WAVELET-BASED ESTIMATION OF LONG-RANGE DEPENDENCE IN VIDEO AND NETWORK TRAFFIC TRACES

Nikola Cackov
ncackov@cs.sfu.ca

Communication Networks Laboratory
http://www.ensc.sfu.ca/cnl
Simon Fraser University

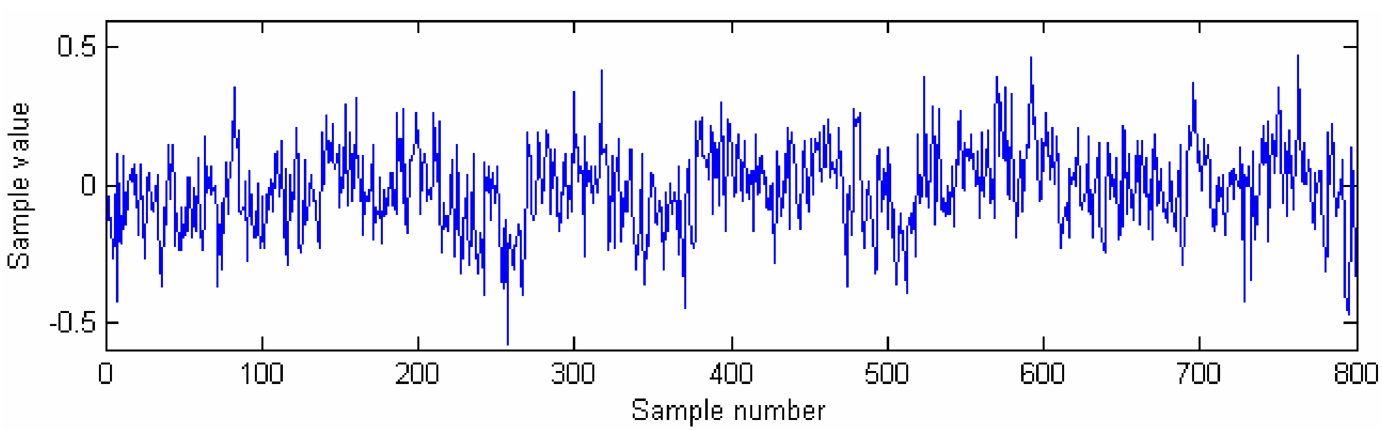
communication networks laboratory

Road map

- Long-range dependence and self-similarity
- Wavelets and estimation of the Hurst parameter
- MPEG video traces
  - performance analysis
- E-Comm traffic traces
  - Hurst parameter estimation of traffic traces collected from a deployed wireless network
- Conclusions

Long-range dependence: properties

- High variability:
  - variance of the sample mean decays more slowly than the reciprocal value of the sample size
- Burstiness over a range of timescales:
  - long runs of large values followed by long runs of small values, repeated in aperiodic pattern



LRD: definition

Slow decay of the autocorrelation function ρ(k) of a (wide-sense) stationary process X(n):

$$\sum_{k=-\infty}^{\infty} \rho(k) = \infty$$ definition

$$\rho(k) = c_\rho k^{\alpha-1}, \; k \to \infty$$ model

$$f(\nu) = c_f \, |\nu|^{-\alpha}, \; \nu \to 0$$ corollary

where $f(\nu)$ is the power spectral density of X(n), c_ρ and c_f are nonzero constants, and $0<\alpha<1$

LRD: Long-range dependence

Second-order self-similarity

X(n): a wide-sense stationary discrete-time process with autocorrelation function $\rho(k)$

Define:

$$X^{(m)}(i) = \frac{1}{m} \sum_{n=m(i-1)+1}^{mi} X(n), \quad m \in \mathrm{N}$$

- X(n) is second-order self-similar if

$$\rho^{(m)}(k) = \rho(k) = \frac{1}{2}\left[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right]$$

where $\rho^{(m)}(k)$ is autocorrelation function of $X^{(m)}(i)$

- H: Hurst parameter

LRD and self-similarity

- X(n): second-order self-similar process

$$\rho(k) \sim H(2H-1)k^{2H-2},\ k \to \infty,\ 0 < H < 1$$

- X(n): LRD process

$$\rho(k) \sim c_\rho k^{\alpha-1},\ k \to \infty$$

- Second-order self-similarity implies LRD and vice-versa, with restriction
  - $0.5 < H < 1$

  with $\alpha = 2H-1$ or $H = 0.5(\alpha+1)$
- Important: X(n) is wide-sense stationary

Road map

- Long-range dependence and self-similarity
- Wavelets and estimation of the Hurst parameter
- MPEG video traces
- E-Comm traffic traces
- Conclusions

Wavelet coefficients

- Discrete wavelet transform of a signal $X(t)$:

$$d_{j,k} = \int_{-\infty}^{\infty} X(t)\psi_{j,k}(t)dt$$ wavelet coefficients

where

$$\psi_{j,k}(t) = 2^{-j/2}\psi\left(2^{-j}t - k\right)$$

- $\psi(t)$: mother wavelet
  - j: octave, k: translation
- Important property:

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \Leftrightarrow \left|\Psi(\omega)\right|^2_{\omega=0} = 0$$ admissibility condition

Reconstruction formula

$$X(t) = \sum_{j=0}^{\infty} \sum_{k} d_{j,k} \psi_{j,k}(t)$$

$$= \sum_{j=j_2+1}^{\infty} \sum_{k} d_{j,k} \psi_{j,k}(t) + \sum_{j=0}^{j_2} \sum_{k} d_{j,k} \psi_{j,k}(t)$$

$$= \sum_{k} c_{j_2,k} \varphi_{j_2,k}(t) + \sum_{j=0}^{j_2} \sum_{k} d_{j,k} \psi_{j,k}(t)$$

approximation at octave j_2

details from octave 0 to octave j_2

$\varphi_{j_2,k}(t)$: scaling function

$\psi_{j,k}(t)$: wavelet function

Wavelets and LRD

- Let X(t) be LRD process (wide-sense stationary)
  - $f(\nu) \sim c_f \, |\nu|^{-\alpha}, \; \nu \to 0$: its power spectral density
- The following holds:

$$\log_2 \mathrm{E}\{d_{j,k}^2\} = \alpha j + c$$

- Estimation of the wavelet power spectrum:

$$\mathrm{E}\{d_{j,k}^2\} = \frac{1}{n_j} \sum_{k=1}^{n_j} d_{j,k}^2$$

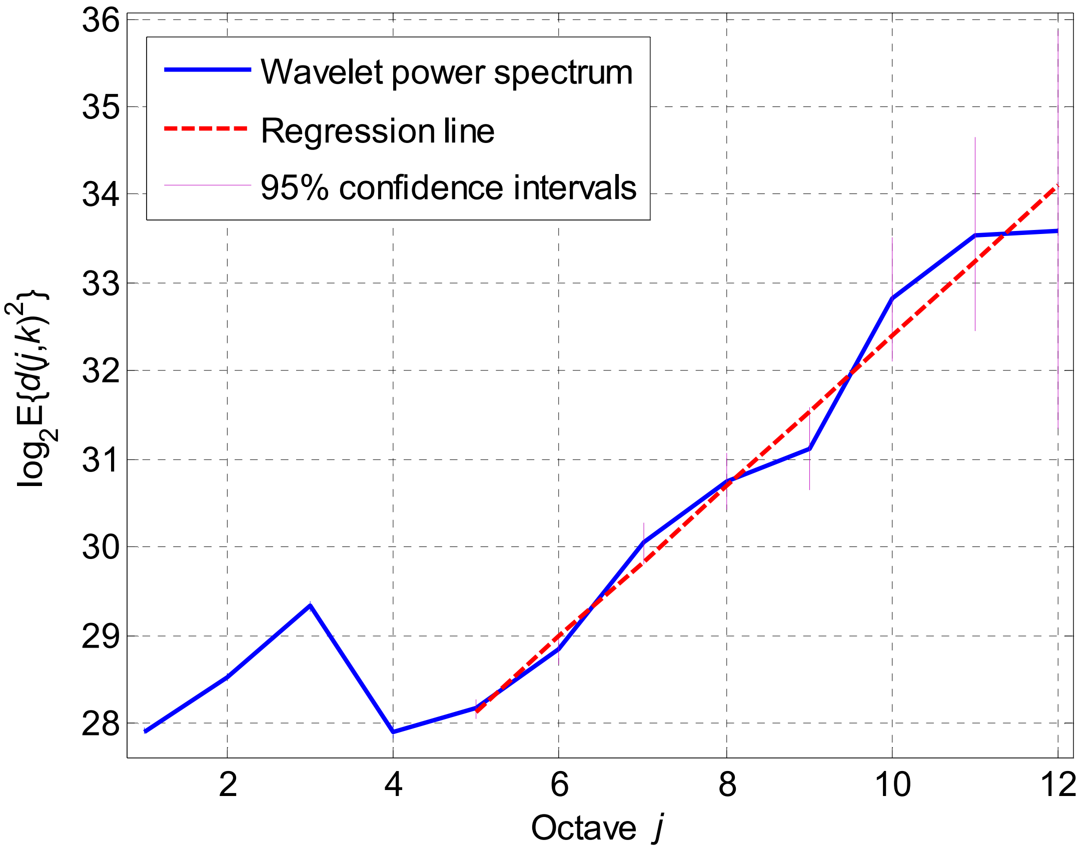
  - n_j: number of wavelet coefficients at octave j

D. Veitch and P. Abry, "A wavelet-based joint estimator of the parameters of long-range dependence," *IEEE Trans. on Information Theory*, vol.45, no.3, pp. 878-897, Apr. 1999.

Estimation of α and H

- Logscale diagram: plot of $\log_2 E\{d_{j,k}^2\}$ vs. j (octave)
- Linear relationship between $\log_2 E\{d_{j,k}^2\}$ and j on the coarsest octaves indicates LRD
- Estimation of α:
  - linear regression of $\log_2 E\{d_{j,k}^2\}$ on j in the linear region of the logscale diagram
- $H = 0.5 (\alpha + 1)$

Logscale diagram: example



- Simpsons (MPEG-1)
- α=0.852, H=0.926 (octaves 5-12)

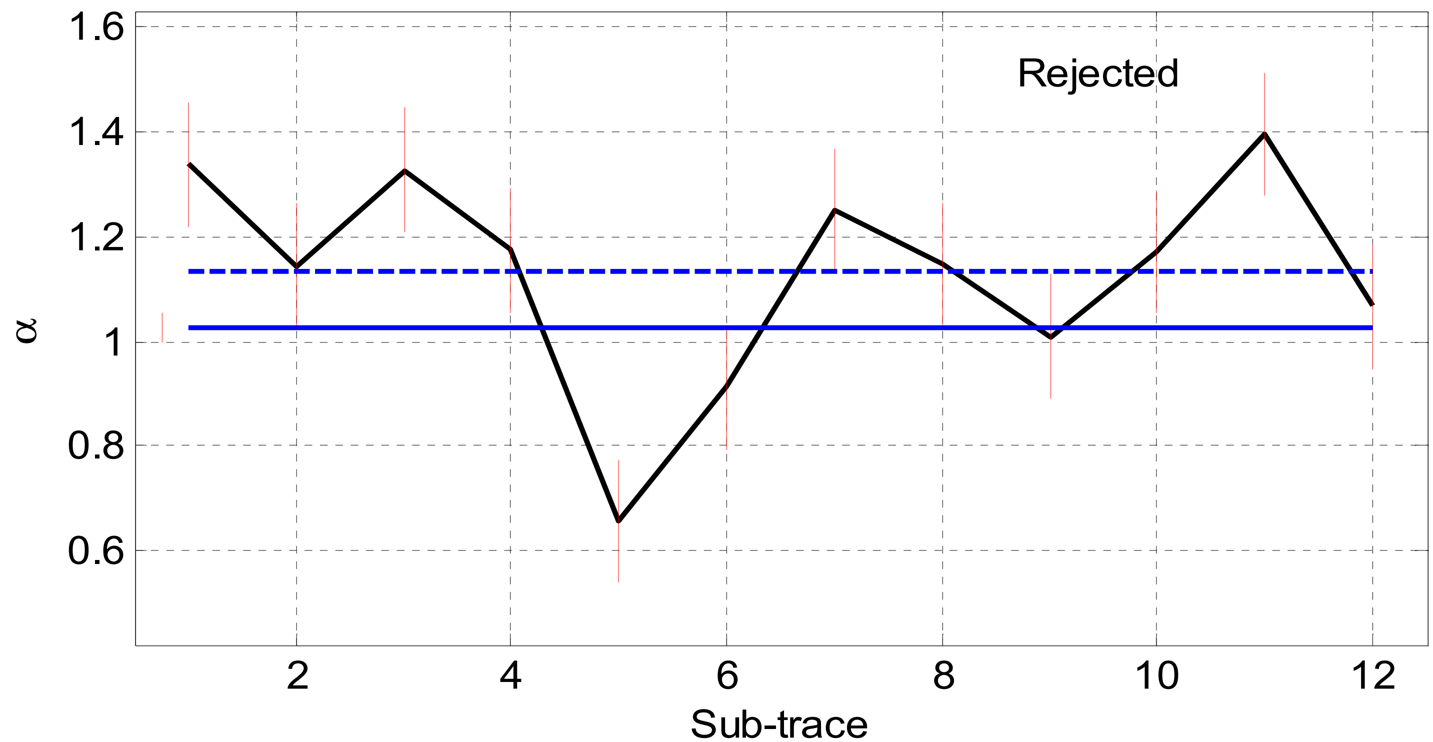
Idealizations

- For fixed j: $d_{j,k}$ are stationary sequences of uncorrelated variables
- Processes $d_{j,k}$ and $d_{j',k}$ ($j \neq j'$) are uncorrelated
- Process X(t) and, hence, the processes $d_{j,k}$ are Gaussian

Test for time constancy of α

- X(n): wide-sense stationary process
  - α does not depend on n
- Is α constant throughout the time series X(n)?
- Approach:
  - divide X(n) into m blocks of equal lengths
  - estimate α for each block
  - compare the estimates
- If α varies significantly, estimating α for the entire time series is not meaningful

Test for constancy: example



- solid line: estimate of α for the entire trace
- dashed line: average of the estimates for the blocks

Star Wars IV (MPEG-4)

Road map

- Long-range dependence and self-similarity
- Wavelets and estimation of the Hurst parameter
- MPEG video traces
- E-Comm traffic traces
- Conclusions

Video traces

- MPEG-1 and MPEG-4 encoded video traces:
  - 22,498–89,998 frames/trace
  - 25 frames/second
  - 15–60 minutes of video
- Each entry: size of the frame (bits)

University of Wuerzburg, index of MPEG traces:
http://www-info3.informatik.uni-wuerzburg.de/MPEG/traces/

University of Berlin, MPEG-4 video traces for network performance evaluation:
http://www.tkn.ee.tu-berlin.de/research/trace/trace.html

Estimates of H

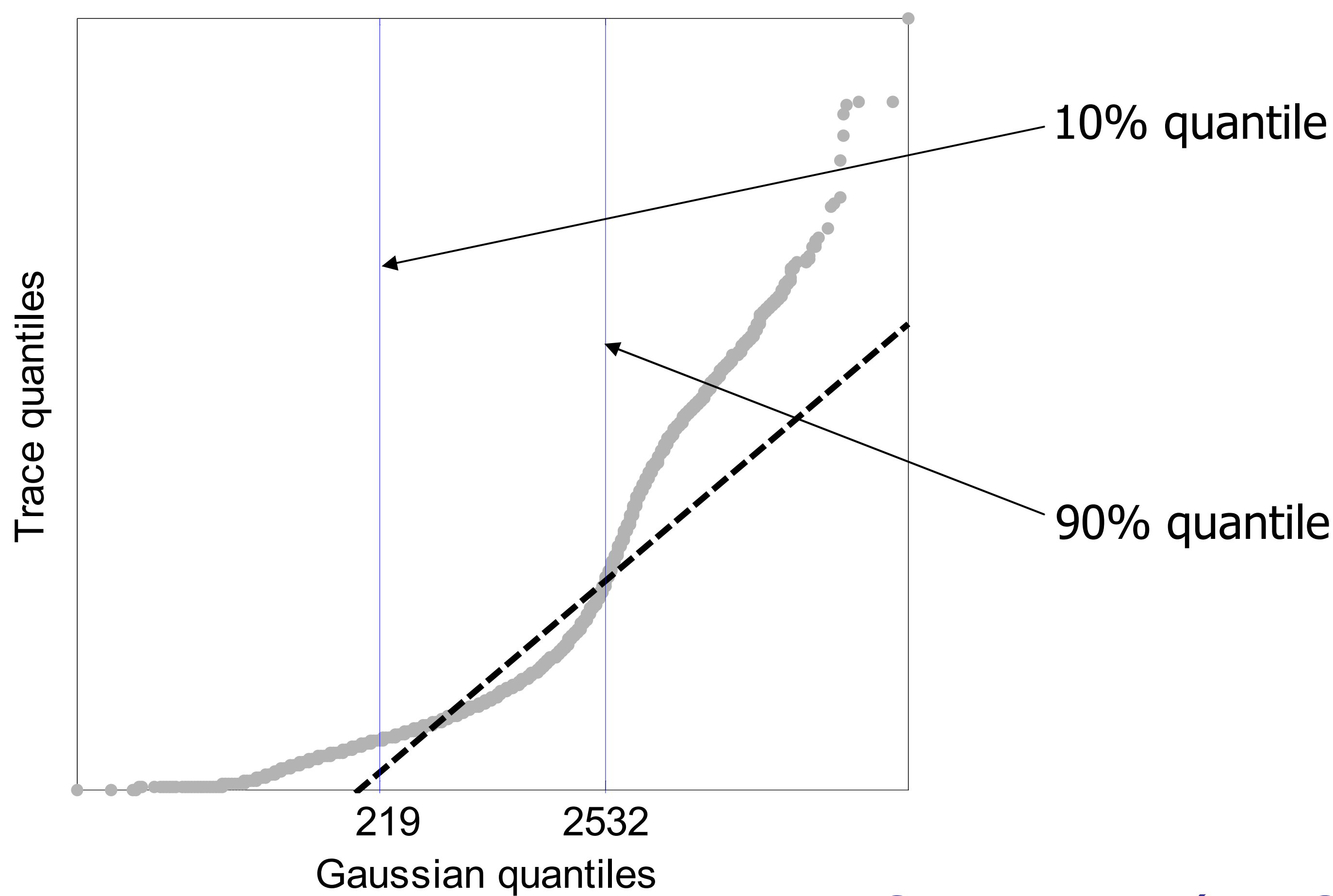
Trace	Encoding	H (wav.)	H (per.)	H (R/S)
MTV	MPEG-1	0.959	0.992	0.89
Jurassic park	MPEG-1	1.096	1.191	0.88
Simpsons	MPEG-1	0.926	0.988	0.89
Mr. Bean	MPEG-1	1.214	1.295	0.85
Silence of the lambs	MPEG-1	1.130	1.171	0.89
Talk show	MPEG-1	1.084	1.174	0.89
ARD news	MPEG-4	1.382	1.310	0.967
Die hard III	MPEG-4	1.190	1.233	0.969
Formula 1	MPEG-4	1.189	1.216	0.867
Futurama	MPEG-4	0.943	1.064	0.877
From dusk till dawn	MPEG-4	1.139	1.186	0.909
First contact	MPEG-4	1.194	1.268	0.931

Unreliable estimates (MPEG)

- Wavelet-based estimator often yields H>1
- Possible causes:
  - non-Gaussianity of data
  - non-stationarity of α
  - LRD-SRD interactions
- Perform additional tests:
  - Gaussianity
  - time constancy of α
  - investigation of LRD and SRD components

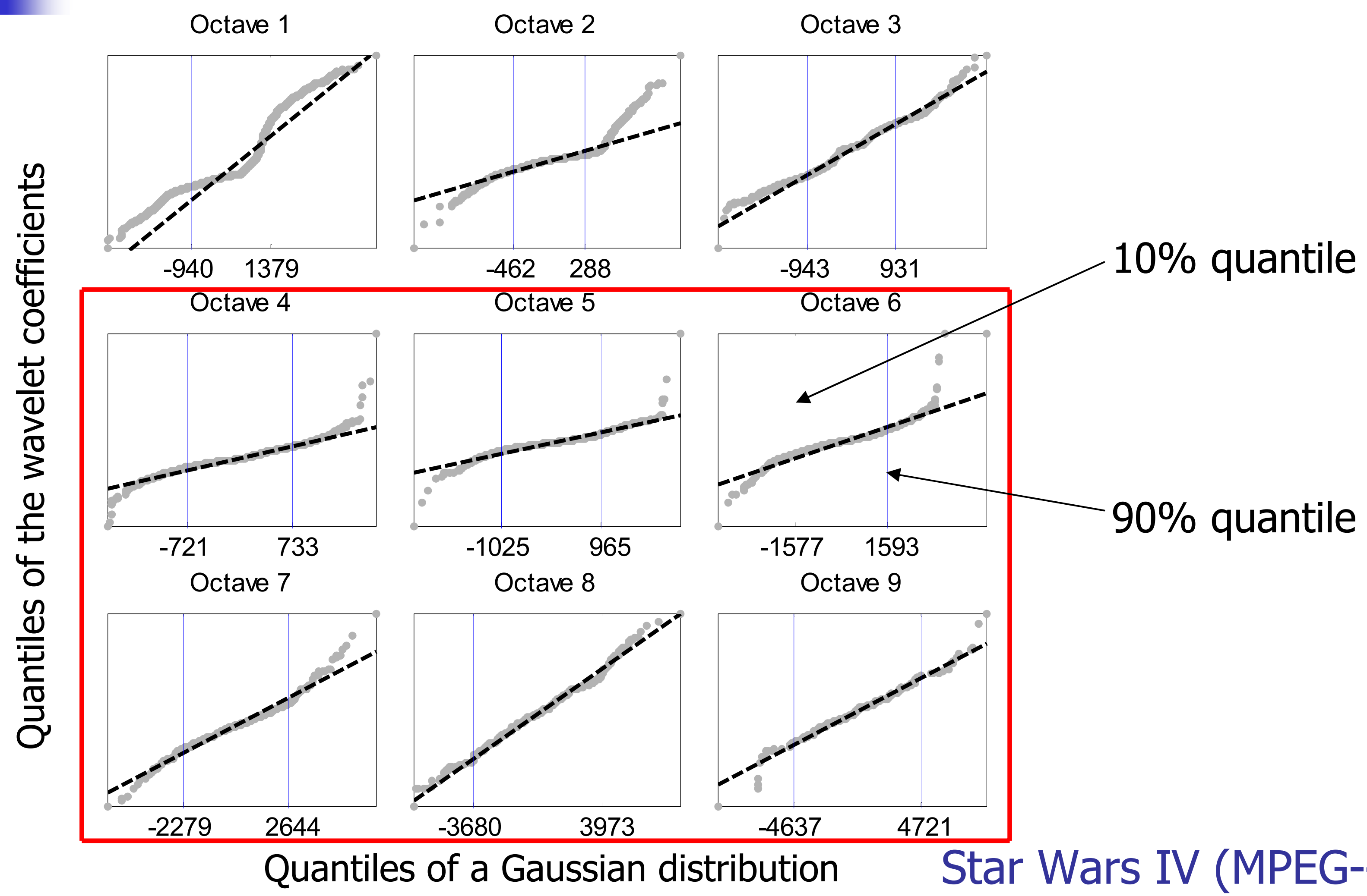
SRD: Short-range dependence

Test for Gaussianity: qq-plot



Star Wars IV (MPEG-4)

Test for Gaussianity: qq-plots



Star Wars IV (MPEG-4)

qq-plots: observations

- Analyzed traces are highly non-Gaussian
- Wavelet coefficients on the finest octaves deviate from Gaussian distribution
- Wavelet coefficients on the middle and coarse octaves have approximately Gaussian distribution

Test for time constancy of α

Trace	Encoding	failed	passed
MTV	MPEG-1	7	0
Jurassic park	MPEG-1	7	0
Simpsons	MPEG-1	2	5
Mr. Bean	MPEG-1	7	0
Silence of the lambs	MPEG-1	7	0
Talk show	MPEG-1	7	0
ARD news	MPEG-4	3	4
Die hard III	MPEG-4	7	0
Formula 1	MPEG-4	4	3
Futurama	MPEG-4	7	0
From dusk till dawn	MPEG-4	6	1
First contact	MPEG-4	7	0

LRD-SRD interactions

- R/S estimates of H are often greater than 0.9
  - strong LRD component
- Conjecture: traces posses strong SRD component
  - similarities within a single scene of video
- Wavelet-based estimator produces unreliable results when applied to traces with both:
  - strong short-range dependent (SRD) and
  - strong LRD components

F. Xue and Lj. Trajković, "Performance analysis of a wavelet-based Hurst parameter estimator for self-similar network traffic," in *Proc. SPECTS '2K*, Vancouver, BC, May 2000, pp. 294–298.

MPEG traces: discussion

- Wavelet coefficients on the coarsest scales have Gaussian distribution
- Generally α is not constant for the entire trace
- LRD-SRD interactions may cause unreliable estimates
- Wavelet- and periodogram-based estimators of H produce similar results
  - estimated power spectral density obeys a power law with exponent α often greater than one

Road map

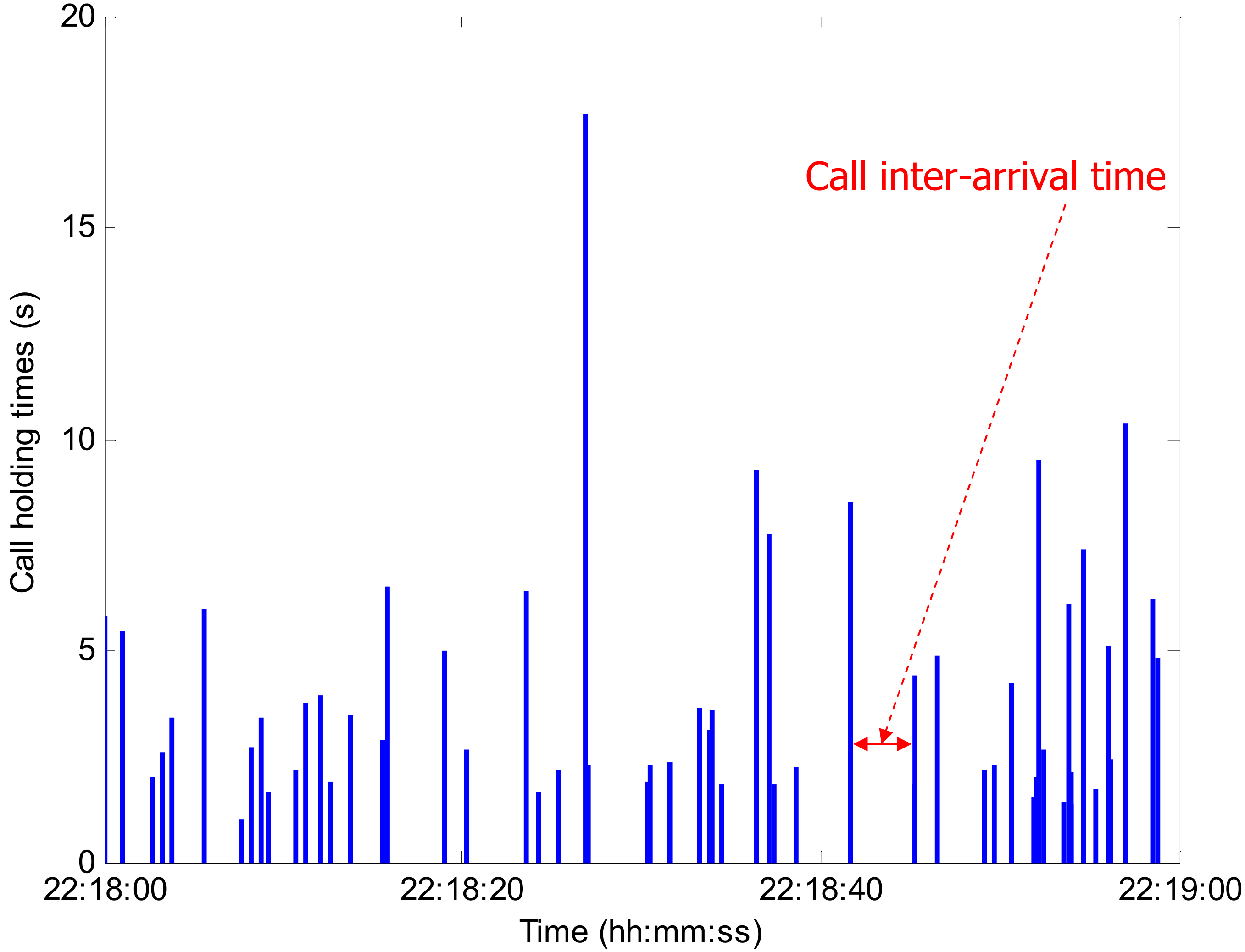
- Long-range dependence and self-similarity
- Wavelets and estimation of the Hurst parameter
- MPEG video traces
- E-Comm traffic traces
- Conclusions

E-Comm traffic traces

- E-Comm: wireless network with cellular infrastructure
  - central site and 11 cells
- We analyzed call holding and call inter-arrival times collected from cell 1 (Vancouver):
  - 1–2 November, 2001
    - daily and hourly traces
  - 1–7 March, 2002
    - weekly, daily, and hourly traces
  - 24–30 March, 2003
    - weekly, daily, and hourly traces

hourly: 5 busy hours

E-Comm traffic trace: example



Traffic analysis

- Test the time constancy of α for various values of m (number of blocks):
  - we consider the estimate of H reliable if the trace passes the test for more than 50% of tests
- Estimate H from the logscale diagram

Analysis of call holding times

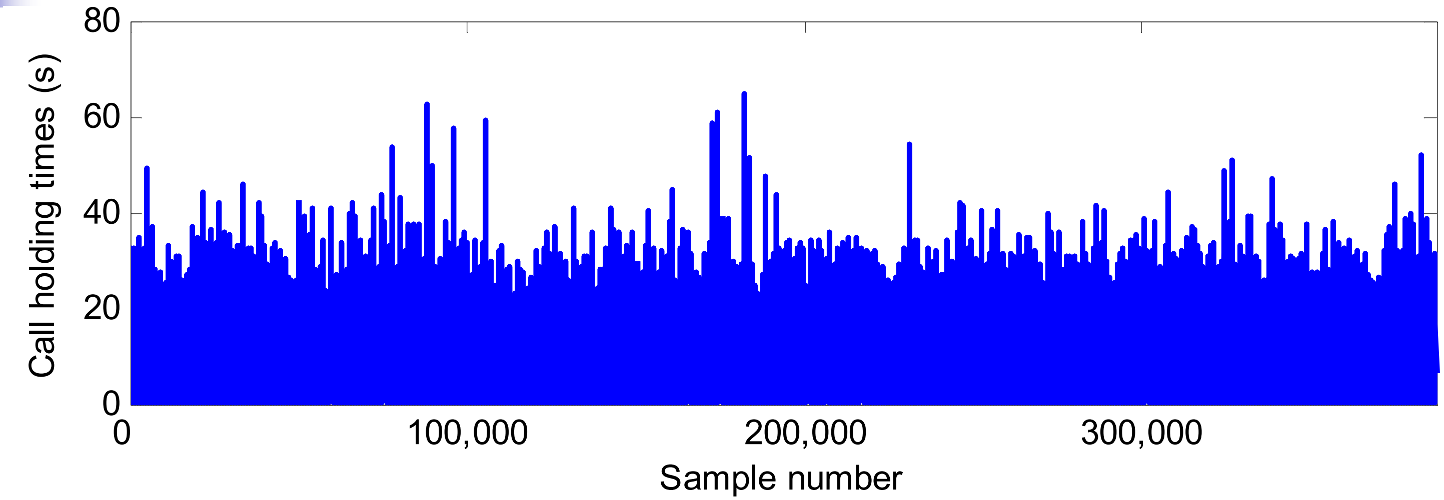
	Trace type	Daily	Hourly
2001	Test for time constancy of α	2/2	5/5
	H	0.561–0.583	0.462–0.493
2002	Test for time constancy of α	7/7	5/5
	H	0.560–0.623	0.460–0.508
2003	Test for time constancy of α	7/7	4/5
	H	0.566–0.617	0.463–0.526

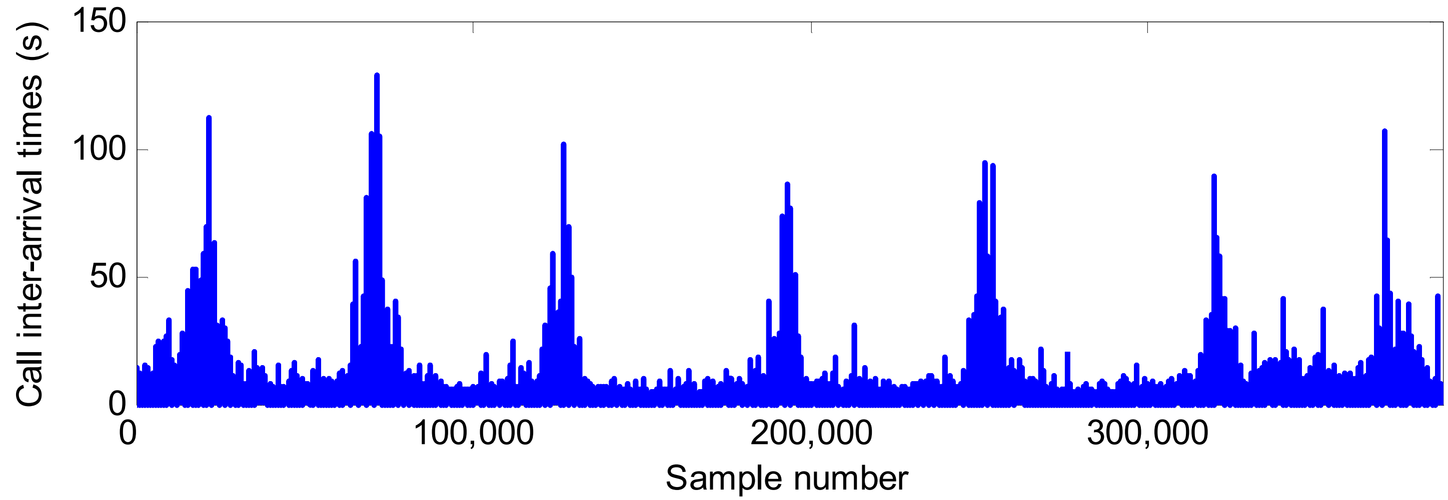
Analysis of call inter-arrival times

Trace type		Daily	Hourly
2001	Test for time constancy of α	0/2	5/5
	H	0.732–0.737	0.663–0.907
2002	Test for time constancy of α	4/7	5/5
	H	0.706–0.784	0.679–0.780
2003	Test for time constancy of α	6/7	5/5
	H	0.685–0.794	0.696–0.832

D. Sharp, N. Cackov, N. Lasković, Q. Shao, and Lj. Trajković, "Analysis of public safety traffic on trunked land mobile radio systems," *IEEE JSAC*, vol. 22, no. 7, pp. 1197-1205, Sept. 2004.

Weekly traces: 24–30 March, 2003

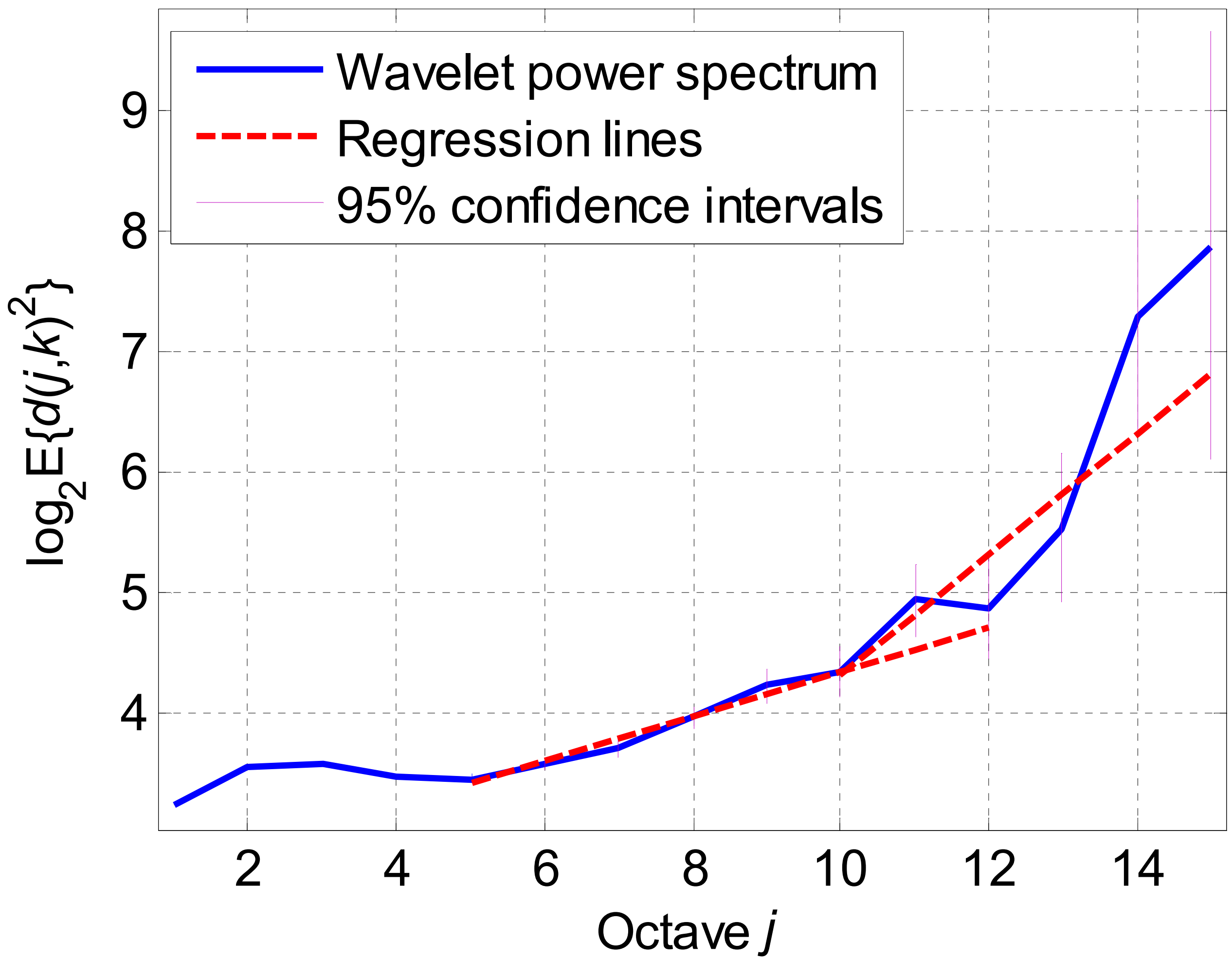




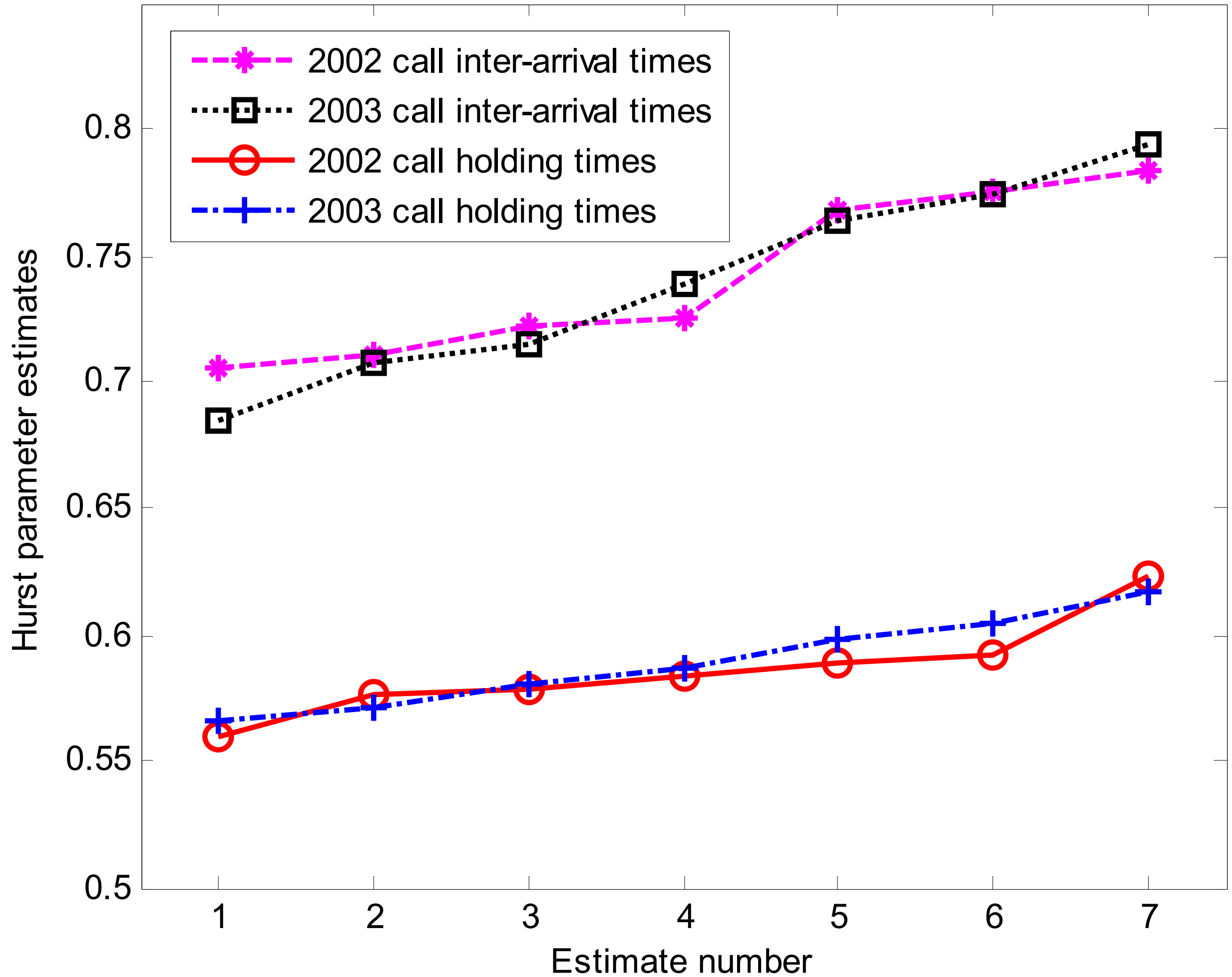
Weekly traces: observations

- Daily cycles are present in call inter-arrival times
- Test for time constancy of α:
  - call holding times: pass
  - call inter-arrival times: fail
- Estimates of the Hurst parameter:
  - call holding times
    - 2002: 0.614
    - 2003: 0.592 [5–12] and 0.751 [10–15] (bi-scaling)
  - call inter-arrival times
    - bi-scaling, with H>1 on the coarser octaves [10-15]

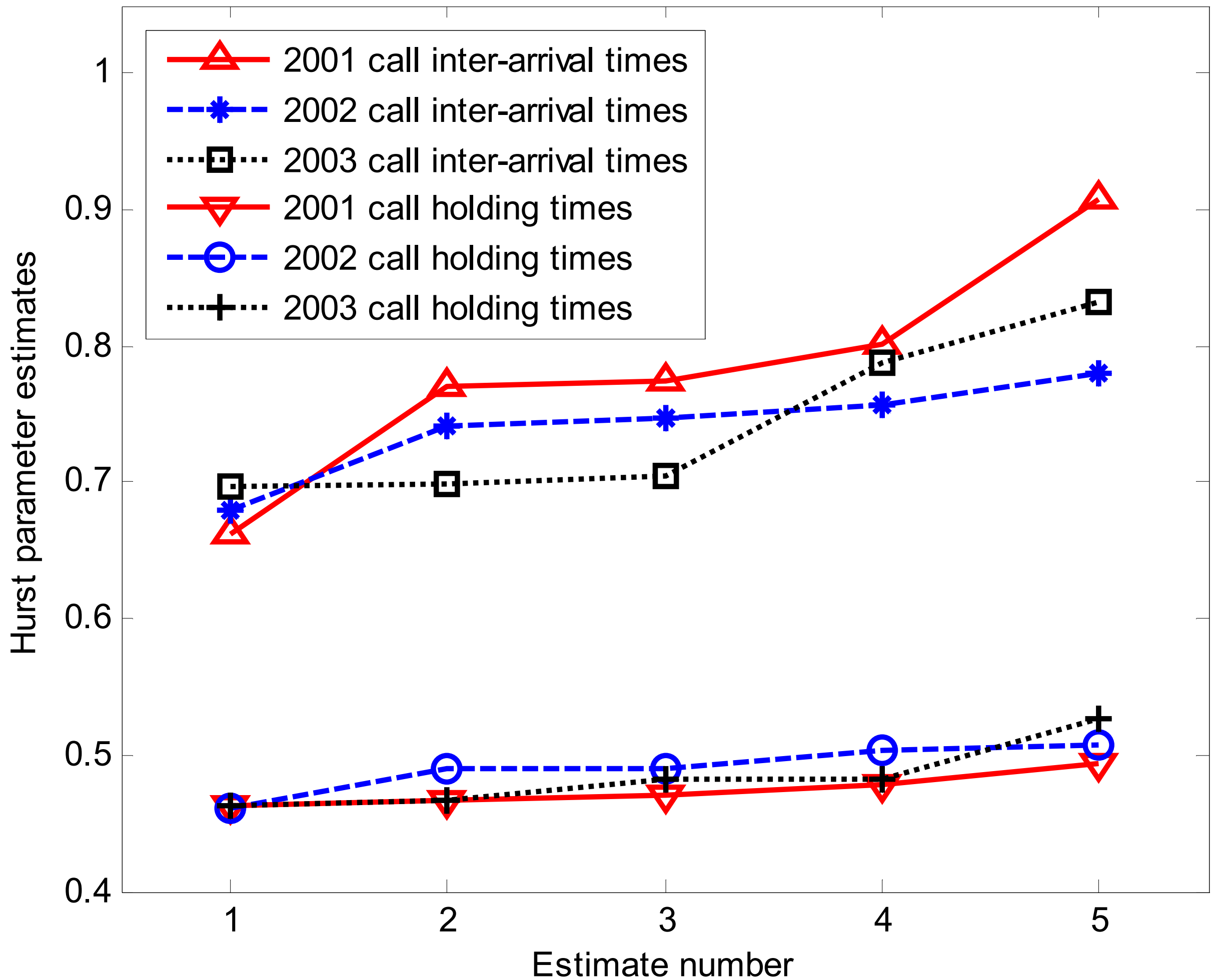
Weekly traces: logscale diagram



Daily traces: comparison of H



Hourly traces: comparison of H



Road map

- Long-range dependence and self-similarity
- Wavelets and estimation of the Hurst parameter
- MPEG video traces
- E-Comm traffic traces
- Conclusions

Conclusions: MPEG traces

- We estimated H for several MPEG traces using the wavelet-based estimator
  - estimates of H were often greater than one
- Wavelet-based estimates of H:
  - differ from those obtained by R/S estimator
  - agree with periodogram-based estimates
- Most MPEG traces fail test for constancy of α
- Unreliability of the estimates may be attributed to the interactions of SRD and LRD components in the trace

Conclusions: E-Comm traces

- We analyzed traffic from cell 1 in the E-Comm network
- Call holding times:
  - mainly pass the test for constancy of α
  - hourly traces: $H \approx 0.5$, independent
  - daily traces: $0.55 < H < 0.6$, weakly LRD
- Call inter-arrival times:
  - weekly traces fail the test for constancy of α
  - daily and hourly traces generally pass the test
  - $0.6 < H < 0.9$: LRD
- These findings agree with previously published results

D. Sharp, N. Cackov, N. Lasković, Q. Shao, and Lj. Trajković, "Analysis of public safety traffic on trunked land mobile radio systems," *IEEE JSAC*, vol. 22, no. 7, pp. 1197-1205, Sept. 2004.

References

[1] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of Ethernet traffic," *IEEE/ACM Trans. Networking*, vol. 2, pp. 1–15, Feb. 1994.

[2] K. Park and W. Willinger, "Self-similar network traffic: an overview," in Self-similar Network Traffic and Performance Evaluation, edited by K. Park and W. Willinger. New York: Wiley, 2000, pp. 1–38.

[3] O. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE Signal Processing Magazine*, vol. 8, issue 4, pp. 14–38, Oct. 1991.

[4] P. Abry, P. Flandrin, M. S. Taqqu, and D. Veitch, "Wavelets for the analysis, estimation, and synthesis of scaling data," in Self-similar Network Traffic and Performance Evaluation, edited by K. Park and W. Willinger. New York: Wiley, 2000, pp. 39–88.

[5] P. Abry and D. Veitch, "Wavelet analysis of long-range dependent traffic," *IEEE Trans. Information Theory*, vol.44, no.1, pp.2–15, Jan. 1998.

[6] D. Veitch and P. Abry, "A wavelet-based joint estimator of the parameters of long-range dependence," *IEEE Trans. Information Theory*, vol.45, no.3, pp.878–897, Apr. 1999.

[7] D. Veitch and P. Abry, "A statistical test for time constancy of scaling exponents," *IEEE Trans. Signal Processing*, vol.49, no.10, pp.2325–2334, Oct. 2001.

[8] D. Veitch, MATLAB code for estimation of scaling exponents: http://www.cubinlab.ee.mu.oz.au/~darryl/secondorder_code.html (accessed June 2004).

References

[9] University of Wuerzburg, Index of MPEG traces: http://www-info3.informatik.uni-wuerzburg.de/MPEG/traces/ (accessed June 2004).

[10] University of Berlin, MPEG-4 and H.263 video traces for network performance evaluation: http://www-tkn.ee.tu-berlin.de/research/trace/trace.html (accessed June 2004).

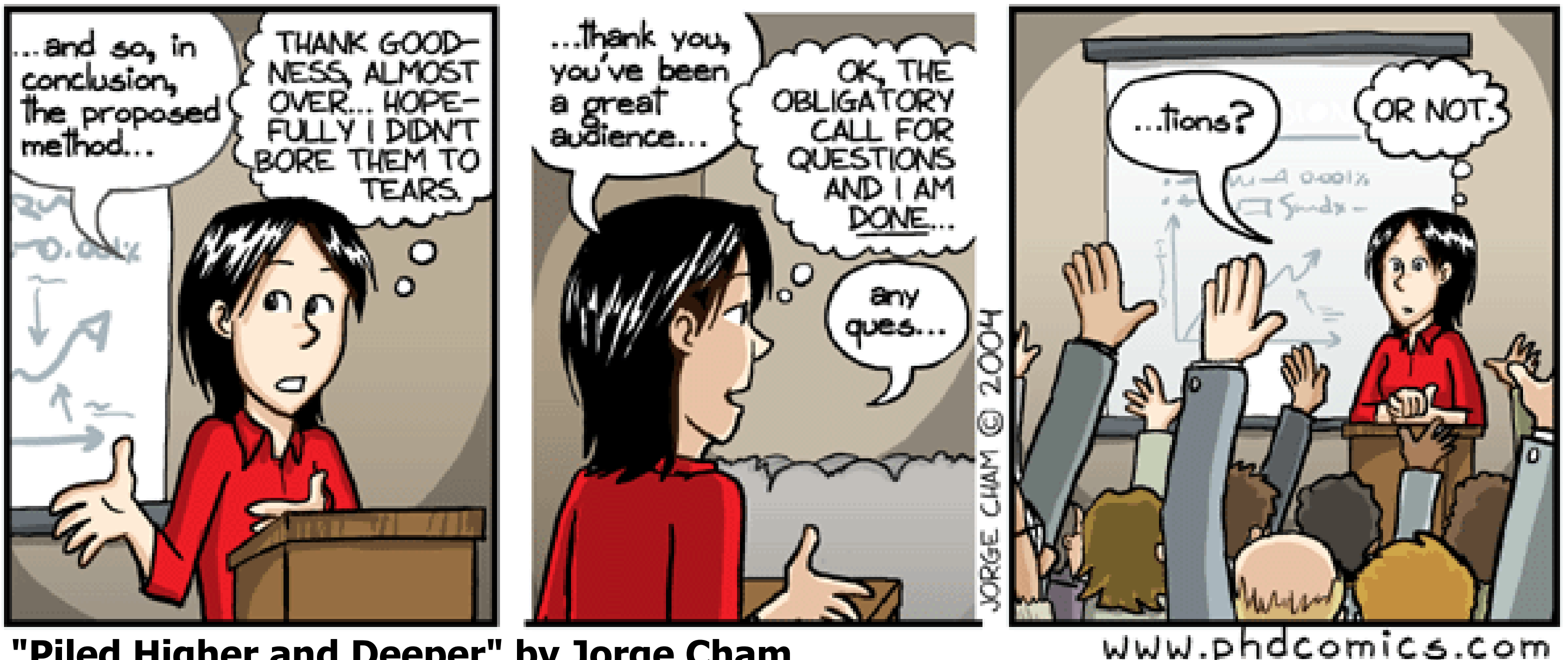
[11] O. Rose, "Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems", University of Wuerzburg, Institute of Computer Science, Report no. 101, Feb. 1995.

[12] D. Sharp, N. Cackov, N. Lasković, Q. Shao, and Lj. Trajković, "Analysis of public safety traffic on trunked land mobile radio systems," *IEEE Journal on Selected Areas in Communications (JSAC) Special Issue on Quality of Service Delivery in Variable Topology Networks*, vol. 22, no. 7, pp. 1197–1205, Sept. 2004.

[13] N. Cackov, B. Vujičić, S. Vujičić, and Lj. Trajković, "Using network activity data to model the utilization of a trunked radio system," *Proc. SPECTS '04*, San Jose, CA, July 2004, pp. 517–524.

[14] N. Cackov, Ž. Lučić, M. Bogdanov, and Lj. Trajković, "Wavelet-based estimation of long-range dependence in MPEG video traces," to be presented at *IEEE Int. Symp. Circuits and Systems 2005*, Kobe, Japan, May 2005.

Questions?



"Piled Higher and Deeper" by Jorge Cham