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

P. Abry ¹, P. Borgnat ¹, P. Owezarski ², METROlogy for SECurity (French) Research Project ³

¹Physics Lab, CNRS UMR 5672, ENS Lyon, France, ²LAAS, CNRS UMR xxx, Toulouse, France, ³www.metrosec.fr

Intimate 2006, July 6th-7th, Paris
Outline

1. **Motivation**
2. **Modeling**
   - Principles
   - Marginals
   - Covariances
   - Results, Estimation and Synthesis procedures
3. **detecting**
   - Intuition on examples
   - Principles
   - DataBase
   - Statistical performance
   - Classification
4. **Conclusions and Perspectives**
5. **Appendix**
Motivation and Goals

1. **Statistical Modeling of Aggregated Internet Traffic Time Series**,  
   - How: Gaussian/Non Gaussian ? Short vs Long range dependencies ?  
   - When: What aggregation level ?  

2. **Anomaly Detection and Classification**  
   - How: Intuitions and Principles  
   - What: Performance evaluation ? ⇒ Need for a Database !  
   - How good: results and performance
Outline

1 Motivation

2 Modeling
   - Principles
   - Marginals
   - Covariances
   - Results, Estimation and Synthesis procedures

3 Detecting
   - Intuition on examples
   - Principles
   - DataBase
   - Statistical performance
   - Classification

4 Conclusions and Perspectives

5 Appendix
Outline

1 Motivation

2 Modeling
   • Principles
   • Marginals
   • Covariances
   • Results, Estimation and Synthesis procedures

3 detecting
   • Intuition on examples
   • Principles
   • DataBase
   • Statistical performance
   • Classification

4 Conclusions and Perspectives

5 Appendix
Aggregated Time Series

Aggregation level: \( \Delta \),
Packet Count,
Byte Count,
Model 1: Intuitions

1. How should we chose 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 ?
   - Depends on data ? on goals ?

4. Proposed Solutions
   1. ⇒ Long Range vs Short range dependencies ? Gaussian vs non Gaussian ?
   2. ⇒ Trade-off: Marginals (1st stat order) and covariances (2nd stat order) **jointly**
   3. ⇒ Modeling covariant with a change of aggregation level ?
Outline

1. Motivation
2. Modeling
   - Principles
   - Marginals
   - Covariances
   - Results, Estimation and Synthesis procedures
3. detecting
   - Intuition on examples
   - Principles
   - DataBase
   - Statistical performance
   - Classification
4. Conclusions and Perspectives
5. Appendix
Model (2/): Marginals

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

- Exponential ? Gaussian ?
- Aggregation level ?
Model (3/): 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** \( \alpha \): From Gaussian to exponential, \((1/\alpha \text{ distance from Gaussian})\),
- **Scale parameter** \( \beta \): Multiplicative factor.
Model (4/) : Gamma Fits

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

![Graphs showing empirical PDFs and gamma fits for different aggregation levels](image)

- Accurately Fits data for all aggregation levels $\Delta$,
- Stability under addition:
  \[
  X_1 : \Gamma_{\alpha_1,\beta}, \quad X_2 : \Gamma_{\alpha_2,\beta}, \quad (X_1, X_2) \text{ Indep.} \quad \implies \quad X_1 + X_2 : \Gamma_{\alpha_1 + \alpha_2,\beta},
  \]
- Aggregation:
  \[
  X_{2\Delta}(k) = X_{\Delta}(k) + X_{\Delta}(k + 1),
  \]
Model (5/): $\hat{\alpha}_\Delta, \hat{\beta}_\Delta$

- Stability under addition and Independence $\Rightarrow$
  \[
  \alpha(\Delta) = \alpha_0 \Delta \\
  \beta(\Delta) = \beta_0.
  \]
- $\hat{\alpha}_\Delta, \hat{\beta}_\Delta$ accommodate SRD!
1 Motivation

2 Modeling
   - Principles
   - Marginals
   - Covariances
   - Results, Estimation and Synthesis procedures

3 detecting
   - Intuition on examples
   - Principles
   - DataBase
   - Statistical performance
   - Classification

4 Conclusions and Perspectives

5 Appendix
Model (5/): Covariances: the wavelet point of view

- $X_\Delta$ stationary stochastic process, with spectrum $f_{X_\Delta}(\nu)$,
- Wavelet Coefficients: $d_X(j, k)$,
- Wavelet Spectrum: $S(j) = \frac{1}{n_j} \sum_{k=1}^{n_j} |d_{X_\Delta}(j, k)|^2$,

$$\mathbb{E}S(j) = \int f_X(\nu)2^j|\psi_0(2^j \nu)|^2 du \simeq \hat{f}_X(\nu = 2^{-j} \nu_0).$$
Model (5/) : Both short and long range dependencies

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

- $\Delta_j$, LBL-TCP-3, $\Delta = 1$ms,

- 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.
Model (6/): FARIMA\((P, d, Q)\) covariance

**Definition of Long Range Dependence**

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

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

**Farima = fractionally Integrated ARMA.**

1. fractional integration with parameter \(d\)
2. short-range correlations as an ARMA(1,1), → params \(\theta, \phi\).

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

- \(d\) controls LRD: with \(\gamma = 2d\),
- \(P, Q\) controls SBD,
Model (7/) : Empirical LDs and FARIMA(P,d,Q) Fits

LBL-TCP-3

- Accurately Fits data for all aggregation levels $\Delta$,
- LRD is persistent, SRD are cancelled out,
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).

<table>
<thead>
<tr>
<th>Data</th>
<th>Date(Start Time)</th>
<th>T (s)</th>
<th>Network(Link)</th>
<th>IAT (ms)</th>
<th>Repository</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAUG</td>
<td>1989-08-29(11:25)</td>
<td>2620</td>
<td>LAN(100BaseT)</td>
<td>2.6</td>
<td>ita.ee.lbl.gov</td>
</tr>
<tr>
<td>LBL-TCP-3</td>
<td>1994-01-20(14:10)</td>
<td>7200</td>
<td>WAN(100BaseT)</td>
<td>4</td>
<td>ita.ee.lbl.gov</td>
</tr>
<tr>
<td>AUCK-IV</td>
<td>2001-04-02(13:00)</td>
<td>10800</td>
<td>WAN(OC3)</td>
<td>1.2</td>
<td>wand.cs.waikato.ac.nz</td>
</tr>
<tr>
<td>CAIDA</td>
<td>2002-08-14(10:00)</td>
<td>600</td>
<td>Backbone(OC48)</td>
<td>0.01</td>
<td><a href="http://www.caida.org">www.caida.org</a></td>
</tr>
<tr>
<td>UNC</td>
<td>2003-04-06(16:00)</td>
<td>3600</td>
<td>WAN(100BaseT)</td>
<td>0.8</td>
<td><a href="http://www.dirt.cs.unc.edu">www.dirt.cs.unc.edu</a></td>
</tr>
<tr>
<td>Metrosec-fc1</td>
<td>2005-04-14(14:15)</td>
<td>900</td>
<td>LAN(100BaseT)</td>
<td>0.8</td>
<td><a href="http://www.laas.fr/METROSEC">www.laas.fr/METROSEC</a></td>
</tr>
</tbody>
</table>
Outline

1. Motivation
2. Modeling
   - Principles
   - Marginals
   - Covariances
   - Results, Estimation and Synthesis procedures
3. detecting
   - Intuition on examples
   - Principles
   - DataBase
   - Statistical performance
   - Classification
4. Conclusions and Perspectives
5. Appendix
Model (9/) : Non Gaussian Long Range Dependent Models

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

for various data, various traffics, various links, various networks,

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

- Detection?
Outline

1. Motivation
2. Modeling
   - Principles
   - Marginals
   - Covariances
   - Results, Estimation and Synthesis procedures
3. detecting
   - Intuition on examples
   - Principles
   - DataBase
   - Statistical performance
   - Classification
4. Conclusions and Perspectives
5. Appendix
Outline

1. Motivation
2. Modeling
   - Principles
   - Marginals
   - Covariances
   - Results, Estimation and Synthesis procedures
3. detecting
   - Intuition on examples
   - Principles
   - DataBase
   - Statistical performance
   - Classification
4. Conclusions and Perspectives
5. Appendix
Aggregated Time Series

IPerf, UDP Flooding.

Non Gaussian Long Memory Traffic Modeling - Anomaly Detection
Abry, Borgnat, Owezarski, METROSEC 23 / 51
**Detection (1/): DDoS Attack (UDP Flooding)**

LogScale Diagrams

- **Black**: before, **Red**: during, **Blue**: After Attack
- **LRD** not caused by nor altered by DDoS Attacks.
Detection (1/): DDoS Attack (UDP Flooding)
Gamma Fits (during attack)

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

1 Motivation
2 Modeling
   ● Principles
   ● Marginals
   ● Covariances
   ● Results, Estimation and Synthesis procedures
3 detecting
   ● Intuition on examples
   ● Principles
   ● DataBase
   ● Statistical performance
   ● Classification
4 Conclusions and Perspectives
5 Appendix
Detection (2/): Principles

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

\[
D_\alpha(l) = \frac{1}{J} \sum_{j=1}^{J} (\hat{\alpha}_{2^j}(l) - \hat{\alpha}_{2^j}(\text{ref}))^2, \quad (1)
\]

\[
D_\beta(l) = \frac{1}{J} \sum_{j=1}^{J} (\hat{\beta}_{2^j}(l) - \hat{\beta}_{2^j}(\text{ref}))^2. \quad (2)
\]

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

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

\[
D_\alpha(l) = \frac{1}{J} \sum_{j=1}^{J} (\hat{\alpha}_{2^j}(l) - \hat{\alpha}_{2^j}(\text{ref}))^2, \quad (1)
\]

\[
D_\beta(l) = \frac{1}{J} \sum_{j=1}^{J} (\hat{\beta}_{2^j}(l) - \hat{\beta}_{2^j}(\text{ref}))^2. \quad (2)
\]

Choose a threshold $\lambda$ to decide when the distance is too large, $D_\alpha(l) \geq \lambda$, 

Detection (3/): Example 1: DDoS Attack
Detection (4/): Example 2: Artificial Traffic Increase
Detection (5/): Statistical performance?

- Receiver Operating Curves:
  How many false positive given the false negative?
  \[ P_D = f(P_F) \text{ or } P_D = f(\lambda), \quad P_F = f(\lambda), \]

\[ \begin{array}{c|c|c}
Prob. False Alarm & 0 & 0.2 & 0.4 & 0.6 & 0.8 & 1 \\
\hline
Prob. Detection & 1 & 0.8 & 0.6 & 0.4 & 0.2 & 0 \\
\end{array} \]

\[ \begin{array}{c|c|c}
Threshold & 0 & 0.5 & 1 & 1.5 & 2 \\
\hline
Prob. Detection and False Alarm & 1 & 0.8 & 0.6 & 0.4 & 0.2 \\
\end{array} \]

⇒ Need for a documented anomaly database !!!!
Outline

1. Motivation
2. Modeling
   - Principles
   - Marginals
   - Covariances
   - Results, Estimation and Synthesis procedures
3. detecting
   - Intuition on examples
   - Principles
   - DataBase
      - Statistical performance
      - Classification
4. Conclusions and Perspectives
5. Appendix
METROSEC partners (all over France),
RENATER network,
# Anomaly DataBase - DDoS Attacks

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

<table>
<thead>
<tr>
<th>Id</th>
<th>$t_i$</th>
<th>$T(s)$</th>
<th>$t_a$</th>
<th>$T_A(s)$</th>
<th>$D$</th>
<th>$V$</th>
<th>$I$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FC performed by human</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC-1</td>
<td>13:45</td>
<td>7200</td>
<td>14:30</td>
<td>1800</td>
<td>-</td>
<td>-</td>
<td>31.27</td>
</tr>
<tr>
<td>FC-2</td>
<td>15:00</td>
<td>7200</td>
<td>15:45</td>
<td>1800</td>
<td>-</td>
<td>-</td>
<td>18.35</td>
</tr>
</tbody>
</table>
Outline

1. Motivation
2. Modeling
   - Principles
   - Marginals
   - Covariances
   - Results, Estimation and Synthesis procedures
3. detecting
   - Intuition on examples
   - Principles
   - DataBase
   - Statistical performance
4. Conclusions and Perspectives
5. Appendix
Detection (6/): Receiver Operational Curves
## Detection (6/): Detection Probability

<table>
<thead>
<tr>
<th>Type of Anomaly</th>
<th>performed with</th>
<th>Id</th>
<th>Intens. (%)</th>
<th>$P_D$ $P_F = 10%$</th>
<th>$P_F = 20%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>FC</td>
<td>Humans</td>
<td>1</td>
<td>31.27</td>
<td>04</td>
<td>14</td>
</tr>
<tr>
<td>DDoS</td>
<td>Iperf</td>
<td>R</td>
<td>33.82</td>
<td>91</td>
<td>93</td>
</tr>
<tr>
<td>DDoS</td>
<td>Iperf</td>
<td>I</td>
<td>17.06</td>
<td>51</td>
<td>64</td>
</tr>
<tr>
<td>DDoS</td>
<td>Iperf</td>
<td>II</td>
<td>14.83</td>
<td>48</td>
<td>54</td>
</tr>
<tr>
<td>DDoS</td>
<td>Iperf</td>
<td>III</td>
<td>21.51</td>
<td>48</td>
<td>58</td>
</tr>
<tr>
<td>DDoS</td>
<td>Iperf</td>
<td>IV</td>
<td>33.29</td>
<td>33</td>
<td>50</td>
</tr>
<tr>
<td>DDoS</td>
<td>Iperf</td>
<td>V</td>
<td>39.26</td>
<td>18</td>
<td>40</td>
</tr>
<tr>
<td>DDoS</td>
<td>Iperf</td>
<td>A</td>
<td>34.94</td>
<td>21</td>
<td>50</td>
</tr>
<tr>
<td>DDoS</td>
<td>Iperf</td>
<td>B</td>
<td>40.39</td>
<td>81</td>
<td>87</td>
</tr>
<tr>
<td>DDoS</td>
<td>Iperf</td>
<td>C</td>
<td>36.93</td>
<td>52</td>
<td>58</td>
</tr>
<tr>
<td>DDoS</td>
<td>Trinoo</td>
<td>tM</td>
<td>4.64</td>
<td>27</td>
<td>50</td>
</tr>
<tr>
<td>DDoS</td>
<td>Trinoo</td>
<td>tN</td>
<td>15.18</td>
<td>54</td>
<td>54</td>
</tr>
<tr>
<td>DDoS</td>
<td>Trinoo</td>
<td>tT</td>
<td>82.85</td>
<td>82</td>
<td>82</td>
</tr>
</tbody>
</table>
Detection (6/): use of other distances

- **Kullback divergence**: 
  \[ KD(p_1, p_2) = \int (p_1 - p_2)(\ln p_1 - \ln p_2) \, dx , \]
- **1D**: 
  \[ K^{(1D)}_\Delta(l) = KD(p_\Delta,l, p_\Delta,Ref), \]
- **2D**: 
  \[ K^{(2D)}_{\Delta,\Delta'}(l) = KD(p_\Delta,\Delta',l, p_\Delta,\Delta',Ref). \]

<table>
<thead>
<tr>
<th></th>
<th>( D_{\alpha} )</th>
<th>( K^{(1D)}_{24} )</th>
<th>( K^{(1D)}_{27} )</th>
<th>( K^{(2D)}_{24,27} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>51 : 64</td>
<td>25 : 64</td>
<td>35 : 67</td>
<td>25 : 51</td>
</tr>
<tr>
<td>II</td>
<td>48 : 54</td>
<td>35 : 58</td>
<td>35 : 61</td>
<td>35 : 61</td>
</tr>
<tr>
<td>III</td>
<td>48 : 58</td>
<td>74 : 93</td>
<td>70 : 83</td>
<td>87 : 93</td>
</tr>
<tr>
<td>IV</td>
<td>33 : 50</td>
<td>56 : 67</td>
<td>56 : 69</td>
<td>34 : 62</td>
</tr>
<tr>
<td>V</td>
<td>18 : 40</td>
<td>87 : 96</td>
<td>34 : 93</td>
<td>90 : 96</td>
</tr>
<tr>
<td>A</td>
<td>21 : 50</td>
<td>50 : 78</td>
<td>37 : 59</td>
<td>53 : 81</td>
</tr>
<tr>
<td>B</td>
<td>81 : 87</td>
<td>78 : 78</td>
<td>09 : 33</td>
<td>78 : 81</td>
</tr>
<tr>
<td>C</td>
<td>52 : 58</td>
<td>91 : 91</td>
<td>91 : 91</td>
<td>91 : 91</td>
</tr>
<tr>
<td>X</td>
<td>93 : 96</td>
<td>93 : 93</td>
<td>93 : 93</td>
<td>93 : 93</td>
</tr>
<tr>
<td>tM</td>
<td>27 : 55</td>
<td>36 : 91</td>
<td>36 : 91</td>
<td>45 : 91</td>
</tr>
<tr>
<td>tN</td>
<td>54 : 54</td>
<td>73 : 91</td>
<td>91 : 91</td>
<td>55 : 73</td>
</tr>
<tr>
<td>tT</td>
<td>82 : 82</td>
<td>100 : 100</td>
<td>100 : 100</td>
<td>100 : 100</td>
</tr>
</tbody>
</table>
Outline

1. Motivation
2. Modeling
   - Principles
   - Marginals
   - Covariances
   - Results, Estimation and Synthesis procedures
3. detecting
   - Intuition on examples
   - Principles
   - DataBase
   - Statistical performance
   - Classification
4. Conclusions and Perspectives
5. Appendix
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

1. Motivation
2. Modeling
   - Principles
   - Marginals
   - Covariances
   - Results, Estimation and Synthesis procedures
3. detecting
   - Intuition on examples
   - Principles
   - DataBase
   - Statistical performance
   - Classification
4. Conclusions and Perspectives
5. Appendix
<table>
<thead>
<tr>
<th>Motivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modeling</td>
</tr>
<tr>
<td>detecting</td>
</tr>
<tr>
<td>Conclusions and Perspectives</td>
</tr>
<tr>
<td>Appendix</td>
</tr>
</tbody>
</table>
### Outline

1. Motivation
2. Modeling
   - Principles
   - Marginals
   - Covariances
   - Results, Estimation and Synthesis procedures
3. detecting
   - Intuition on examples
   - Principles
   - DataBase
   - Statistical performance
   - Classification
4. Conclusions and Perspectives
5. Appendix
Wavelet Transform

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

![Wavelet Transform Diagram]
Model (8/): Jointly 1st and 2nd order statistics

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

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

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

Procedure.

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

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

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

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

  $$\gamma_Y(k) = \sqrt[4]{\gamma_X(k)/4\alpha}.$$  \hspace{1cm} (4)

- **Procedure**: generate $2\alpha$ Gaussian processes with covariance $\gamma_Y$ 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 $\Gamma$-laws.

Data

Model

Non Gaussian Long Memory Traffic Modeling - Anomaly Detection
Abry, Borgnat, Owezarski, METROSEC 50 / 51
Empirical PDF and $\Gamma_{\alpha,\beta}$ models
Metrosec-fc1

2nd order stat. Log-Scale Diagram

- $\Delta_0 = 4\text{ms}$
- $\Delta = 32\text{ms}$
- $\Delta = 256\text{ms}$