Traffic Decomposition & Characterization

Angelos K. Marnerides

Infolab21,Computing Department
Lancaster University
Lancaster, UK
a.marnerides@comp.lancs.ac.uk

Under guidance and collaboration with
David Hutchison - Lancaster University
Dimitrios P. Pezaros - University of Glasgow
Hyun-chul Kim -Seoul National University

Lancaster University
Computing Department
Faculty of Applied Sciences

InfoLab21

Overview

- Motivation & Research Questions
- Our approach
- Theory Recap : Bispectrum,bicoherence & Hinich Algorthms
- Data & Results
- Towards protocol-specific anomaly detection
- ICMP example
- WVD comparison with Wavelets
- Conclusions & Future Work

Motivation & Research Questions

- Have ever the “dynamic” protocol characteristics of operational traffic been statistically visualized and fully justified?
- Traffic modeling assumptions not thoroughly investigated
  - Linearity?
  - Gaussianity?
  - Stationarity?
- Current statistical techniques involving identification of linearity and gaussianity involve simple descriptors such as 1st and 2nd order moment sequences of a process (i.e. mean , variance, autocorrelation sequence, etc..)
- “Bucket” traffic modeling sets limits to tasks such as anomaly detection.
- Macroscopic vs. microscopic traffic view

Our approach

- Employment of microscopic traffic view
  - -Volume-based analysis on short duration traces
- Traffic Decomposition
  - -Protocol modelling
- Introducing Traffic characterization using Higher Order Spectral Analysis
  - - Polyspectra (mainly Bispectrum and Bicoherence)
  - - Hinich Algorithms
  - - Cohen Class Energy Distributions for anomaly detection.
  - - Instant frequency and group delay for stationarity.

Bispectrum, Bicoherence & Hinich algorithms

- Bispectrum * defined as the FT of the 3rd order cumulant sequence for a real process X(k)

$$C(\omega_1,\omega_2)=\sum_{\tau_1=-\infty}^{+\infty}\sum_{\tau_2=-\infty}^{+\infty}C_3(\tau_1,\tau_2)\exp\{-j(\omega_1\tau_1+\omega_2\tau_2)\}$$

- Bicoherence * : squared normalized version of the bispectrum

- Hinich algorithms (Linearity/ Gaussianity test)

  -IF 3rd order cumulant =0 => bispectrum and bicoherence =0

  -IF bispectrum != 0 => non-Gaussian process

  -IF process linear and non-Gaussian => bicoherence !=0 and constant

* interested people on proofs and definitions please refer to: **Mendel JM. "Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications."** ***Proceedings of the IEEE*, 79, 3, 278-305** *

Hinich Algorithms (cont..)

- **Step 1: Hypothesis testing for non-zero bispectrum**

H1 : bispectrum $y(n) \neq 0$

H2: bispectrum $y(n) = 0$

IF H1==TRUE we can test for linearity

- **Step 2: Hypothesis for bicoherence**

H1` : bicoherence $b(n) \neq const$

H0` : bicoherence $b(n) = const$

IF H0`==TRUE process is linear

Data & Results

- Hour–long full pcap trace from a Gb Ethernet Link at KEIO University, JP

  - divided in 30-min bins (KEIO1,KEIO2)

  - extracted # of bytes and pkts for each unidirectional flow for TCP,UDP, ICMP

- Hour-long full pcap trace from a US-JP link (WIDE) 100 Mbps FastEthernet link (SamplePoint B – MAWI Working group)

  – divided in 4, 15-min bins (USJP1,USJP2,USJP3,USJP4)

  - extracted # bytes and pkts for each unidirectional flow for TCP,UDP,ICMP

Data & Results (cont..)

KEIO1 (duration 30 mins)

	TCP	UDP	ICMP
Bytes	Linear & NG	Non-Linear & NG	Non-Linear & NG
Packets	Linear & NG	Non-Linear & NG	Non-Linear & NG

KEIO2 (duration 30 mins)

	TCP	UDP	ICMP
Bytes	*Non-Linear & NG*	Non-Linear & NG	*Linear & NG*
Packets	Linear & NG	Non-Linear & NG	*Linear & NG*

US-JAPAN Link – Trace 1 (duration 15 mins)

	TCP	UDP	ICMP
Bytes	Linear & NG	Linear & NG	Linear & NG
Packets	Linear & NG	Linear & NG	Non-Linear & NG

US-JAPAN Link – Trace 2 (duration 15mins)

	TCP	UDP	ICMP
Bytes	Linear & NG	*Non-Linear & NG*	Linear & NG
Packets	Linear & NG	Linear & NG	Non-Linear & NG

US-JAPAN Link – Trace 3 (duration 15mins)

	TCP	UDP	ICMP
Bytes	Linear & NG	Non-Linear & NG	*Non-Linear & NG*
Packets	*Non-Linear & NG*	Non-Linear & NG	Linear & NG

US-JAPAN Link – Trace 4 (duration 15mins)

	TCP	UDP	ICMP
Bytes	*Non-Linear & NG*	*Linear & NG*	Non-Linear & NG
Packets	*Linear & NG*	*Linear & NG*	*Non-Linear & NG*

Towards protocol-specific anomaly detection

- Non-linearity & non-stationary case: applicability of energy distributions in contrast to traditional linear (a.k.a atomic) TF representations as Wavelets and the Short Fourier Transforms (STFT).

- We use particularly the Wigner-Ville distribution (WVD), a member of the Cohen Class distributions defined as:

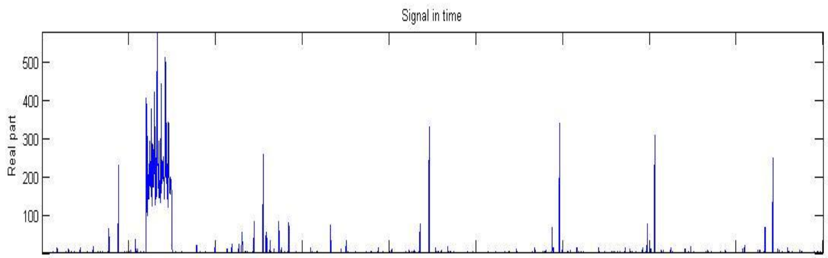
$$W_x(t,\nu) = \int_{-\infty}^{+\infty} x(t+\tau/2)\, x*(t-\tau/2)\, e^{-j2\pi\nu\tau}\, d\tau$$

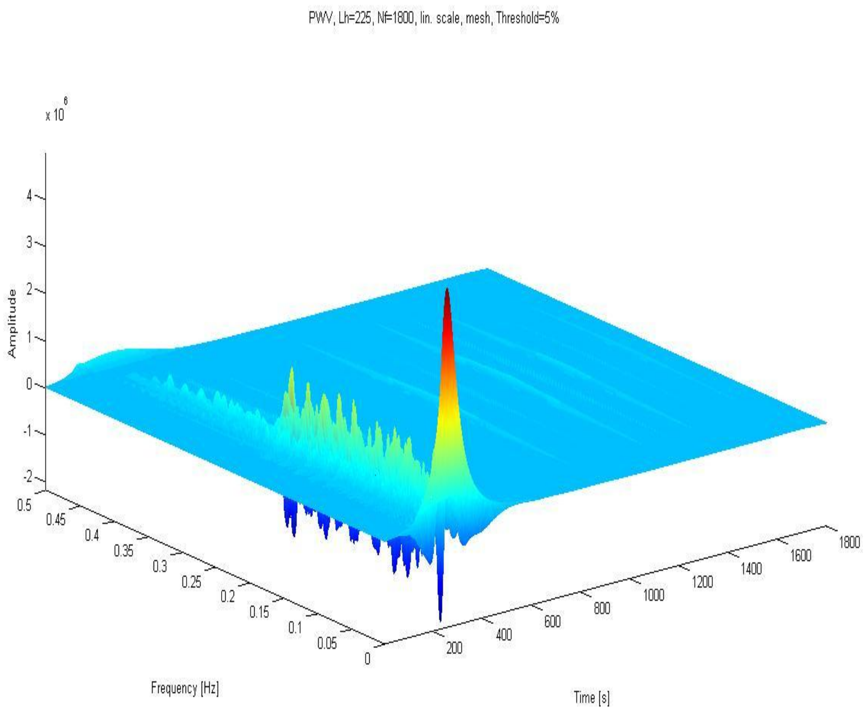
- Motivation :

- much better TF localization than atomic solutions

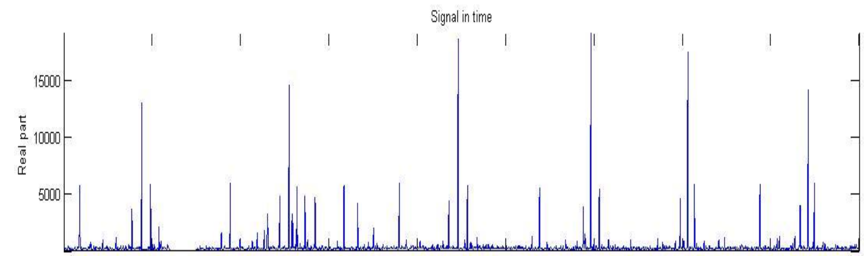
- less time costly

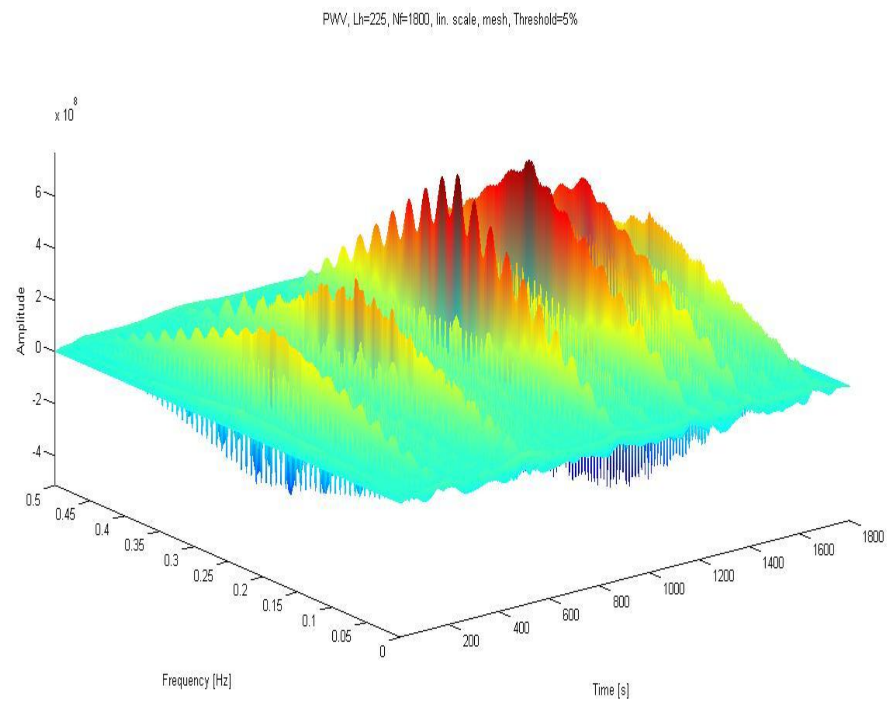
ICMP example (KEIO1)





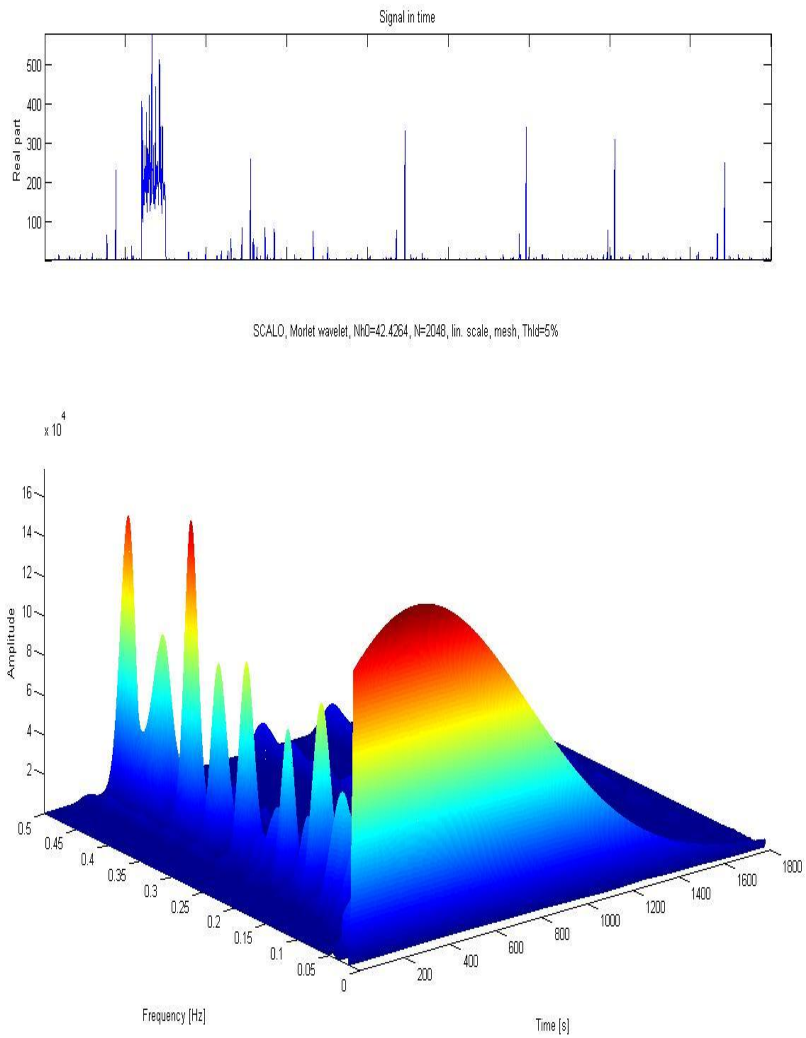
WVD_packets : Time processing cost: **2.14 sec**



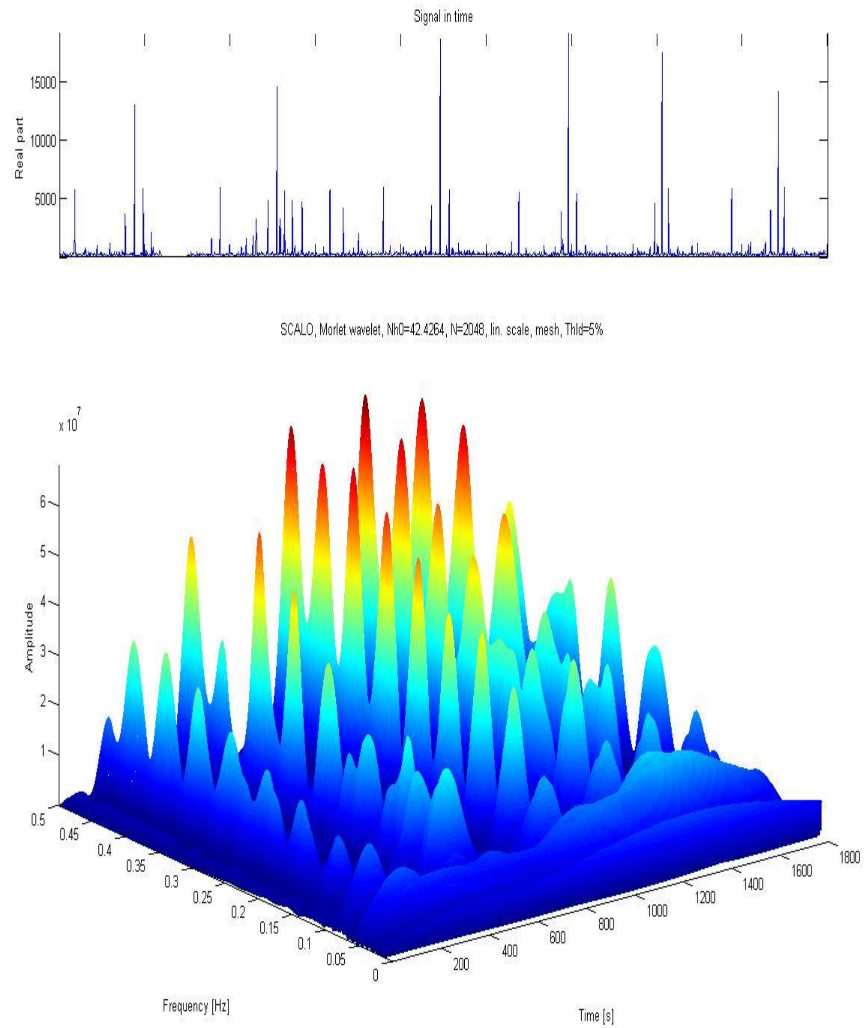


WVD_bytes: Time processing cost: **2.42 sec**

Comparison with Wavelets (KEIO1)



Morlet Wavelets_pkts : Time processing cost: **97.22 sec**



Morlet Wavelets_bytes : Time processing cost: **131.19 sec**

Conclusions

- Higher-Order Spectral analysis is a valuable and reasonably accurate tool for the demanding task of traffic modeling.
- Traffic decomposition enables tracking of protocol-specific anomalies.
- Energy distributions, in contrast to already used atomic solutions, offer a new approach consuming less processing time for detecting anomalous events making them applicable candidates for future real-time detection.

On-going & Future Work

- Extended analysis on more network traces (WIDE project).
- Investigation of energy distributions for general traffic classification.
- Refinement of scaling and smoothing factors on WVD as well as their marginal distribution properties.
- Investigation for additive noise analysis on traffic signals.
- Back-tracking validation

Thank you ☺