Extracting Knowledge from Traffic Monitoring and Analysis in the Scope of the Future Internet

David Muelas (dav.muelas@uam.es), **Jorge E. López de Vergara**

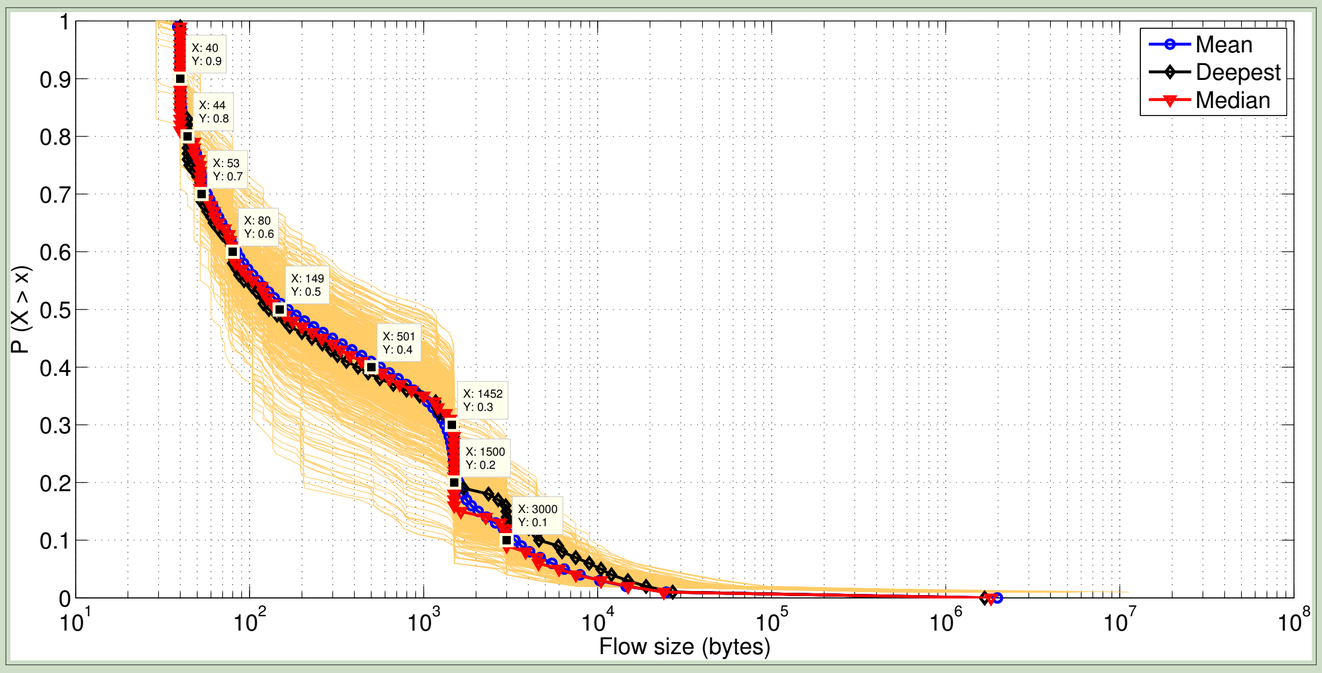
Dpto. de Tecnología Electrónica y de las Comunicaciones,
Escuela Politécnica Superior, Universidad Autónoma de Madrid

Hypothesis

Traffic Monitoring and Analysis must face new challenges and take advantage of the Future Internet elements. The advances in Data Analysis provide opportunities to lead this evolution:

- Software Defined and Virtualized Networks
- Multitenant infrastructures / Cloud
- Big Data era: growth of data and computing capacities
- More complex data-flows to extract knowledge

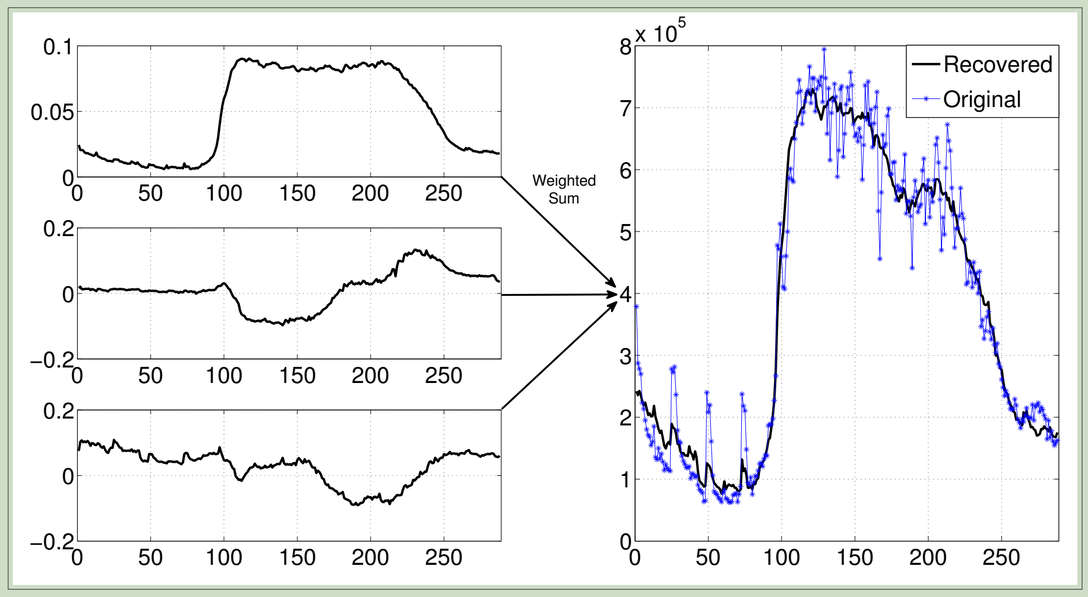
Network Data Mining Using Functional Data Analysis



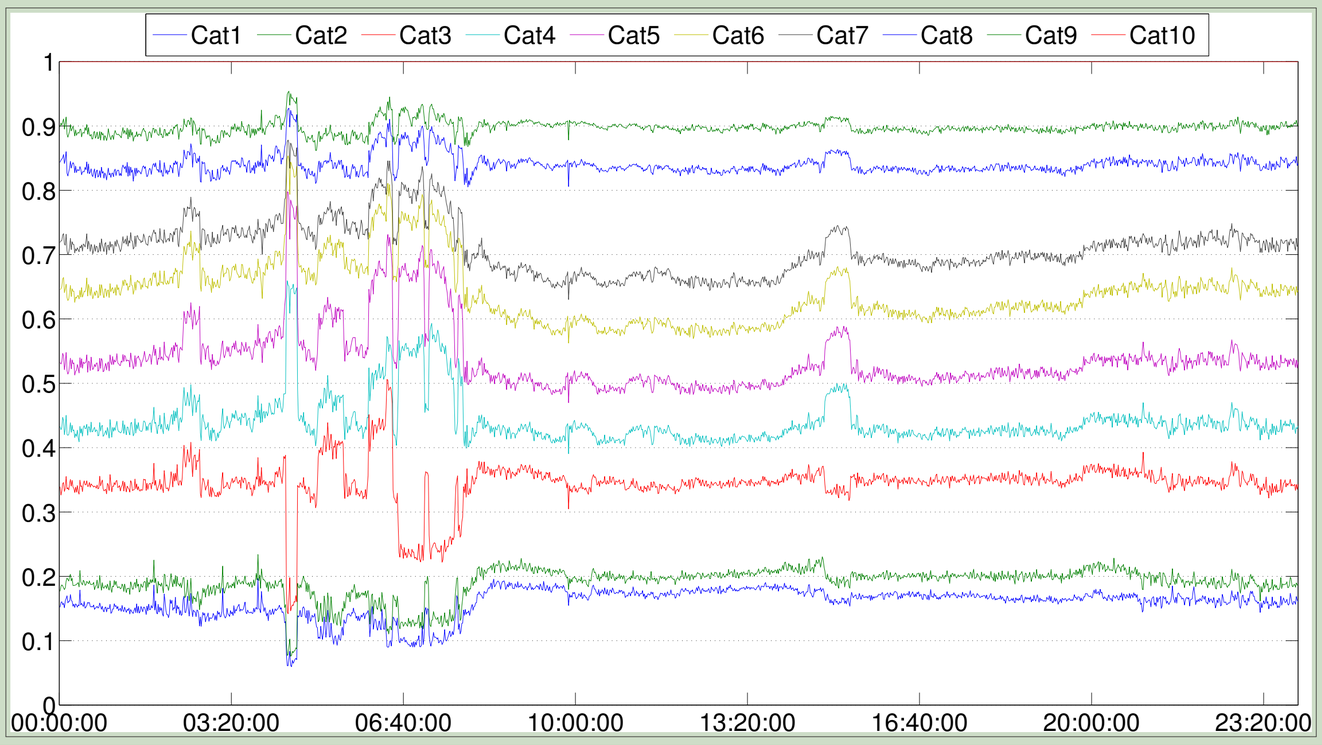
- Network management must suit the heterogeneity and flexibility of emerging technologies.
- In Functional Data Analysis (FDA), random variables are trajectories of stochastic processes.
- By considering functions that represent network state (time series, ECDFs), FDA allows the exploitation of the information in such functions:
  - Many methods do not require hypothesis on the marginal distributions of time series.
  - Analysis is not constrained only to periods where network behavior is stationary.
- Also allows robust analysis of multi-dimensional functions and derivatives of observations.

Network Data Reduction and Feature Selection

- FDA includes feature selection methods that can be applied to Traffic Monitoring and Analysis.
- Functional Principal Components Analysis is one of such methods:
  - Reduces the data volume required to represent network behavior.
  - Detects the components that best explain the variance of the network measurements.
  - Provides a space for data projection that improves classification / clustering of measurements.



Visualization of Network Flow Characteristics



- Big network flow data must be represented to extract knowledge: that is, obtain comprehensive and easy to understand visual summaries.
- The instrument: **Dictyogram** (from *δίκτυο*, network in Greek): Method to graphically trace the network flow behavior versus time. Its graphical results can be like a network electrogram, showing its vital signs.
- The basic idea is to define flow categories using the empirical distribution function of characteristics —*e.g.*, flow size.

Other ongoing research lines

- Application of FDA techniques to network behavior forecasting.
- Exploration of the virtualization of network monitoring and analysis elements.

Authors thank the concession of a Student Travel Grant (STG) that supported the participation of David Muelas in the 6th PhD School on Traffic Monitoring and Analysis