ITA - Instituto Tecnológico de Aeronáutica

A practical approach to forecast QoS considering outliers

Dr. Ilka Miloucheva

Salzburg Research

Alleassandro Anzaloni, ITA Brazil

E. Müller, SR

This research is done in the framework of European INTERMON IST project and in international research cooperation between ITA Brazil- and Salzburg Research Austria

Goal

Autoregressive integrated moving average (ARIMA) models

-> end-to-end QoS prediction in inter-domain environment

-> a preliminary detection and elimination of outliers in the time series data

-> feasibility for short and medium term forecasting (daily and monthly)

-> practical real world end-to-end delay QoS measurement scenario

-> outlook - technology usage in behind European context (Austria, Spain, Brazil,..)

Outliers

- reflect QoS behaviour due to abnormal events, anomalies and exceptions such

  -> route failures [BID 02]

  ->BGP-4 policy misconfigrations [MWA 02]

  ->convergence [NM 02] after failure

  ->route flap damping and oscillation [MGVK 02]

  ->DoS attacks [CKT 02].

-described by patterns; which do not appear to follow the characteristic distribution of the rest of the time series data data

- indicate significant fault events

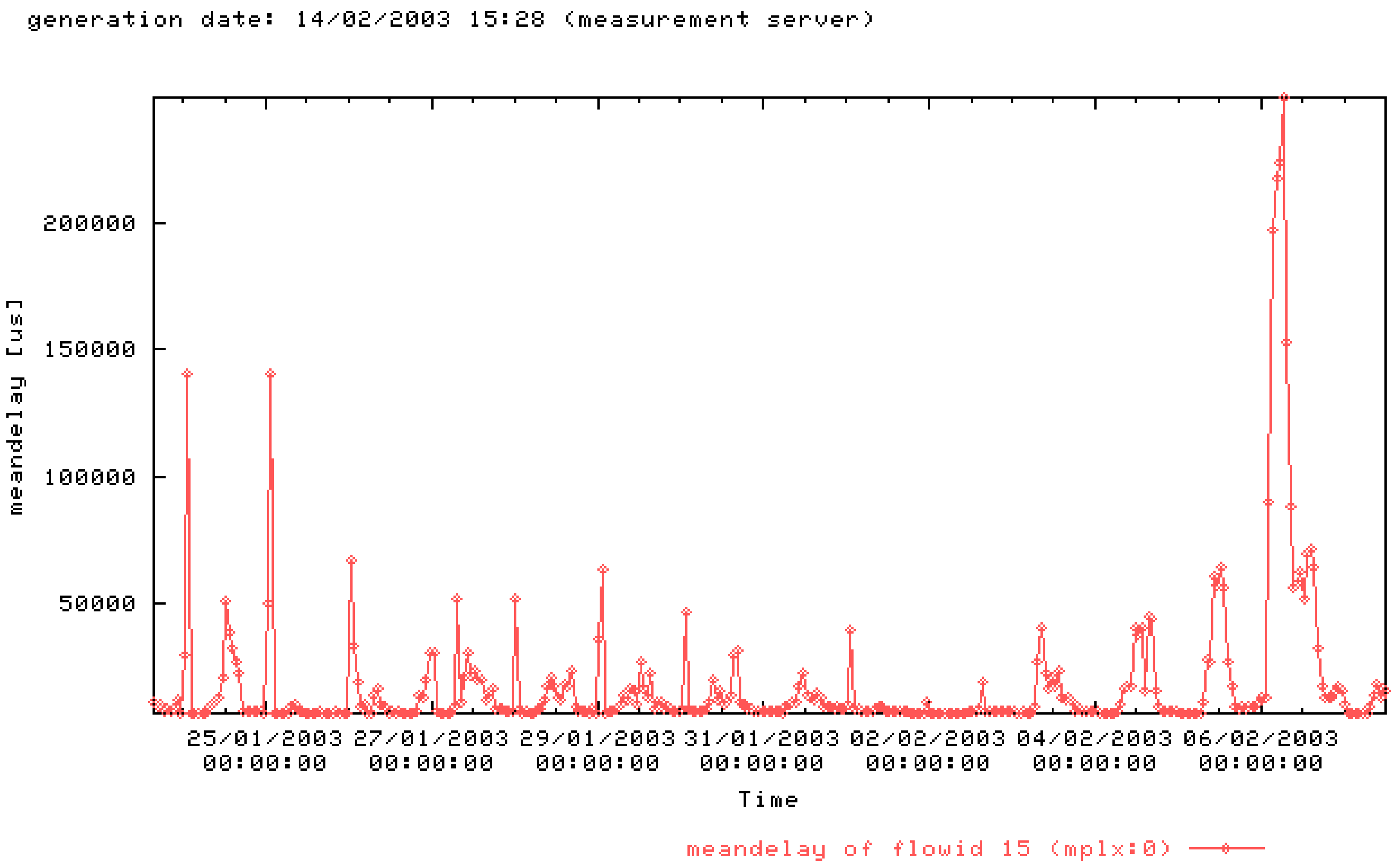
- provide useful knowledge for operation and management

ITA - Instituto Tecnológico de Aeronáutica

INTERMON

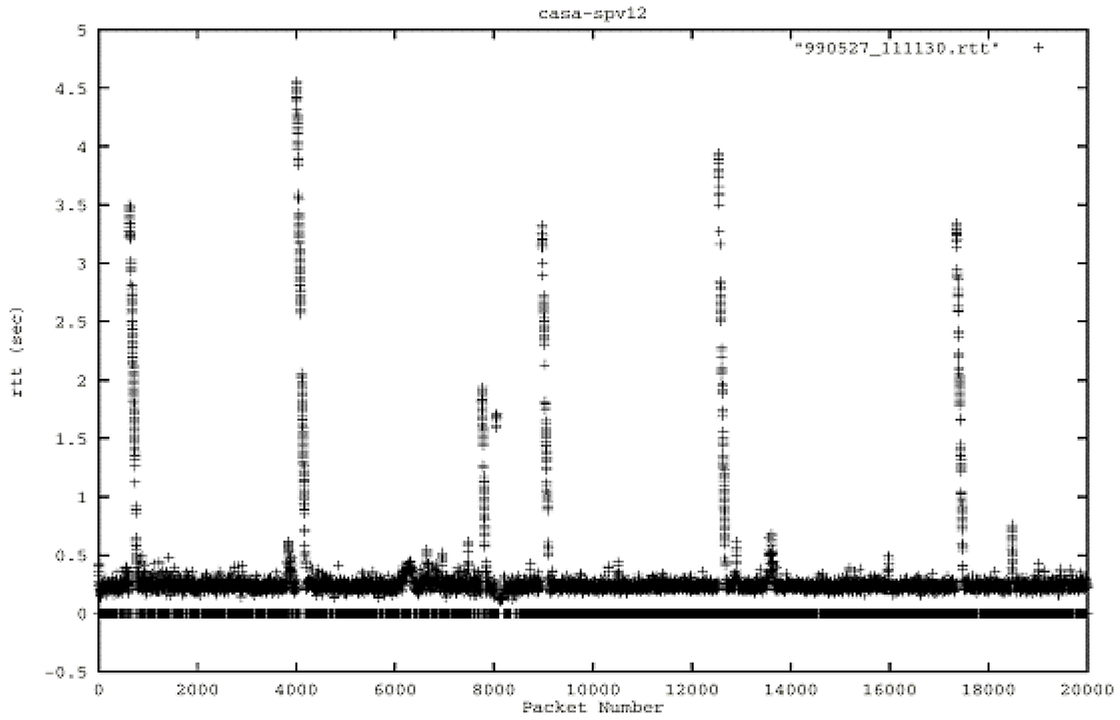
Outliers - examples

Coffee breaks outliers

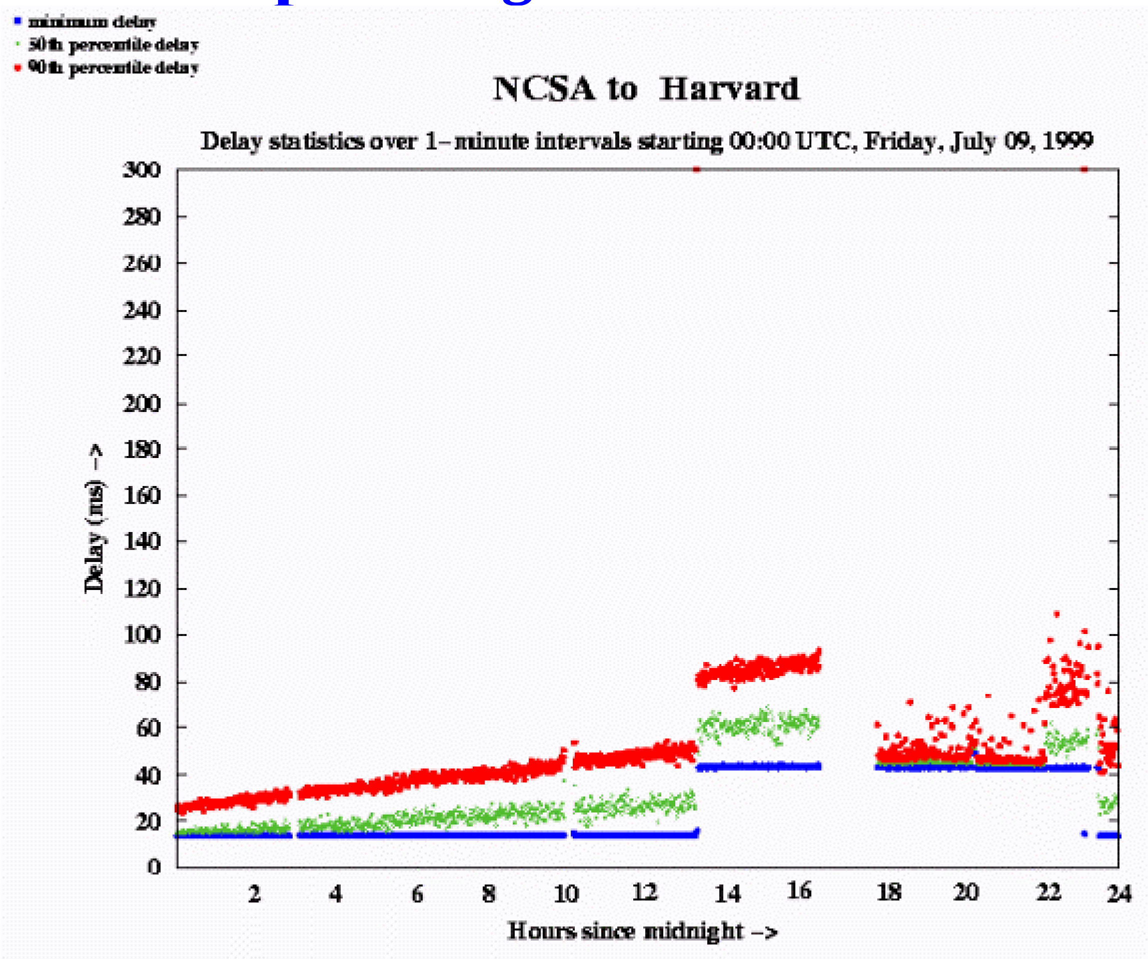


Outliers - examples

Coffee breaks outliers - [SDG 00] S. Salza, M. Draoli, C. Gaibisso, A. Laureti Palma**,** R. Puccinelli, Methods and Tools for the Objective Evaluation of Voice-over-IP Communications **,** INETProceedings, 2000**,**



Outliers - example Congestion outlier



Outliers - understanding

-----> Identification of outlier classes and their patterns dependent on abnormal events

Identification of abnormal events and their classification

↓

Impact of abnormal event on QoS behaviour – detection of outlier „pattern“, i.e. QoS

↓

Outlier identification and grouping based on the different types of „outlier“ patterns

↓

Method for enhancing forecasting models with identification and discarding of the specific „outlier patterns“

Anomalies - causes for outliers

- Grouping anomalies into different categories is useful for identification of different ”outlier” sources in order to study ”outlier” behaviour (i.e. pattern).

- Anomaly groups:

- **”Network Operation Anomalies”** - network device outages, configuration changes, traffic reaching environmental limits; distinguished by instantaneous changes in the QoS parameter values.
- **”Flash Crowd Anomalies”** - due to software release or external interest in a Web; Rapid rise in traffic flows of a particular type (e.g. FTP flows) to some destination with a gradual drop off over time.
- **”Network Abuse”** - DoS flood attacks and port; Could be observed multiple times per week; Distinct from network operation and flash crowd anomalies in that they are not always readily apparent in bit or packet rate measurements.

Forecasting Methodology ARIMA

- Box-Jenkins methodology, or Auto-Regressive Integrated Moving Average (ARIMA) modelling technique [BJR 94], [BD 02], [Pow 99] ----- > class of linear time series forecasting, i.e. linear dependency of the future value on the past values

- A long-term Traffic Prediction based on ARIMA for NSFNET Backbone [GP 94].

- Fractional ARIMA for prediction of long-range dependent traffic [Il 00]

- univariate and multivariate time series prediction of performance data describing large wide area data transfers [VSF 02].

- usage in econometrics for financial time series data [GM 96]

A practical approach for ARIMA forecasting with preliminary outlier identification

- consider outliers as extreme values which could be modified and dropped by the use of "statistical control principles"

- values that are above or below a certain range (expressed in terms of multiples of sigma, the standard deviation), i.e. they could be modified or dropped before final estimates for the ARIMA model (seasonality, etc) are computed.

- ”additive” or ”spike” outliers in the previous section.

- similarity to forecasting of co-evolving time sequence data [YSJ 00]

ARIMA general concept

-An ARIMA (p,d,q) model is an extension of a set of time series models called

- autoregressive of order p denoted by AR(p)
- moving average of order q denoted by MA(q), and
- autoregressive moving average denoted by ARMA(p,q).

ARMA model of order (p,q) is combination of AR(p) and MA(q)

- predicts the current value of the time series based on **p previous values and q previous shocks (i.e. random Gaussian noise).**

- Where the AR(p), MA(q) and ARMA(p,q) models require that the data be stationary, the ARIMA(p,d,q) model **transforms the nonstationary data into stationary series by differencing parameter d**

- ARIMA model of order (p,d,q) is simply an **ARMA (p,q) model that is differenced d times,** i.e. d is parameter giving the degree of differentiation

ARIMA definition

- ARIMA forecasts the time series value based on the weighted sum of the p previous values of the process plus q which parameter is a random Gaussian noise (white noise).

- The Gaussian noise process is characterised by zero mean and a finite variance.

- The series data $\{X_t\}$ modelled with ARIMA (p,d, q) satisfies a difference equation of the form

$$f^*(B)\, X_t \,^{o}\; f\,(B)\,(1-B)\,^{d}X_t = q(B)\, Z_t,\ \{Z_t\} \sim WN\,(0,\ s^2)$$

where $\phi^*(z)$ and $\theta(z)$ are polynomials of degree p and q, respectively and $\phi\,(z) \neq 0$ for IzI <=0. The polynomial $\phi^*(z)$ has a zero of order d at z= 1. The process $\{X_t\}$ is stationary if and only if d = 0, in which case it reduces to ARMA(p,q) process.

Outlier detection and elimination based on ARIMA model

- if the ARIMA model parameters are correct, the estimation error will follow the Gaussian distribution $\{Z_t\}$ with 0 mean and standard deviation equal to σ

- Outlier is a value that is very different from what we expected.

- Considering that the estimation error in ARIMA follows a Gaussian distribution with a standard deviation σ, than we label as **"outlier" every sample of $\{X_t\}$ that is >= 2 s away from the estimated value and replace it with estimated value considering ARIMA**

Reason ----> in a Gaussian distribution, 95% of the probability mass is within + /- 2σ from the mean.

Usage ---> QoS outliers based on sudden unexpected high values (spikes), i.e. implementation biases in the routers [PMFT 02].

Experimental forecasting software environment

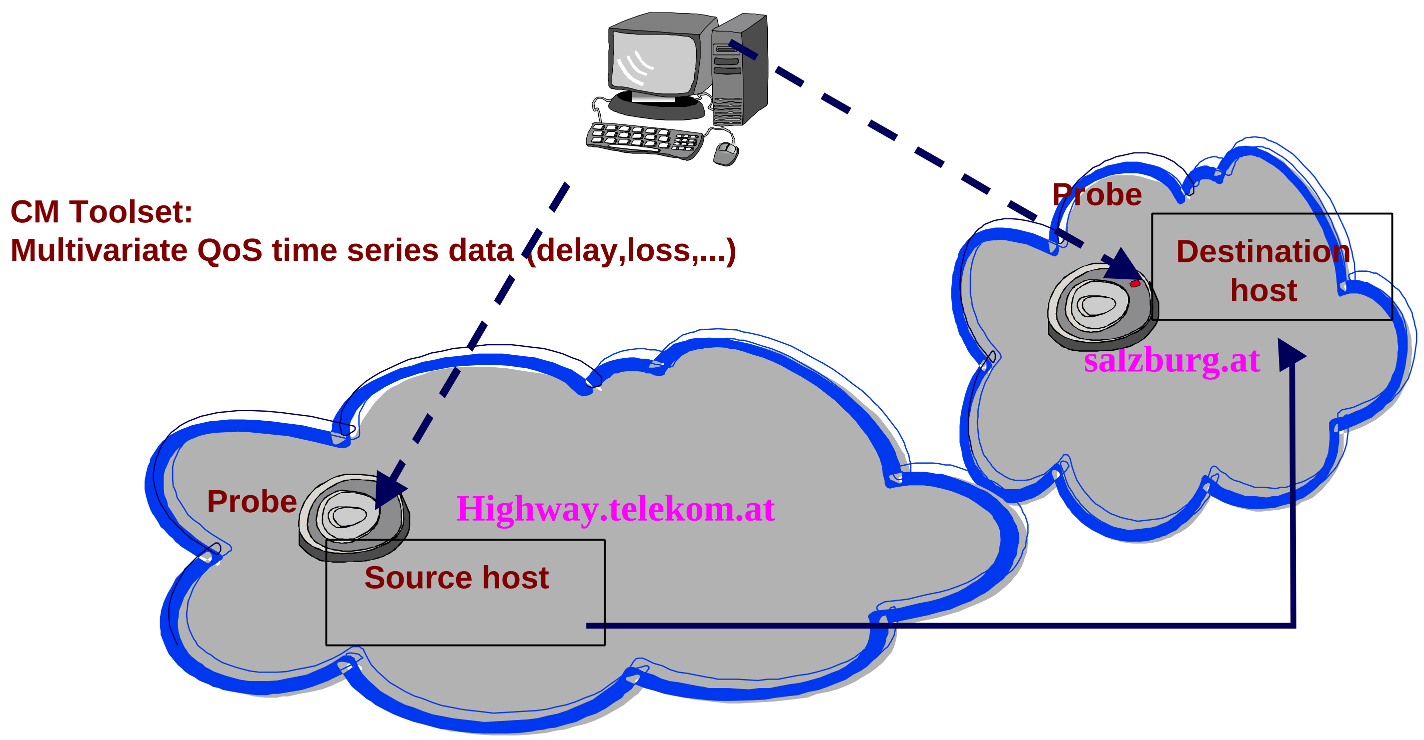
- TSW public domain software (www.bde.es) used in econometrics
- TRAMO/TERROR programs for financial purposes [MC 02] [CMS 01].

The background of this software is a methodology (and an associated program) for automatic (or manual) identification of ARIMA models, when observations may be missing and the series may be contaminated by outliers and by special effects (in particular, calendar effects).

- TRAMO ("Time Series Regression with ARIMA noise, Missing values, and Outliers") tested and proved fast and reliable [FP00].
- TRAMO's automatic model identification procedure ---> adopted by official X12 ARIMA US-Bureau of the Census program [Mon 02].

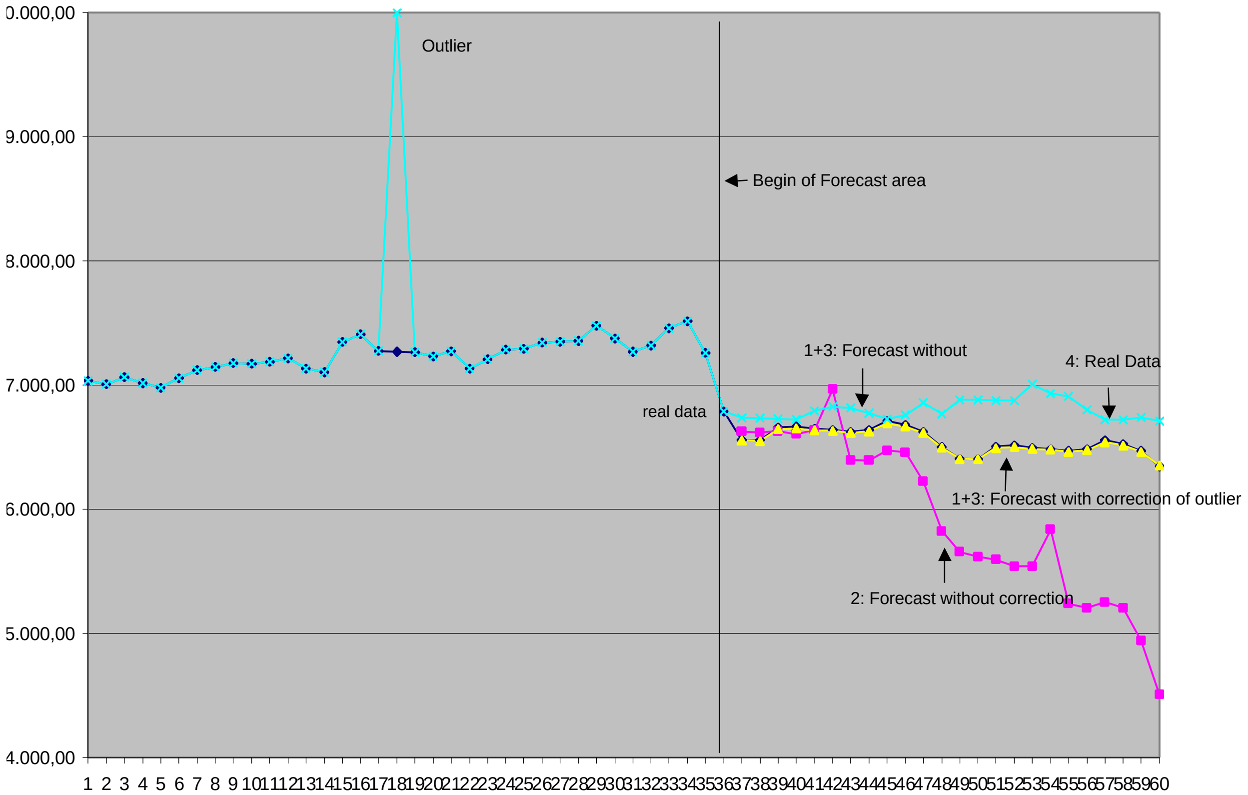
End-to-end QoS measurement scenario

- CM Toolset [AQUILA], [MH 01], [SDH 01}, [HMSP 02]. CM Toolset is an agent based monitoring tool of QoS parameters of application flows.



ITA - Instituto Tecnológico de Aeronáutica

Discussion of forecasting experiment and results



Conclusions

- Outliers could corrupt the forecasting values for the QoS parameters (delay).

- ARIMA can work in the short and mean range of QoS forecasting periods

- Integration of ARIMA forecasting considering outliers with tools for event detection and monitoring of fault events (route failure, operation anomalies, DoS attacks, misconfiguration, etc) is highly desirable to detect the source of the anomalies

Further work

- Analysis of outlier templates (patterns) that can appear in specific networking scenarios

- Understanding of sources of outliers (anomaly grouping and detection [BP 01], [SS 01])

- Study of dependencies between outliers and their sources using data mining techniques such as pattern dependency analysis [OSGC 95], [OC 96]

- New algorithms for outlier detection and elimination considering practical experiments in networking scenarios

- Study of the feasibility of new forecasting techniques for QoS prediction in networking scenarios like non-linear forecasting methods [BD 02], [CE 92]