Infinite Divisibility and Traffic Data

Patrice Abry* and Darryl Veitch**

* CNRS–ENS Lyon, France,
  http://www.ens-lyon.fr/~pabry

** EMULab, University of Melbourne, Australia
  http://www.emulab.ee.mu.oz.au/~darryl

Thanks to
Pierre Chainais, Stephane Roux, Patrick Flandrin (CNRS–ENS Lyon),
Lidong Huang (EMULab), Jörg Micheel (WAND, Waikato, NZ)
Telecommunications Networks and Their Traffic

Networks: A Deep Hierarchy of Systems

- Physical Layer technology (coding, error recovery..)
- Circuit vs Packet Switched Paradigms
- Connection oriented vs connectionless philosophy
- Protocols and their encapsulation, eg: Ethernet[ IP [ TCP[HTTP] ] ]
- Routing, signaling, admission control...

Tele–Traffic: A Turbulent River of Myriad Sources

- Geographic Complexity
  - Network Edge: distributed sources, destinations,
  - Internally: confluence and splitting of streams at switches
- Offered Traffic Complexity
  - User ‘sessions’, durations, arrivals ..
  - Applications, underlying protocols
  - Underlying data objects: files, video, audio..
- Temporal Complexity: the offered traffic is time-varying
  - Time scale rich: $\mu$s to months, 1kbps to Terabits/s
- Burstiness a key feature
  - Temporal: dependence in time, often LRD
  - Amplitude: large fluctuations, often highly non-Gaussian
In General

- Highly complex, and users ‘independent’, so model as stochastic.
- Can model Source or Aggregate behaviour.
- Black box statistical model or structural hierarchal models, or both.
- Can analyse many different time series derived from the (sub)streams:
  - Time indexed: # new connections, # connections current, ...
  - Connection/Flow indexed: durations, # of pkts/bytes, inter-arrival times
- Historically Markovien models dominated (tractability, low burstiness).
- Currently a lot of measurement driven work.
- Now accepted that scaling a robust feature of packet traffic.

In This Talk

- Consider aggregate Internet traffic focusing on TCP (2Mpbs link).
- Describe scaling features: both temporal and amplitude burstiness.
- Model as Black Box, ‘enlightened black box’, and discuss network origins.
- Introduce Infinitely Divisible Cascades as a modelling class.
- Use wavelets as the natural analysis tool for scaling data.
Examples of Scaling in Traffic

**ETHERNET (LAN)**

**INTERNET (WAN)**

**NB DE PAQUETS**

**(bloc 10 MS)**

**NB DE CONNESSIONS ACTIVES**

**(bloc 10 MS)**

![Graphs showing traffic patterns](image-url)
Available Scaling Models

Philosophy of scaling:

No characteristic scale, but invariant relationships between scales

Families of Scaling Processes

- $1/f$ processes ($1/f^\alpha$)
- (exactly) Self-similar (SS) processes
- Long-Range Dependent (LRD), and anti-persistent processes
- Locally self-similar processes
- Processes with fractal sample paths
- Multifractal measures/processes
- Multiplicative Cascades (eg Conservative MCs)
- Infinitely Divisible Cascades (IDC)

Use as Traffic Models

- **Using LRD:** $X(t)$ is an On/Off Process, with infinite variance for On/Off.
- **Using SS:** $Y(t) = m(t) + \sigma(t)Z_H(t)$, with $Z_H(t)$ a fractional BM, where

  \[
  X(t) \quad \text{is the traffic rate at time } t
  \]

  \[
  Y(t) \quad \text{is the total traffic arriving in } [0, t], \text{ ie } Y_t = \int_0^t X(t).
  \]
Examples of Scaling in Traffic: 2nd Order Wavelet Analysis

**ETHERNET**: bytes per 12ms bin.

**INTERNET**: new connections per 10ms bin.

Logscale Diagram, $N=2$  
$[ (j_1,j_2)= (3,15), \ \alpha-est = 0.59, \ Q= 0.011384 ], \ D-init$

Logscale Diagram, $N=2$  
$[ (j_1,j_2)= (8,19), \ \alpha-est = 0.59, \ Q= 0.81665 ]$
Wavelets are ideal for scaling processes, as they are localised time-scale tools.

- From the mother wavelet, \( \psi_0(u) \), satisfying \( \int \psi_0(u)du = 0 \) one defines the
- Wavelet ‘bases’ \( \psi_{a,t} = \frac{1}{|a|} \psi_0\left(\frac{u-t}{a}\right) \) note dilation with scale!
- Continuous Wavelet Transform of \( X(t) \) is coefficients: \( T_X(a, t) = \langle X, \psi_{a,t} \rangle, a \geq 0 \).
- Discrete Wavelet Transform are the same coefficients, but only those on the dyadic grid: \( d_X(j, k) = T_X(a = 2^j, k2^j) \). These can be calculated in an \( O(n) \) algorithm.
- The number of vanishing moments, i.e., the largest \( N \geq 1 \) such that \( \int t^k \psi(t) dt \equiv 0 \) for \( k = 0, 1, \ldots N - 1 \) controls the ability to ‘cancel’ LRD and even non-stationarity.
Properties

Definition: If $X = \{X(t), t \in \mathcal{R}\}$, then $\{X(t)\} \overset{d}{=} \{c^{-H}X(ct)\}$ for any $c > 0$.

Wavelet detail processes: $\{d(j, k), k \in \mathcal{Z}\} \overset{d}{=} \{2^{j(H+1/2)}d(0, k), k \in \mathcal{Z}\}$.

Unproblematic detail dependencies:

$\{d(j, k), k \in \mathcal{Z}\}$ stationary for each $j$ fixed

$\text{Cov} [d(j, k)d(j, k')] \leq 2^j |k - k'|^{2(H-N)}$, for $|2^j k - 2^j k'| \to \infty$.

Scaling Exponent revealed in moments: $\mathbb{E}|d(j, k)|^q = 2^{jqH}\mathbb{E}|d(0, k)|^q$

Exploitation in Estimation

Moments Estimated via: $S_q(j) = \frac{1}{n_j} \sum_k |d(j, k)|^q$

Exponent Estimated via: weighted regression of $\log_2[S_q(j)]$ vs $j$, slope is $qH$.

This simple SS scaling seen in diverse traffic time series over large scales
Smaller Scales, Richer Models, ‘Multiscaling’
Beyond Mono-Parameter Scaling

Self-Similarity: \[ \mathbb{E}|d(j, k)|^q = C_q(2^j)^{qH} = C_q \exp(qH\ln(\alpha)) \]
- A single scaling parameter \( H \)
- Power-laws

Multi-Scaling: \[ \mathbb{E}|d(j, k)|^q = C_q(2^j)^{H(q)} = C_q \exp(H(q)\ln(\alpha)) \]
- A collection of parameters: \( H(q) \)
- Power-laws
Beyond Power-Law Scaling

Self-Similarity: \( E|d(j, k)|^q = C_q(2^j)^{qH} = C_q \exp(qH \ln(a)) \)
- A single scaling parameter \( H \)
- Power-laws

Multi-Scaling: \( E|d(j, k)|^q = C_q(2^j)^{H(q)} = C_q \exp(H(q) \ln(a)) \)
- A collection of parameters: \( H(q) \)
- Power-laws

Infinitely Divisible Cascade: \( E|d(j, k)|^q = C_q \exp(H(q) n(a)) \)
- Two collections of parameters: \( H(q), n(a) \)
- No Power-Law!
- but separability of order \( q \) and scale \( a \).

But! relative power-laws remain:
\[ E|d_X(j, k)|^q = C_{p,q}(E|d(j, k)|^p)^{H(q)/H(p)} \]
- tested by plotting \( \log_2 S_q(j) \) vs \( \log_2 S_p(j) \), if passes,
- \( H(q) \) and \( n(a) \) can be estimated.
Testing for IDC in TCP Connection Arrivals

Estimation of $H_p(q) = H_1(0.5)$, $N = 3$

Estimation of $H_p(q) = H_1(1)$, $N = 3$

Estimation of $H_p(q) = H_1(1.5)$, $N = 3$

Estimation of $H_p(q) = H_1(2)$, $N = 3$

Estimation of $H_p(q) = H_1(2.5)$, $N = 3$

Estimation of $H_p(q) = H_1(3)$, $N = 3$

Estimation of $H_p(q) = H_1(4)$, $N = 3$

Estimation of $H_p(q) = H_1(5)$, $N = 3$

Estimation of $H_p(q) = H_1(6)$, $N = 3$
Estimating the IDC in TCP Connection Arrivals

Unlike MS, IDC hypothesis is valid over both scaling ranges.
Assuming the IDC formalism, we have the following relationship between scales $a_1 > a_2$:

$$E|T_2|^q = e^{-H(q)(n(a_2) - n(a_1))}E|T_1|^q$$

Trick: let $T$ be a positive r.v. with PDF $F_T$, generating function $\hat{F}_T$, and $U = \ln T$.

$$E|T|^q = Ee^{q\ln|T|} = Ee^{qU} \equiv \hat{F}_U(q)$$

We can rewrite in terms of moment generating functions of the $U_a = \ln |T_a|$:

$$\hat{F}_{U_2}(q) = e^{-H(q)(n(a_2) - n(a_1))}\hat{F}_{U_1}(q)$$

which describes the connection between the marginals at two scales, and in fact any set of scales $a_1 > a_2 > \cdots a_{n-1} > a_n$.

Think of convolution, interpret $\exp\{-H(q)(n(a_2) - n(a_1))\}$ as the MGF of a distribution function $G_{a_1,a_2}$, the propagator, possessing the semi-group property: $G_{a_1,a_3} = G_{a_1,a_2} * G_{a_2,a_3}$.

**Special Cases**

- **H-SS**: $G_{a_1,a_2} = \delta(-H\ln(a_1/a_2))$, translation in scale: $F_1(d) = F_2\left(\frac{d}{\alpha_0}\right)$, $\alpha_0 = \left(\frac{a_1}{a_2}\right)^H$

- **MS**: $n(a_2) - n(a_1) = \ln\left(\frac{a_2}{a_1}\right)$
Results for the Inter-Arrival Series of TCP Connections

EMULab

H(q)/H(p), p=1

log₂(T)

log(P(log(T)))

log T

log(P(log(T)))

log T
So Where is the Infinite Divisibility?

The kernel is a family of DF’s forming a convolution semi-group, with generator corresponding to $-H(q)$.

Connection to Infinite Divisibility:

*The class of DFs from semi-groups of convolution is the class of ID distributions.*

- Natural to let the $\{U_a\}$ be ID, though perhaps not necessary.
- Connection to marginals of $\{X(t)\}$, and other constraints? needs further work.
- A Markov process in scale $\Rightarrow$ Fokker Planck, evolution along decreasing scale.

Interpretation for $X(t)$

- A scaling creating, multiplicative, operation is performed, of a nature given by $H(q)$
- $n(a_2) - n(a_1)$ is the number of times performed between scales $a_1$ and $a_2$. 
Returning to Traffic

Where did we find IDC type scaling?

- Time indexed: arrivals, departures, active. (For UDP also)
- Connection indexed: inter-arrivals

We did not find it in:

- Connection indexed: durations, # bytes, # packets (modelled as i.i.d. pareto).

Main Results:

- When two scaling regimes are found, IDC is usually verified, showing that in fact the two are not unrelated.
- The generator $-H(q)$ does not change past the change point, but $n(a)$ does, the same scaling laws are being applied, but at a different ‘speed’.
- The deviation of $H(q)$ from linearity is weak, ‘barely multifractal’.
- Very similar statistics between arrivals and departures.
- Details inside connections essential to explain burstiness.
Current Directions

On the Cascades

- Understanding of the implications of the formalism on \( \{X(t)\} \), what is allowed?
- Precise examples of (stationary) IDC processes, and their generation.
- Understanding the connection between arrivals and departures,
  Wild Conjecture: A IDC point process is invariant under random reordering.

On Estimation Theory

- Refinement of wavelet estimators of \( H(q), n(a) \).
- Accurate calculation of confidence intervals at each \( q \).

On the Modelling

- Identification of the network mechanism modifying \( n(a) \),
  Conjecture: the logarithmic rescaling of TCP slow start, acting up to a (distribution of) round trip times.
- Understanding the connection between arrivals and departures – can we exclude network based causes?