

Data mining approach to study Quality of Voice over IP applications

I. Miloucheva¹, D. Hetzer² & A. Nassri³

¹*Salzburg Research Austria*

²*T-Systems Germany*

³*Siemens Austria*

Abstract

Patterns focus on “interesting” structures in data behaviour, which can be abstracted and used for efficient understanding and analysis. This paper presents a data mining technology aimed to assess characteristics of network connections for Voice over IP (VoIP) communication based on analysis of structures of Quality of Service (QoS) parameter time series data (end-to-end delay and packet loss). The data mining approach for QoS parameter data is designed and developed in the framework of EU IST INTERMON project [1]. It is used for spatio-temporal analysis in the large scale inter-domain Internet.

We propose automated techniques to study network QoS parameter patterns for VoIP communication, such as:

- “Outliers” defining QoS parameter values exceeding “thresholds” corresponding to the rating factor R of the ITU-T E-model.
- End-to-end delay structures, used for appropriate choice of playback delay adjustment algorithms.

Based on patterns, the network administrator can automatically evaluate the ability of network connections for VoIP communication, and select appropriate parameters for playback algorithms. The discussed data mining facilities for the QoS parameter study for VoIP applications are shown using real measurement data for inter-domain Internet connections collected in the INTERMON data base.

Keywords: data mining, pattern, QoS parameter, end-to-end delay, packet loss, VoIP, outlier.



1 Introduction

Efficient QoS support is a challenge of today large scale inter-domain Internet. For this purpose better understanding of large sets of QoS parameter measurement data obtained for different network connections and automated integration of QoS parameter data for management decisions is required.

To automate the analysis of QoS parameter data for different kind of applications in large scale Internet, we propose a technology based on:

- data mining interfaces to study network QoS parameters data considering demands of specific quality based application classes, such as VoIP, streaming multimedia and Grid.
- exploration of QoS parameter data sets for applications by automated processing of useful patterns.

This paper is aimed to present data mining user interface and pattern concept for efficient analysis of VoIP quality in large scale inter-domain Internet. It is based particularly on the work of INTERMON project [1] developing architecture for QoS analysis in inter-domain environment using integrated monitoring, modelling and simulation tools with common data bases for study of topology, traffic and QoS parameters. The general facilities of spatio-temporal QoS pattern analysis in INTERMON, as well as their interaction with the inter-domain topology and routing discovery, to obtain impact of routing and failures on QoS based applications like VoIP using patterns, is discussed in [2] and [3].

We present a pattern analysis technology of network QoS parameters for VoIP which supports efficient mining and understanding of complex network delay characteristics (patterns) of network connections for VoIP communication, using "heuristics" for playback delay adjustment algorithms [4]. It addresses the needs of ISP providers to support emerging standards like E-Model [7] by using of outlier patterns. We describe the theoretical background of pattern analysis for VoIP quality study and the automated processing of network QoS parameter data integrated into a VoIP Quality analyser. We give an example for using of the technology based on real measurement data.

2 Patterns of QoS parameter data for VoIP Quality Analysis

Patterns focus on "interesting" structures in data behaviour, which can be abstracted and used for efficient understanding, knowledge discovery and processing [2], [4], [5], [6]. Patterns are defined dependent on the area, where they are applied. Patterns like "Head and Shoulders" and "Symmetric Triangle" are useful technique for describing behaviour of stock prices. In ECG (electrocardiogram), the "Heart Beat" pattern is a known procedure.

In the area of management of QoS in large scale Internet, we studied patterns aimed to describe network QoS parameters of Internet connections useful to study quality demands and tuning of applications.



2.1 Introduction to QoS parameter time series data

The IETF Performance Measurement working group (IPPM) has defined a set of performance metrics, sampling techniques and associated statistics for transport and connectivity layer measurements. In this paper, we refer to end-to-end one way delay [8] and packet loss metrics [9].

QoS parameter values could be aggregated for specific intervals obtaining as result new time series data of mean, maximum and minimum QoS parameter values, describing the intervals. The resulting QoS parameter sequences build aggregated QoS parameter sequences which differ from the measured row QoS parameter data. Dependent on the data mining goal, patterns based on row or aggregated QoS parameter data could be useful. Row QoS parameters are useful for fault detection, because they show more exactly the structure of QoS parameter behaviour, but use significant more storage and processing overhead. Aggregated QoS parameter analysis is used for comparison and forecasting in the area of planning.

2.2 Basic and composite patterns of time series data of QoS parameters

From a mathematical point of view, the time series data of QoS parameter values (row or aggregate) is mapped by the patterns into a simple approximated process with plain and increase/decrease phases. We distinguish between basic and composite patterns [2], [4]. Basic patterns, which can be used to study VoIP communication, are “increase/decrease”, “plain”, “outlier” and “spike”. The composite patterns consist of sequences of basic patterns.

Increase and decrease patterns represent continuous increase/decrease in the data values of the given sequence. Let the QoS values $x_1, \dots, x_i, \dots, x_n$ belong to $\{X_i\}$, which is an increase basic pattern with

$$0 = x_1 = x_{\min} < \dots x_i < \dots < x_n = x_{\max}$$

and decrease basic pattern with

$$x_{\max} = x_1 > \dots x_i > \dots > x_n = x_{\min} > 0.$$

For each neighbour x_i, x_j hold that

$$\text{diff} = |x_i - x_j| \geq d$$

E.g. the difference between neighbour values, should be greater, than some minimum value d , which is chosen as interval base to detect plain patterns.

The gradient of increase/decrease pattern defines the rate for rise/fall of the values in the given time interval. The gradient m of an increase or a decrease pattern is defined as:

$$m = \Delta x / n, \text{ where}$$

$$\Delta x = x_n - x_1 \text{ is the difference between maximum and minimum value,}$$

n is the number of data in the time series sequence.

The goal is, based on gradient evaluation, to obtain information, how efficient VoIP services is supported by the measured QoS parameters of the connection.

Plain pattern is a sequence of time series data with same or similar QoS values ranging within a given interval. Let $\{X_i\}$ be time series data consisting of QoS parameter values. With $x_1, \dots, x_i, \dots, x_n$, belonging to $\{X_i\}$, the QoS values of a plain pattern structure is given, for which hold that



$$\text{diff} = |x_i - x_j| < d.$$

For each neighbour pair x_i, x_j , the difference between neighbour values should be smaller than given coefficient d . The plain pattern is typical for network connections with small traffic variations and is appropriate for VoIP traffic.

Outliers are patterns which show abnormal behaviour [10].

Spikes are sudden increases in time series data values, which are of small duration Δt . Spikes could influence the VoIP quality and playback algorithm.

2.3 Network patterns for VoIP quality analysis

We focus on two kinds of network patterns useful for study of VoIP quality:

- Outliers for VoIP quality analysis
- End-to-end delay patterns for choice of playback algorithm and parameters.

2.3.1 Outliers for VoIP quality analysis

Outliers are abnormal QoS parameter values, which are detected based on some model such as ARIMA or in respect to some application dependent threshold [10]. The E-model is used in our work to select the threshold values for automated detection and analysis of network delay and packet loss “outliers” for VoIP, i.e. all these QoS parameter data, which exceeds the threshold level corresponding to the given R-factor.

For VoIP quality analysis, we define a QoS parameter outlier based on the given level of R-factor of the E-model and corresponding threshold value d_{thr_m} [7]. A sequence of $x_1, \dots, x_i, \dots, x_n$, belonging to the time series data $\{X_t\}$ build an “outlier” related to the threshold value d_{thr_m} , when for all $x_i, i = 1..n$, hold that $x_i > d_{\text{thr}_m}$.

Outlier as pattern P_i is analysed by:

- Occurrence and frequency of “outlier” sequences in a given time interval and for a given period (daily, monthly), i.e. the number of observations of “outlier” patterns $P_i, i = 1..n$, related to the time interval Δt :

$$F = \sum P_i / \Delta t.$$
- Relative length of all outliers segments (i.e. duration) in the observed interval Δt and per period Δt_p .

$$L = \sum L_i / \Delta t, \text{ where } L_i \text{ is the duration of the outlier pattern } P_i.$$
- Histogram of outlier values
- Basic and composite structures of decrease, increase, plain and spikes pattern types.

E-model, recommended by the Telecommunication industry associations (e.g. [7], [13]), is a computational model standardised by ITU-T using R-factor for prediction of VoIP quality based on “impairment” parameters such as packet loss and delay transmission parameters to predict the subjective quality of voice. Impairments, due to codecs, delay and percent packet loss, are defined in ITU-T Recommendation G.113, Transmission impairments, Appendix I ([13]).

The impact of the packet loss on the VoIP quality depends on the end-point’s playback algorithm intelligence and encoding scheme. Practical research and experiences have discussed impairment of end-to-end delay and packet loss, on R-factor for different encoding and conversation types (e.g. [11], [12]).



2.3.2 Network delay patterns for selection of VoIP playback algorithm

For VoIP quality tuning, heuristics for playback delay adjustment are defined according network QoS parameter structures. The role of end-to-end delay variation for adjustment of playback buffers is considered in a number of playback algorithms based on exponential average, fast exponential average and spikes adaptation [14],[15],[16],[17]. Exponential average algorithms to adjust the playback delay, are based on formulas using parameter α :

$$\text{Delay change} \rightarrow d_{\Delta} = \alpha \cdot d_{\Delta(i-1)} + (1 - \alpha) \cdot n_i$$

[14], [16] studied different values for α and observed that:

- Higher values of α are good only in the situations where network conditions are stable, e.g. slow delay and jitter increase.
- Smaller values of α are appropriate when network conditions are changing rapidly (sudden increase/decrease in delay).

Our approach is based on increase and decrease patterns with specified threshold gradient which show slow and rapidly changing end-to-end delay, which is useful for selection of parameter α .

Further we consider “spikes”, e.g. sudden and large increase in the values of end-to-end delay for tuning of playback delay. The aim is to handle short-lived network “spikes” dependent on their duration. [15] proposes algorithms to react more effectively to delay “spikes”. During the “spike”, the first packet of the spike is used as the playback delay. After the “spike”, the playback delay is chosen considering the delay distribution of the last received N packets. [17] discusses the learning of “spikes” for playback delay adjustment focussing on their values and frequency of occurrence above a certain threshold.

We detect similar patterns of end-to-end delay including “spikes” based on analysing of histograms of “spike” values, duration of “spike” patterns, distance between “spike” patterns and frequency of “spike” patterns.

2.4 Techniques for pattern detection

The patterns of interest are detected by modelling approaches and algorithms [5]. Data mining rules are used to relate the observed process with the obtained patterns dependent on the context. There are different modelling approaches and algorithms to find data mining rules for pattern detection, for instance usage of:

- Algorithms based on knowledge of internal system processes
- Learning algorithms e.g. neural networks to discover “hidden” patterns.

The data mining technology for VoIP Quality analysis is based on algorithms to process models of basic, composite patterns and outlier patterns as discussed in the previous sections. Data mining rules for detection and similarity analysis of patterns are defined according the experiences of the network administrator, who is responsible for setting of:

- parameters for increase, decrease and plain pattern detection,
- tables for VoIP quality thresholds according the R-factor of the E-Model
- parameters for spikes and gradient based analysis to detect composite patterns appropriate for the specific playback algorithm.



For similarity detection of increase, decrease and plain patterns, we developed algorithms based on the linear approximation [18], [19]. Piecewise linear segmentation, which attempts to model the data as sequences of straight lines, has advantages as a representation and similarity search of long-term QoS parameter Time Series Data (TSD). [19] point out that it provides a useful form of data compression and noise filtering.

We use linear approximation to optimise the trade-off between high level of aggregation of measurement results and low level of information loss for accurate gradient estimation of QoS parameter structures. The algorithms are configured by parameter d , called abstraction coefficient, depending on the user preference for structure details of the QoS parameter patterns. Parameter d considers differences of neighboured ordered TSD to approximate values.

3 Data mining technology for VoIP Quality Analysis

The study of connection QoS parameters for VoIP communication using patterns requires an efficient data mining user interfaces related to a measurement infrastructure. The tool for automated pattern detection and similarity analysis is using QoS parameter data base of CMToolset [4] developed for active end-to-end QoS monitoring and topology discovery in INTERMON.

The QoS parameter data base collects measurement data for scenarios of emulated VoIP traffic. Pattern analysis is based on raw and aggregated QoS parameter data. A dedicated data base is used to store detected patterns. The following figure shows the integrated technology for VoIP Quality analysis:

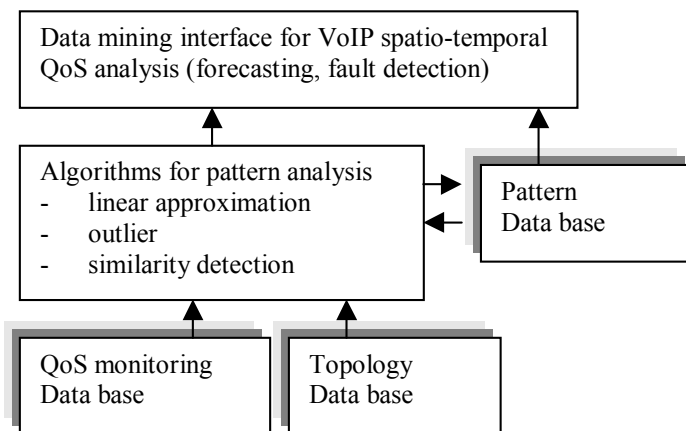


Figure 1: Technology for analysis of network QoS for VoIP traffic.

The data mining user interface for VoIP analysis is shown in figure 2.

The VoIP traffic emulation [20] is based on:

- Call model: long distance calls, business calls.
- Modelling of call arrival strategies.
- Processing encoding technique: G.711, G.723.1, G.723.2, G.728 and G.729.

The CMToolset integrates traffic generation functions using on-off patterns to model talk-spurts and silence gaps of VoIP traffic which we use in the scenarios.

The data mining functions integrated in VoIP Quality Analyser allows to:

- Define interval of times, i.e. segments, as well as network connections, to obtain basic and composite patterns of QoS parameters
- Compare patterns based on their specification
- Study frequency of pattern in a given time interval and per specific period.
- Learn dependencies of patterns, considering multivariate time series data patterns and their temporal and spatial relationships.

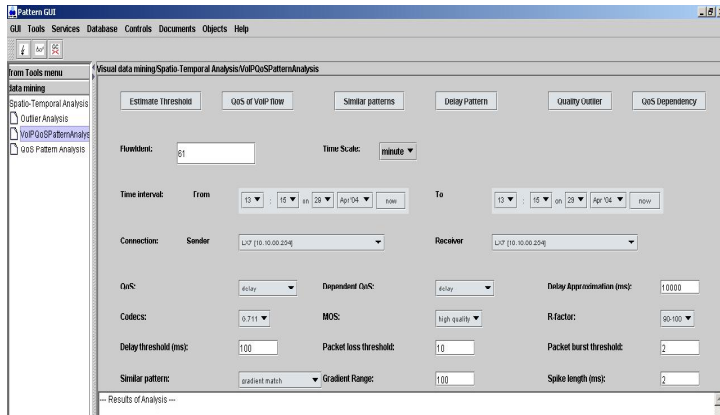


Figure 2: Data mining tool of network QoS pattern analysis for VoIP.

4 Scenarios for study of “outliers” impacting VoIP Quality

The developed technology was used to study inter-domain QoS parameter patterns of European network connections for VoIP.

Analysis of QoS patterns of the inter-domain connection Madrid – Salzburg is based on emulation of VoIP packets, 64 Bytes long, sent in 20 ms intervals.

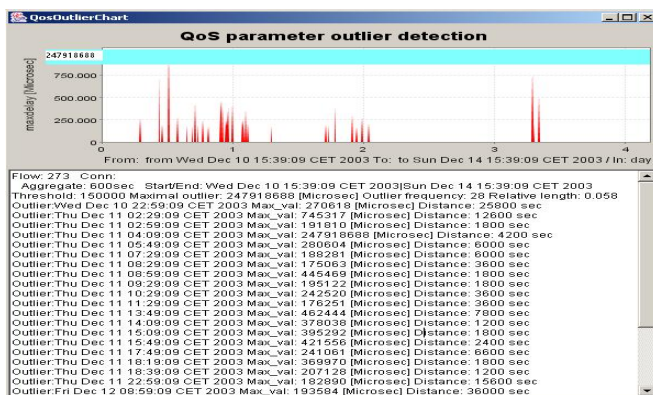


Figure 3: Daily end-to-end delay outlier patterns of aggregated data.



The next figure gives the end-to-end delay “outliers” of emulated VoIP traffic, considering threshold of 150 ms and aggregated QoS parameter data (maximum delay) in time scales of 10 minutes.

We use “multiple scale” pattern presentation, because of the great difference in the values between typical outliers of 150-1000 ms, and “extreme” outliers. The detection of the structure of outliers found in these four days shows surprisingly in many cases strong “decreasing” patterns for outliers with “extreme” values. This is shown in the next figure.

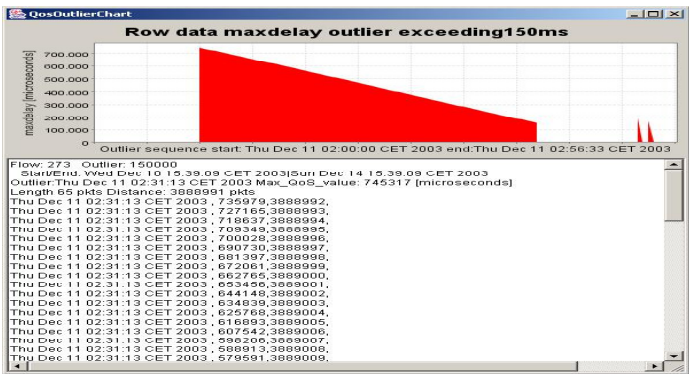


Figure 4: “Decreasing” end-to-end delay pattern of “extreme” outlier.

An outlier, starting Thu Dec 11 02:31:13 CET 2003 is shown, reaching 745317 microseconds as highest value with duration of 65 samples presenting a strong “decrease” pattern. The same pattern has the “highest” outlier starting Thu Dec 11 02:31:13 CET 2003 with maximal value of 247918684 microseconds including 87 samples with strong “decreasing” values.

Multivariate QoS parameter patterns [6] are important for cause detection.

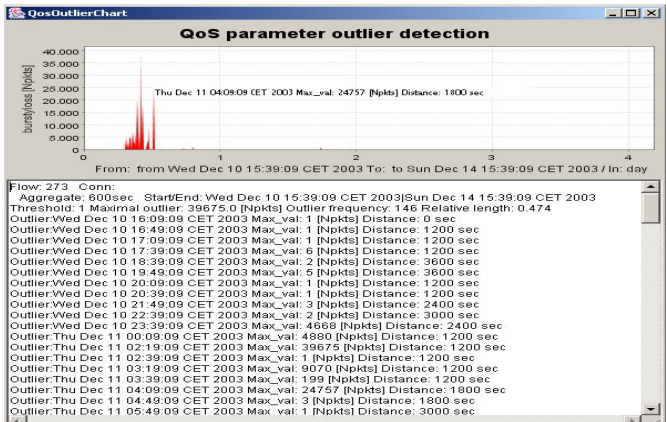


Figure 5: Burst packet loss patterns of inter-domain connection.



For instance, packet loss burst considered in the same time of end-to-end delay measurements, could give more insight in the dependencies of “outliers”. The length of the “highest” outlier (duration, x-axis) is negligible (860 ms); but it is immediately followed by a significant packet loss burst of 24.757 Pkts which is shown in the next figure:

The duration of the maximum packet loss burst is comparable with the values of end-to-end outlier. A possible cause for the highest “extreme” outlier pattern is stopping of transmission. Comparing different multivariate data such as traffic and topology, causes for patterns could be detected and classified. Data mining rules could be defined to find similar patterns of multivariate QoS parameters such as “end-to-end delay outlier followed immediately by burst packet loss”.

Data mining rules could be also defined to obtain appropriate parameters for playback adjustment algorithms based on analysis of end-to-end delay patterns using approximations. The following figure shows the end-to-end delay patterns based on linear approximation parameter of 20 ms.

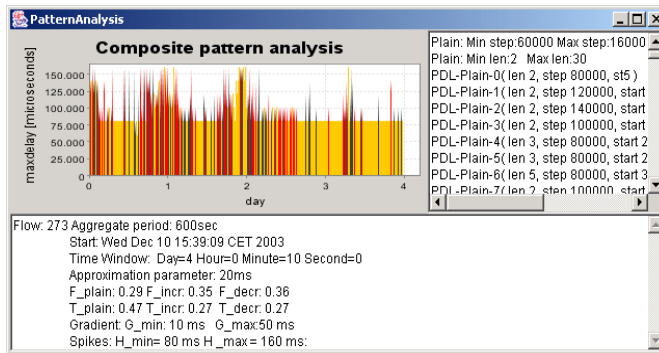


Figure 6: End-to-end delay patterns based on approximation.

The linear approximation in figure 6 allows more efficiently to select the phases of the playback buffer adjustment considering prediction of gradients of increase and decrease patterns as well as “spike” occurrence and long plain patterns.

5 Conclusions

This work presents a new approach for evaluating the QoS parameter of Internet connections for VoIP applications using automated detection of patterns impacting quality and parameters of VoIP application. This approach, integrated in the INTERMON toolkit, allows:

- Mapping of network delay and packet loss “outliers” to the R-factor demands for VoIP quality.
- Network delay patterns for selection of playback delay adjustment.
- Detecting pattern causes based on inter-domain routing and traffic.



Further research is aimed to use the VoIP quality analysis based on patterns of real measurements. The initiated work on study of inter-domain connections between Madrid (Spain) and Salzburg (Austria) has shown interesting “outlier” patterns and dependencies of them. Continuing this research, we want to focus on revealing of “hidden” QoS parameter patterns and similarity search in order to show multivariate dependencies of QoS parameter values and networking events [5], [6]. For integration of the VoIP technology in operational management, more efficient data base design to reduce filtering time and multiple-scale visualisation will be studied for improvement of the current technology.

References

- [1] INTERMON - Advanced architecture for INTER-domain quality of service MONitoring, modelling and visualisation, <http://www.ist-intermon.org>.
- [2] I. Miloucheva, U. Hofmann, P.A.A. Gutiérrez, Spatio-temporal Analysis in large scale Internet environment, MIPS Workshop, Napoli, Italy, Nov., 2003.
- [3] P.A.A. Gutierrez, I. Miloucheva, Integrating Inter-domain Routing Analysis in Novel Management Strategies, NEWAN Conference, February, 2004.
- [4] I. Miloucheva, A. Nassri, A. Anzaloni, "Automated analysis of network QoS parameters for Voice over IP applications", Inter-Domain Performance and Simulation Workshop, Budapest, Hungary, 22-23 March, 2004.
- [5] D. Hand, H. Mannila, P. Smith, Principles of Data Mining, Massachutes, MIT Press, 2001.
- [6] F. Höppner, Learning dependencies in Multivariate Time Series, European Conference on Artificial Intelligence, Lyon, July 2002.
- [7] ITU-T Recommendation G.107, The E-Model, a computational model for use in transmitting planning, December 1998.
- [8] G. Almes, S. Kalidindi, M. Zekauskas: A One-way Packet Loss Metric for IPPM. RFC 2680, September, 1999.
- [9] G. Almes, S. Kalidindi, M. Zekauskas: A One-way Delay Metric for IPPM. RFC 2679, September, 1999.
- [10] I. Miloucheva, A. Anzaloni, E. Müller, A practical approach to forecast Quality of Service parameters considering outliers, Inter-domain performance and simulation Workshop, Salzburg, 20-21 February, 2003.
- [11] R.G. Cole, J.H. Rosenbluth, Voice over IP Performance Monitoring, ACM SIGCOMM, Computer Communication Review, vol.31, April, 2001.
- [12] J. Tyre, R. Britt, Performance Characteristics of Voice over IP Networks, NORTEL, <http://qos.ca.nortel.com/apnnotes>, 2001.
- [13] Provisional planning values for the equipment impairment factor I_e , ITU-T Recommendation G.113, Appendix I, 1999.



- [14] M. Narbutt, L. Murphy, VoIP Playout Buffer Adjustment using Adaptive Estimation of Network Delays, Proc. of the 18th International Teletraffic Congress (ITC-18), p. 1171-1180, Berlin, Germany, Sept. 2003.
- [15] S. Moon, J. Kurose, D. Towsley, "Packet audio playout delay adjustment: performance bounds and algorithms", ACM/Springer Multimedia Systems, vol. 6, pp. 17-28, Jan. 1998.
- [16] M. Narbutt, L. Murphy, "Adaptive Playout Buffering for Audio/Video Transmission over the Internet" 17th UK Teletraffic Symposium, May, 2001.
- [17] F. Tobagi, A. Markopoulou and M. Karam, "Is the Internet ready for VoIP?", in Proc. IWDC, Capri, Italy, Sept., p. 49-57, 2002.
- [18] E. J. Keogh, S. Chu, D. Hart, M. J. Pazzani, Segmenting Time Series: A Survey and Novel Approach, Data mining in Time Series Databases, 2003.
- [19] E. Keogh, M. Pazzani, An enhanced representation of time series. 4th Conference on Knowledge Discovery and Data Mining. New York, 1998.
- [20] W. Jiang, H. Schulzrinne, Assessment of VoIP Service Availability in the Current Internet, Passive and Active Measurement Workshop, April, 2003.

