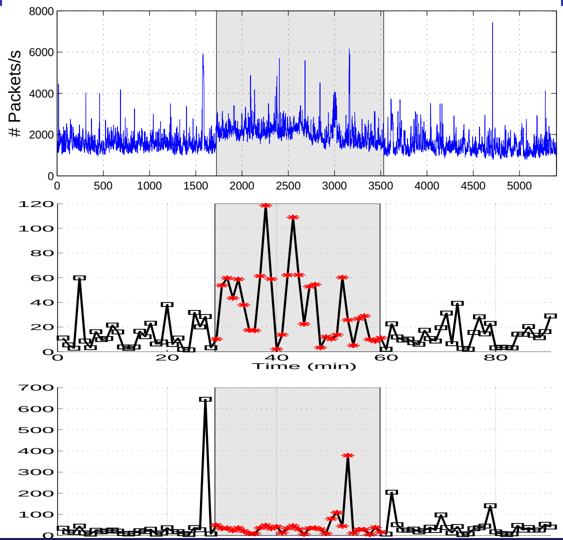
$$D_\beta(I) = \frac{1}{J}\sum_{j=1}^{J}\left(\hat{\beta}_{2^j}(I) - \hat{\beta}_{2^j}(ref)\right)^2. \quad (2)$$

  - Choose a threshold λ to decide when the distance is *too large*, $D_\alpha(I) \geq \lambda$.

Example 1 : DDoS Attack

$D_\alpha(l)$

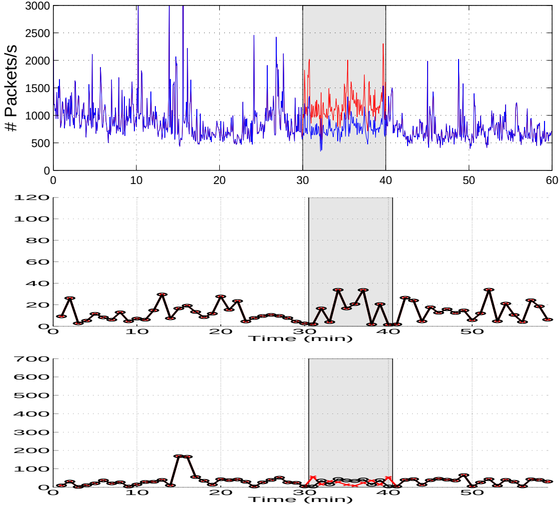
$D_\beta(l)$



Example 2 : Artificial Multiplicative Traffic Increase

$D_\alpha(l)$

$D_\beta(l)$

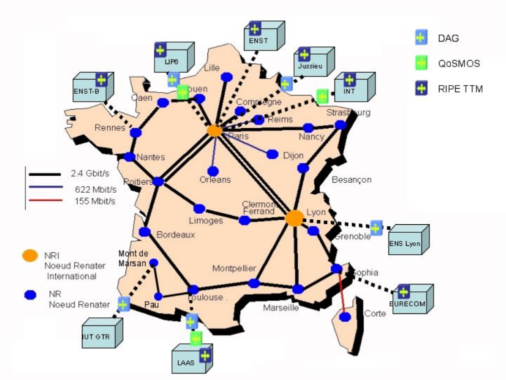


Statistical performance ?

- Receiver Operating Curves:
  How many false positive given the false negative ?
- $P_D = f(P_F)$ or $P_D = f(\lambda), P_F = f(\lambda)$,
- $\Rightarrow$ Need for a documented anomaly dataBase !!!

Outline

Anomaly DataBase: Topology



- METROSEC partners (all over France),
- Lyon, Nice, Paris, Mont-de-Marsan, Coimbra $\Rightarrow$ Toulouse,
- RENATER network,

Anomaly DataBase: Typology

- UDP Flooding, IPperf, Trinoo.
- Increase Link/Routeur load $\Rightarrow$ Decrease QoS.
- Emulate a small Leave of a huge Botnet Tree:

  Moraly close to the source,
  Low Intensity Level Attack,
  Before Effective Impact on QoS,
  $\Rightarrow$ Difficult to detect.
- In progress: TFN2K : SYN, ICMP flooding, Smurf, Targa3.

Anomaly DataBase - DDoS Attacks -2004 - 2006

Id	t_i	T(s)	t_a	T_A(s)	D	V	I (%)
DDoS performed with Iperf							
R	17:30	60000	20:00	20000	0.5	60	33.82
I	09:54	5400	10:22	1800	0.25	1500	17.06
II	14:00	5400	14:29	1800	0.5	1500	14.83
III	16:00	5400	16:29	1800	0.75	1500	21.51
IV	10:09	5400	10:16	2500	1.0	1500	33.29
V	10:00	5400	10:28	1800	1.25	1500	39.26
A	14:00	5400	14:28	1800	1	1000	34.94
B	16:00	5400	16:28	1800	1	500	40.39
C	10:03	5400	10:28	1800	1	250	36.93
DDoS performed with Trinoo							
tM	18:21	5400	18:58	601	0.1	300	4.64
tN	18:22	3600	18:51	601	0.1	300	15.18
tT	18:22	3600	18:51	601	8	300	82.85

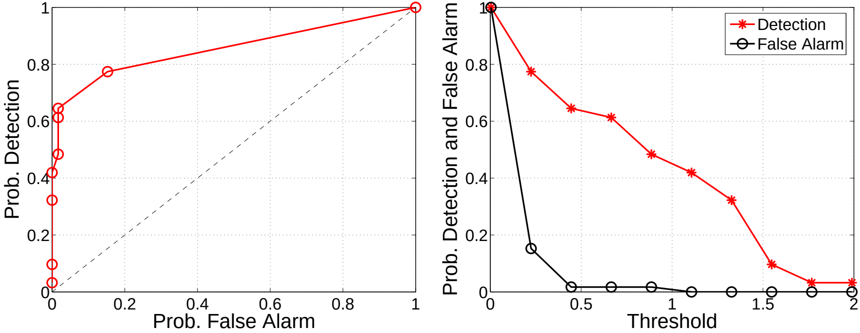
Anomaly DataBase - Flash Crowds - 2005 - 2006

Id	t_i	T(s)	t_a	T_A(s)	D	V	I (%)
FC performed by human							
FC-1	13:45	7200	14:30	1800	—	—	31.27
FC-2	15:00	7200	15:45	1800	—	—	18.35

Outline

Stat. Perf.: Receiver Operating Curves

- How many false positive given the false negative ?
- $P_D = f(P_F)$ or $P_D = f(\lambda)$, $P_F = f(\lambda)$,



Stat. Perf.: Detection Probability

Type of Anomaly	performed with	Id	Intens. (%)	P_D P_F =10%	P_F =20%
DDoS	Iperf	R	33.82	91	93
DDoS	Iperf	I	17.06	51	64
DDoS	Iperf	II	14.83	48	54
DDoS	Iperf	III	21.51	48	58
DDoS	Iperf	IV	33.29	33	50
DDoS	Iperf	V	39.26	18	40
DDoS	Iperf	A	34.94	21	50
DDoS	Iperf	B	40.39	81	87
DDoS	Iperf	C	36.93	52	58
DDoS	Trinoo	tM	4.64	27	50
DDoS	Trinoo	tN	15.18	54	54
DDoS	Trinoo	tT	82.85	82	82

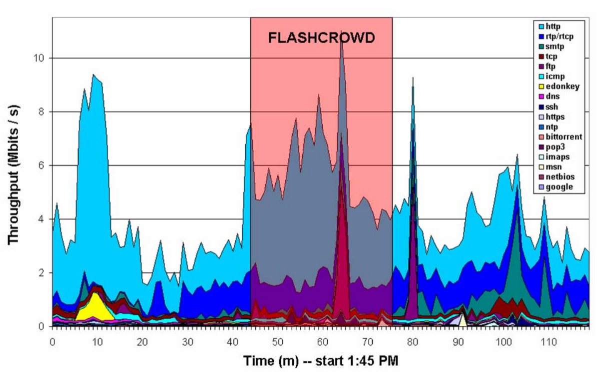
Stat. Perf.: use of other distances/other thresholds

- Kullback divergence :
$$KD(p1, p2) = \int (p_1 - p_2)(\ln p_1 - \ln p_2) dx$$
- 1D : $K_{\Delta}^{(1D)}(I) = KD(p_{\Delta,I}, p_{\Delta,Ref})$
- 2D : $K_{\Delta,\Delta'}^{(2D)}(I) = KD(p_{\Delta,\Delta',I}, p_{\Delta,\Delta',Ref})$
- Multiple consecutive threshold bypasses.

	D_α	$K_{2^4}^{(1D)}$	$K_{2^7}^{(1D)}$	$K_{2^4,2^7}^{(2D)}$
I	51 : 64	25 : 64	35 : 67	25 : 51
II	48 : 54	35 : 58	35 : 61	35 : 61
III	48 : 58	74 : 93	70 : 83	87 : 93
IV	33 : 50	56 : 67	56 : 69	34 : 66
V	18 : 40	87 : 96	34 : 93	90 : 96
A	21 : 50	50 : 78	37 : 59	53 : 81
B	81 : 87	78 : 78	09 : 33	78 : 81
C	52 : 58	91 : 91	91 : 91	91 : 91
X	93 : 96	93 : 93	93 : 93	93 : 93
tM	27 : 55	36 : 91	36 : 91	45 : 91
tN	54 : 54	73 : 91	91 : 91	55 : 73
tT	82 : 82	100 : 100	100 : 100	100 : 100

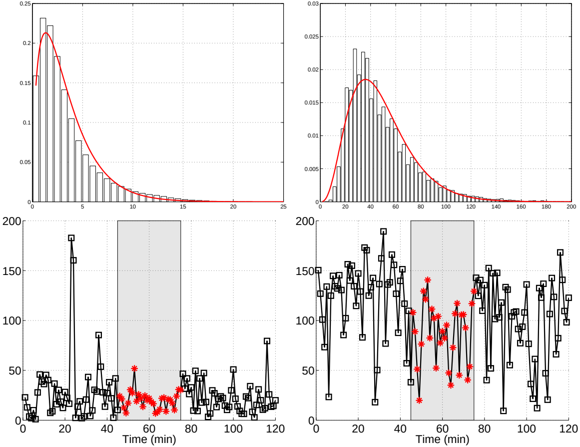
Outline

Flash Crowd



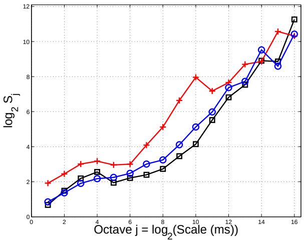
- Operated by Humans.

Flash Crowd and Gamma Fits



- Model fits data with anomaly equally satisfactorily
- But Flash Crowd does not change the SRD.

Flash Crowd and LogScale Diagrams



- LRD not caused by nor altered by Flash Crowd,
- SRD not altered by Flash Crowd,
- Medium Range Dependencies altered,
- Distances on LDs $\Rightarrow$ Detection and Classification.

Outline

Conclusions

- Conclusions:
  - Non Gaussian/ Gaussian,
  - Short vs Long range Dependence,
  - Versatile and parsimonious modeling,
  - Detection/Classification oriented,
  - Synthetic traffic generation,
  - Performance Evaluation methodology,

Perspectives

- Perspectives:
  - Comparison against other tools, IDS ?
  - A richer DataBase ?
  - QoS Impact ?
  - Adaptive Reference ?
  - Detection far from sources ?
  - Multivariate data ? Multi-Point Analysis ?
  - Robustness: Split traffic into OD pairs ?
  - Partial sampling ?
  - Joint Topology and Time Series Detection ?

Further Information

- Mails:

Patrice.Abry@ens-lyon.fr
Pierre.Borgnat@ens-lyon.fr
owe@laas.fr

- URL (reprints and preprints):

perso.ens-lyon.fr/patrice.abry
ens-lyon.fr/PHYSIQUE
www2.laas.fr/METROSEC/

Outline

Long Range Dependence

Definition of Long Range Dependence

Covariance is a non-summable power-law $\rightarrow$ spectrum $f_{X_\Delta}(\nu)$:

$$f_{X_\Delta}(\nu) \sim C|\nu|^{-\gamma}, \ |\nu| \rightarrow 0, \quad \text{with } 0 < \gamma < 1.$$

Long Range Dependence and Wavelets

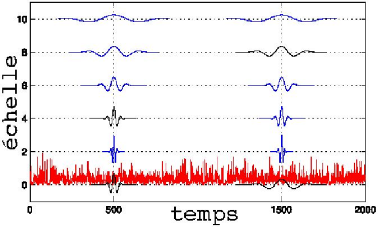
$$\mathbb{E}S(j) = \int f_X(\nu) 2^j |\Psi_0(2^j \nu)|^2 du \simeq \hat{f}_X(\nu = 2^{-j}\nu_0).$$

$$\text{LRD} \Longrightarrow \mathbb{E}S(j) \sim C2^{j(\gamma-1)}, 2^j \rightarrow +\infty.$$

Back

Wavelet Transform

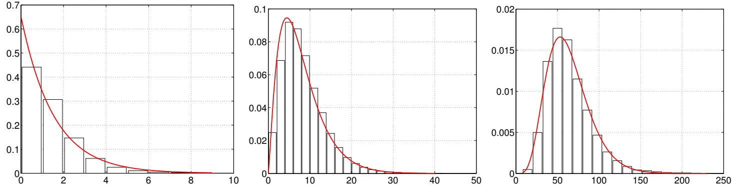
- Let ψ_0 denote an elementary mother wavelet,
- Shifted and dilated templates of ψ_0:
  $\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k)$,
- Wavelet Coefficients: $d_{X_\Delta}(j,k) = \langle \psi_{j,k}, X_\Delta \rangle$.



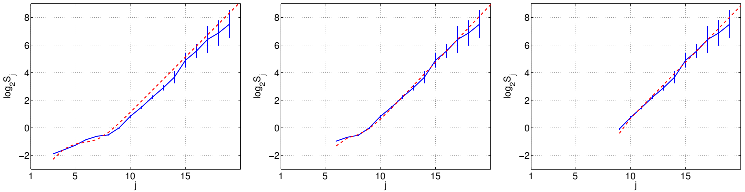
Back

Model (8/) : Jointly 1st and 2nd order statistics

1st order stat. Marginals fitted by Γ-laws.



$\Delta = 4$ms $\Delta = 32$ms $\Delta = 256$ms



2nd order stat. Covariance fitted by a FARIMA(P, d, Q)

Back

TableData

- A variety of traces from major repositories were tested.
- Data collected on the french Renater network, by the METROSEC project (Metrology for Security on the Internet).

Data	Date(Start Time)	T (s)	Network(Link)	# Pkts	IAT (ms)	Repository
PAUG	1989-08-29(11:25)	2620	LAN(10BaseT)	1	2.6	ita.ee.lbl.gov/index.html
LBL-TCP-3	1994-01-20(14:10)	7200	WAN(10BaseT)	1.7	4	ita.ee.lbl.gov/index.html
AUCK-IV	2001-04-02(13:00)	10800	WAN(OC3)	9	1.2	wand.cs.waikato.ac.nz/wand/wits
CAIDA	2002-08-14(10:00)	600	Backbone(OC48)	65	0.01	www.caida.org/
UNC	2003-04-06(16:00)	3600	WAN(10BaseT)	4.6	0.8	www-dirt.cs.unc.edu/ts/
MTS-ref1	2004-12-09(18:30)	5000	LAN(10BaseT)	3.9	1.5	www.laas.fr/METROSEC/
MTS-ref2	2004-12-10(02:00)	9000	LAN(10BaseT)	2.1	4.3	www.laas.fr/METROSEC/
MTS-ref3	2006-03-20(11:00)	3600	LAN(10BaseT)	2.8	3.7	www.laas.fr/METROSEC/
MTS-ref4	2004-12-21(15:00)	3600	LAN(10BaseT)	2.9	3.9	www.laas.fr/METROSEC/

Back

Synthesis of a Γ-farima process

Procedure.

- **Mapping – 1st order stat.**: if $Y_j(k)$ is a Gaussian r.v. with variance $\beta/2$, then

$$X(k) = \sum_{j=1}^{2\alpha} Y_j(k)^2 \tag{3}$$

is a $\Gamma_{\alpha,\beta}$ r.v.

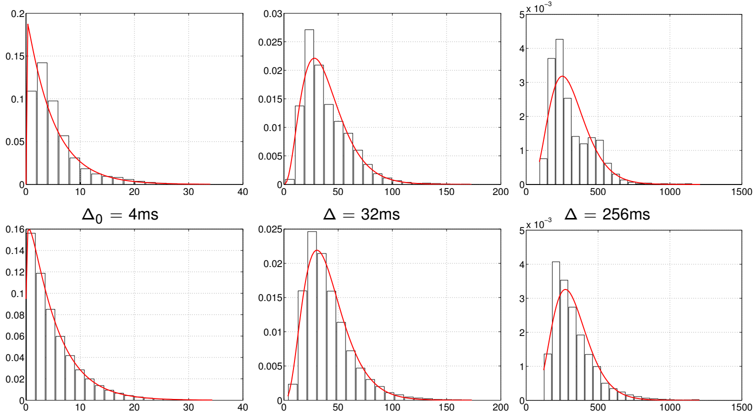
- **Mapping – 2nd order stat.**: as a consequence,

$$\gamma_Y(k) = \sqrt{\gamma_X(k)/4\alpha}. \tag{4}$$

- **Procedure**: generate 2α Gaussian processes with covariance γ_Y derived with (2) from the farima covariance, then obtain X from (1).

Empirical PDF and $\Gamma_{\alpha,\beta}$ models
Metrosec-fc1

1st order stat. Marginals fitted by Γ-laws. Data



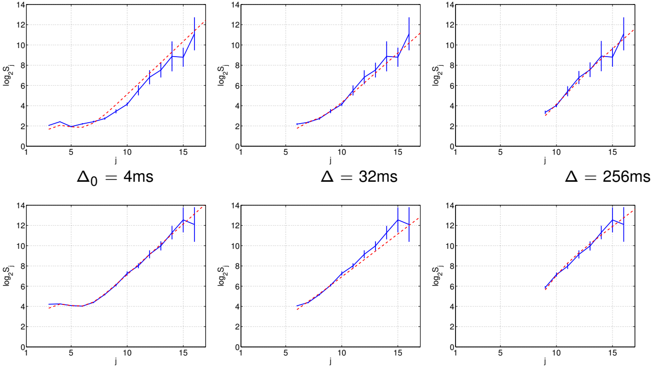
Model

Empirical PDF and $\Gamma_{\alpha,\beta}$ models

Metrosec-fc1

2nd order stat. Log-Scale Diagram

Data



Model

Back