Spatio-Temporal QoS Pattern Analysis in large scale Internet environment

I.Miloucheva[1], U. Hofmann[1], P.A.Aranda Gutiérrez[2]

[1]Salzburg Research, 5020 Salzburg, Austria
{ilka.miloucheva, ulrich.hofmann}@salzburgresearch.at
[2]Telefónica I+D, Madrid, Spain
paag@tid.es

Abstract. For enhanced Quality of Service (QoS) provision of multimedia applications in Internet environment, there is a need of data mining tools supporting the automated analysis of QoS behaviour and dependencies for the purpose of modelling and forecasting, QoS planning and anomaly detection. This paper presents a data mining technology for spatio-temporal QoS pattern analysis including automated extraction and description, similarity matching and dependency maintenance of patterns in telecommunication networks. The technology is based on selection and definition of patterns from measured time series data sequences of QoS parameters using an appropriate pattern description language, pattern matching algorithms with different options for pattern similarity analysis and data mining interface supporting the network engineer in QoS pattern analysis considering temporal and spatial constraints. The building of an appropriate archive of detected similar patterns in given spatio-temporal context, i.e. pattern analysis data base, is directly concerned with for solving of specific QoS data mining task. The described technology is developed in the framework of INTERMON project for integrated QoS analysis in a large scale inter-domain environment based on interaction with monitoring, topology discovery and traffic analysis tools.

Keywords : QoS, pattern, data mining, time series data, spatio-temporal analysis

1 Introduction

Pattern analysis is a particular data mining method based on heuristics to study, interpret, model and forecast behaviour and dependencies of time series data using patterns (e.g. chart structures) [1]. Patterns like “Head and Shoulders” and `Symmetric Triangle“ are useful technique for describing behaviour of stock prices [2]. In ECG (electrocardiogram), the detection of “Heart Beat” pattern is a known approach. Pattern based data mining tools were developed for different areas, as example variable time boxes [5], cluster based pattern analysis tool [6], Query Scetch for graph centric queries [7]. Although the significance of patterns as data mining technique for understanding traffic and QoS is good understood and addressed in current telecommunication research [3], there is a lack of comparable tools for automated pattern engineering. New technologies for automation of pattern detection, similarity analysis and management of significant patterns are required to increase the efficiency of QoS data mining in solving specific network engineering problems. To support the network engineer in the tasks of QoS data mining, a spatio-temporal QoS pattern analysis tool for automated processing of patterns extracted from QoS measurement data bases is developed in the framework of INTERMON project [4]. The particular focus of the developed data mining technology is to obtain patterns of time series data describing QoS parameters, e.g. generic QoS patterns, useful for QoS planning, performance study and anomaly detection in large scale networking environment.

This paper presents the developed technology for spatio-temporal pattern analysis:

- Pattern definition technique using Pattern Description Language and basic pattern types describing QoS parameter behaviour
- Pattern matching algorithms processing different options for pattern analysis useful for solving of data mining tasks in large scale network environment
- Data base for archiving of patterns and their relationships
- Data mining user interface supporting the network engineer in automated pattern detection, storage and similarity analysis in spatio-temporal context.

The automation of pattern detection, modelling and similarity search supports the network engineering in the process of interpretation of patterns and pattern based modelling. The technology is applied in the framework of INTERMON architecture [4], [9],[15],[16],[17],[18],[19] for integrated QoS data mining based on interaction with tools aimed at QoS measurement, connectivity monitoring, inter-domain routing analysis, traffic flow measurement and traffic matrix exploration.

This paper is organized as follows. Section 2 presents the pattern modelling approach using the pattern description language. The framework of basic types of patterns and pattern composition are explained. Examples for interpretation of basic pattern types for solving of specific data mining tasks are given. Design and facilities integrated in the spatio-temporal pattern analyser – architecture, pattern matching algorithms and data base for pattern analysis are explained in section 3. Usage of the spatio-temporal pattern analyser in the INTERMON architecture for integrated QoS data mining based on the interaction with monitoring and topology discovery tools is addressed in section 4. Section 5 concludes this paper and outlines further work.

2 QoS Pattern Description

2.1 Definition of QoS patterns in telecommunication networks

Pattern is a general term for any recognisable regularity (structure) in the data [1]. The data mining community has developed a wide variety of techniques for algorithmically extracting useful patterns from time series data and detecting pattern similarities in combination with specific applications. Spatio-temporal aspects of patterns are studied using models derived from application context [10].

This paper is focussed on patterns obtained from measured time series data sequences (QoS, traffic) in telecommunication networks. Considering the current state-of-the-art, QoS patterns are used to describe by visual representation (charts) some specific

behaviour of measured QoS parameters. „Outlier" is for instance a pattern, describing an abnormal QoS behaviour caused by anomaly events and routing problems [9]. Delay "outlier" due to route failure is addressed in [8].
Sudden "step" changes of QoS values (throughput) referred also as "congestion" pattern [11], are usually associated with a network reconfiguration, e.g. a new route, or a link upgrade. "Break" patterns describing Voice over IP delay behaviour are shown in [11]. "Diurnal" patterns [13] are oscillations in the throughput on a daily basis, showing for instance "rush hour" behaviour, "weekend" patterns, etc.
Although patterns are used in the current research to explain specific QoS or traffic behaviour in the telecommunication networks, however, there is a lack of systematic approach to define and extract pattern structures from measured QoS parameter data.
This work proposes a model for pattern definition, e.g. structures of time series data sequences (QoS, traffic, resources) in communication networks. The model includes:

- Basic pattern types specific for the measured QoS parameters in telecommunication network scenarios.
- Composite patterns built from basic patterns
- Common shape structure parameters applied for all basic patterns contained in a composite pattern structure.
- Pattern description language used to describe the pattern composition.

The pattern definition model is used for similarity search by pattern matching algorithms. Figure 1 illustrates the pattern definition approach.

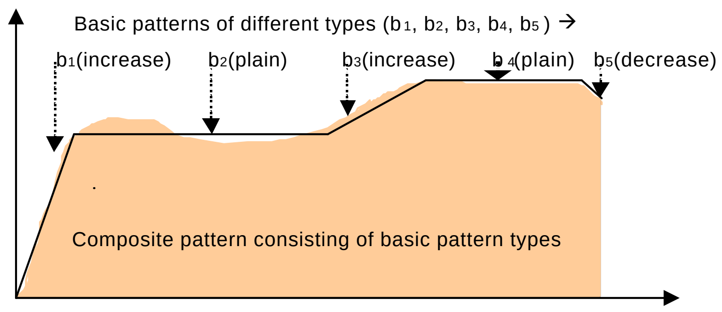

Fig. 1. Pattern definition approach

2.2 Pattern Definition Language

Depending on the application area, patterns could be extracted focussing on different structures and functions operating with the patterns. The Shape Definition Language [14] is an example for pattern description based on language concept.
This work defines a Pattern Definition Language (PDL), which considers the specifics of the pattern structures obtained from time series data sequences of measured QoS parameter values in telecommunication networks. The PDL description includes the concept of basic pattern types commonly found in the telecommunication networks which could be explained in some context and which could be used for building arbitrary composite pattern structures. PDL main syntax includes:

Composite pattern structure = <Shape parameters>, <QoS description>, <Basic pattern sequence>
Shape parameters = 'shape' (distance measure, < definition options>)
QoS description = 'QoS' (QoS metric, metric unit, monitoring scale)
Basic pattern sequence = Basic pattern structure < ,Basic pattern structure>
Basic pattern structure = <basic pattern type> '('<length > <, range >')'
Basic pattern type = 'extreme' I ‘plain’ I 'incr' I 'decr'

Basic patterns structures include extreme value, plain, increase and decrease pattern types. Shape parameters are aimed to describe structure options for all basic patterns included in the composite pattern structure. Distance measure is shape parameter, which defines an interval d used to obtain the particular basic pattern type. Outlier definition option defines the methods for calculation of extreme value patterns.

2.3 Basic QoS pattern types

In this section, basic pattern types are studied, which structure is characterised by:

- Type of basic pattern
- Range parameters describing the basic pattern types by specific values of the time series data sequence (for instance minimum, maximum, mean)
- Length defining the number of time series data included in the basic pattern.
- Shape definition parameters.

Extreme value. Extreme values are called outliers. In [9], we described extreme value detection using ARIMA models. Considering that the estimation error in ARIMA follows a Gaussian distribution with a standard deviation σ, than as “extreme value” is labelled every sample of $\{X_t\}$ with $X_t >= 2\ \sigma$ away from the estimated value.
The reason is that in a Gaussian distribution, 95% of the probability mass is within the interval [mean-2σ, mean+2σ] from the mean. The extreme value could be calculated based on multiple of σ or QoS parameter threshold values. Figure 2 shows extreme value patterns, which we extracted from inter-domain measurement scenarios [9].

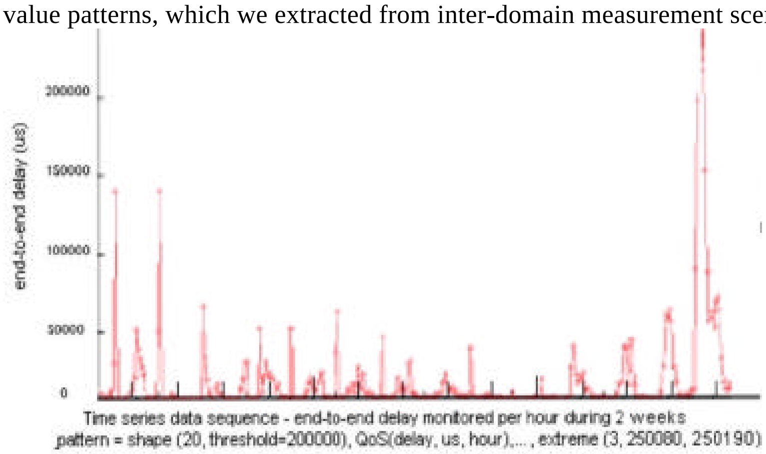

Fig. 2. Extreme value patterns

The number of consecutive extreme QoS values defines the length of the extreme value pattern. The extreme value detection method could be applied to detect sudden unexpected very high delay or delay jitter values due to anomalies, for instance router changes or implementation biases in the routers [8].

Plain. Plain pattern is a sequence of time series data with same or similar QoS values ranging within a given interval**.** Let $\{X_t\}$ be time series data consisting of QoS parameter values. With $x_1,...x_i,...x_n$, belonging to $\{X_t\}$, the QoS values of a plain pattern structure is given for which hold that diff = $| x_i - x_j | < d$
For each neighbour pair x_i, x_j, the difference between neighbour values should be smaller than some distance measure parameter d. The range parameter of the plain pattern is given by the mean QoS parameter value of the data included in the pattern.
The plain pattern is typical for network connections with small traffic variations. It is useful to predict performance of multimedia traffic requiring QoS guarantee.
In figure 3, plain pattern structures are obtained from active QoS measurement scenarios using CM Toolset [15]. Mean end-to-end delay is monitored for a given day of the week. The time series data is collected per hour.

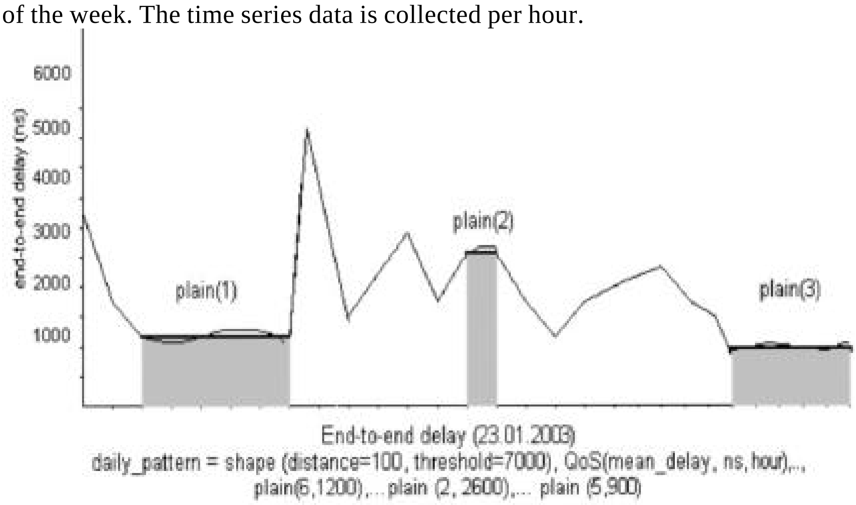

Fig. 3. Plain pattern description

Increase/decrease pattern. QoS time series data structure, which is characterized with either increasing, or decreasing values of the QoS parameters is called increase res. decrease pattern. Let $\{x_t\}$ be time series data consisting of QoS parameter values. The QoS values $x_1,...x_i,...x_n$ belonging to $\{x_t\}$ build

- Increase basic pattern, when $0 =< x_1 = x_{min} < \ldots x_i < \ldots < x_n = x_{max}$ with start QoS value x_{min} as the minimum and last QoS value as maximum value of the sequence;
- Decrease basic pattern, when $x_{max} = x_1 > \ldots x_i > \ldots > x_n = x_{min} >= 0$ with start QoS value as the maximum and last QoS value as minimum value of the sequence.

For each neighbour x_i , x_j hold that

diff = | x_i - x_j | $\geq$ d

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

The increase/decrease patterns are described by range parameters showing minimum and maximum time series data value. Increase/decrease patterns are useful for describing the dynamic of QoS behaviour, e.g. “rush hours“, “congested networks“.

From a mathematical point of view the real measured process with time discrete continuous values is mapped by the plain and increase/decrease patterns into a simple approximated process with plain phases (gradient = 0) and increase/decrease phases (gradient (max-min)/length >/< 0).

Figure 4 shows an example for decrease and increase basic pattern types obtained in QoS measurement scenarios using CM Toolset [15].

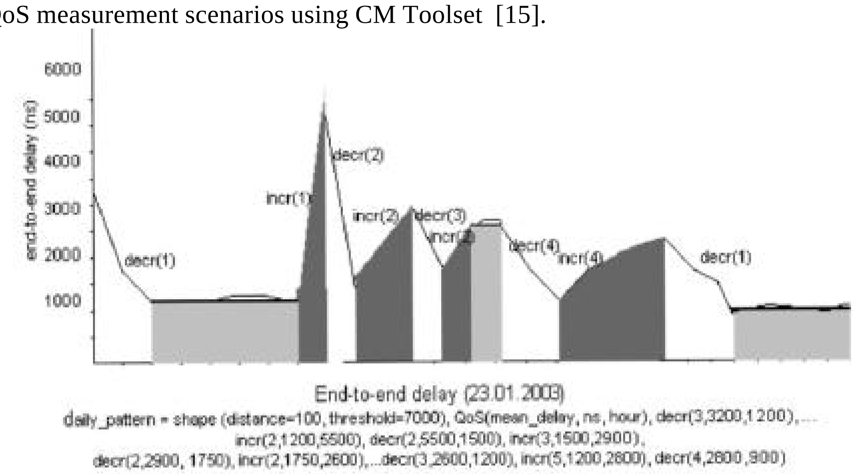

Fig. 4. Increase / decrease patterns

2.4 Composite patterns

In the area of QoS analysis is useful to study more complex patterns. Composite patterns are defined by sequences of basic pattern types.

An example is the “step” pattern, which is defined by sequence of plain patterns with different mean values. Considering a network connection with small QoS variations (i.e. plain QoS pattern), the selection of a new route could increase or decrease the QoS provision, shown with the “step” pattern definition.

“Peak” pattern is another composite pattern, defined by sequence of an increasing pattern followed by a decreasing pattern. “Valley” pattern as the opposite of “peak” is defined by decreasing pattern followed by increasing pattern.

More complex composite patterns are built considering the range parameters of the sequences of basic patterns. For instance, “plateau” pattern includes in the sequence of basic patterns a plain pattern which mean value is bigger that the range values of the other included basic patterns.

3 Spatio-temporal QoS pattern analysis

3.1 Goals and architecture for spatio-temporal QoS pattern analysis

To support the network engineer in deriving of significant patterns from measurements for the purpose of QoS data mining, the spatio-temporal QoS pattern analyser is developed in the framework of INTERMON [4] project. The developed technology is based on automated detection, analysis, storage and processing of QoS patterns considering spatial (network topology) and temporal (time windows and periods) context. The network engineer selects the measurement units and pattern structure parameters (distance measures, extreme value calculation options) based on his experience (on route change, abnormal QoS values) and requested QoS/SLA.
The tool interprets the parameters provided for automated pattern detection and transformss them into PDL specifications. Dependent on the data-mining task to be solved, the network engineer selects the options for pattern similarity detection in the spatio-temporal context. The system applies the algorithms as required and stores the results of the similarity analysis in the pattern data base. The interpretation of detected patterns and deriving of models and heuristics for solving of specific data mining tasks using the pattern data base is done by the network engineer.
The usage of the spatio-temporal QoS pattern analyser for automation of the data mining tasks is shown in Figure 5.

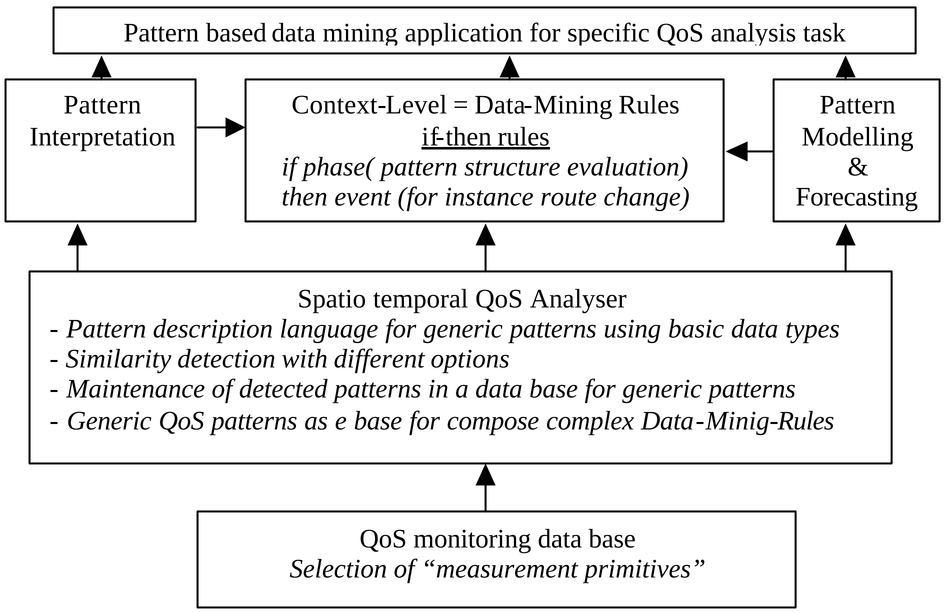

Fig. 5. Data mining approach using the spatio-temporal QoS pattern analyser

In particular, following functions are supported for enhanced planning and engineering based on patterns:

- Selection of patterns from operational QoS monitoring data base and storage of selected patterns in a pattern data base,
- Pattern similarity analysis using algorithms with different options,
- Temporal QoS pattern analysis by matching similar patterns in specific time window or periods,
- Spatial dependency analysis by detecting similar patterns considering different flows and topologies,
- Flexible definition and query of patterns by graphical user interface
- Specification of pattern structures using pattern description language.

The general architecture of the spatio-temporal QoS pattern analyser is shown:

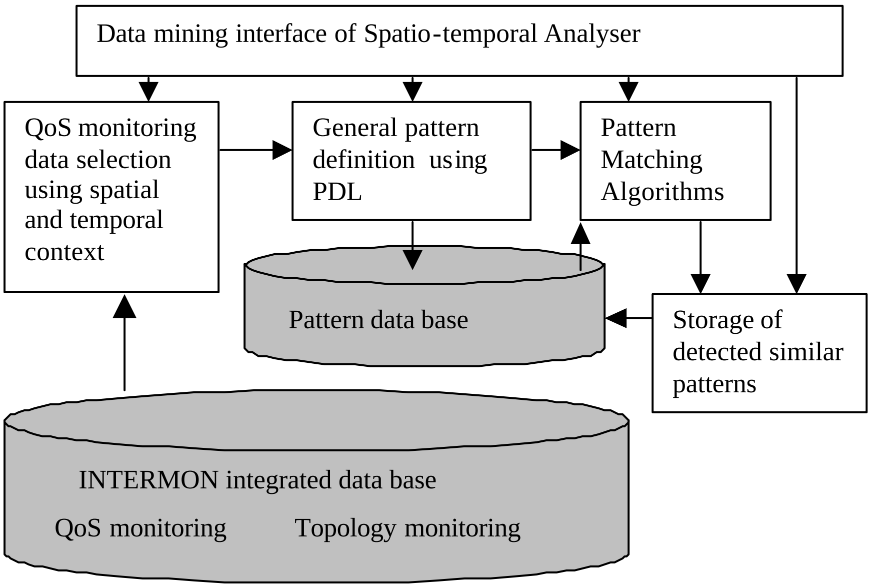

Fig. 6. General architecture of spatio-temporal QoS pattern analyser

The developed architecture uses a data base for management of detected significant patterns with their relationships to other similar patterns found by operational QoS monitoring tools in a framework considering different network connections and flows The particular goal is to support the network engineer in the area of:

- Analysis of temporal QoS pattern dependencies in different network connections (to show impact of load on common topological structures, influence of Denial of Service attacks). ·
- ·Analysis of spatial compositions of QoS parameters based on patterns. Detection of a "dominant" spatial connection which QoS patterns impact mainly the QoS patterns of the end-to-end connection.
- Extraction of QoS patterns for inferencing of QoS parameters considering different cases of "hidden topologies" in network tomography scenarios [16].

3.2 Pattern matching algorithms and data base for pattern analysis

The pattern matching algorithms are aimed to detect similar patterns using different options. Assuming a given pattern structure Q= [q(1),....q(m)] which is called the query pattern structure, the pattern matching (discovery) algorithms try to find all similar patterns in a search time series data sequences X = [x(1),...x(n)] (i.e. measured QoS parameters) considering the spatial and temporal constraints of the measurement scenarios. The algorithms are characterized with following functions:

- Detection of the set of search time series data sequences in spatial context $X_{sp} = \{X1,\ldots Xn\}$. X_{sp} includes all relevant time series data sequences obtained for flows and connections in the considered network topology.
- Detection of patterns included in the spatial composition of end-to-end connection. Assuming that the time series data $Q_{end\_conn,\Delta}$ describes the QoS of the end-to-end connection C_{e2e} for time window Δ, then the spatial QoS composition of $Q_{end\_conn,\Delta}$ is defined by the sum of time series data describing the connections C_i, i = 1...n, which are part of the end-to-end connection.

  $$C_{e2e}= C_1....., C_i.......C_n$$

  $$Q_{end\_conn,\Delta}=Q_{C1,\Delta}+ Q_{Ci,\Delta} + Qc_{n,\Delta}$$

- Temporal based analysis considering time windows and periods. Time window (t_1, t_n) means search of pattern in a time series data sequence X = [x(1),...x(n)] with start element x(1) at time t_1 and last element x(n) at t_n.
- Similarity detection using different options for similarity definition.

  "Hard" similarity matching requires that the matched subsequence within X defines a pattern with exactly same PDL description as the description of the query pattern Q.

  1. "Amplitude free" similarity match does not consider the range parameters in the detection of similar patterns
  2. "Offset free" matching means that the length of the basic patterns is not considered in the comparison for similarity.
  3. "Gaps" similarity check allows matching of similar patterns discarding specific subsequence(s) from the time series data structure called "gaps".

- Mapping of detected similar patterns to spatial context (e.g. network connections and flows) using topology information.
- Support of data mining including pattern based modelling with detection of network connections (flows) with similar patterns for specific time period, detect of similar patterns of a given network connection in specific time windows (periods), matching of pattern dependencies in spatial compositions.

In order to support data mining based on pattern analysis, a pattern data base is used which collects for each significant pattern in systematic way the results of the pattern matching algorithms.

The pattern analysis data base giving information on similar patterns in spatio-temporal context is used for solving of tasks in the area of capacity planning, forecasting, bottleneck detection, Denial of Service Attacks and other network management tasks.

The pattern analysis data base is provided as MySQL data base operating on top of the CM Toolset QoS monitoring data base, using the QoS parameter measurements for specific scenarios managed by the CM Toolset data base [16].
Figure 7 shows the main design of data items and relationships included in the pattern data base for pattern dependency analysis:

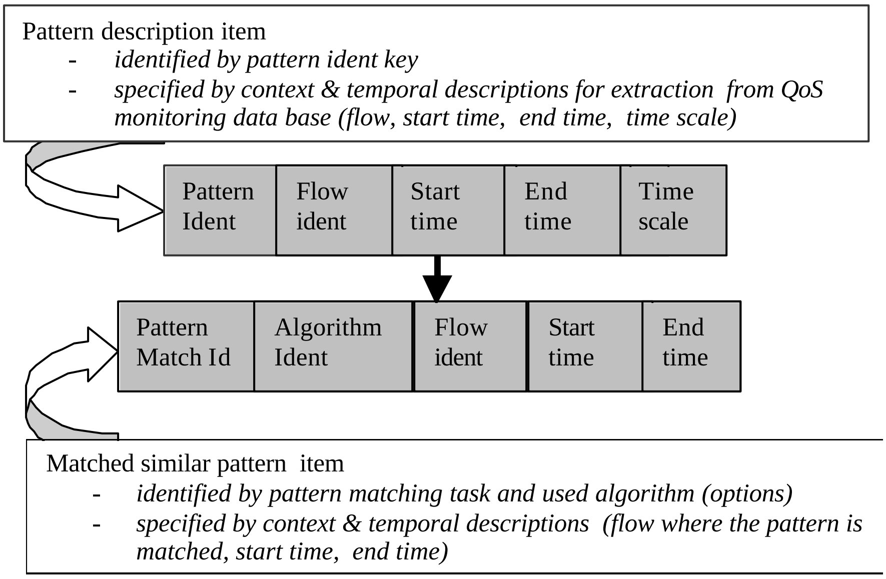

Fig. 7. Design of data base for pattern dependency analysis

4 Integrated QoS pattern and inter-domain routing data mining in the framework of INTERMON toolkit

The INTERMON architecture is developed to combine different kinds of techniques for automated QoS analysis in large scale inter-domain environment based on monitoring. modelling and visualisation [4].
Particular tools and technologies of INTERMON architecture are:

- Inter-domain route analysis and connectivity monitoring – InterRoute Explorer[15].
- Active QoS measurement per traffic flows using CM Toolset combined with active monitoring of inter-domain connection topology [16].
- Traffic measurement using IETF IPFIX traffic flow export concept [17].
- Border router monitoring tools collecting MIB information [4].
- Visual Data Mining (VDM) system for inter-domain QoS analysis aimed at automated processing, filtering, modelling and visualisation of different kind of QoS, traffic and topology data [18].

- Traffic matrix explorer [19] for inter-domain traffic engineering.
- Spatio-temporal QoS pattern analyser for data mining used in the area of network planning, anomaly analysis and QoS engineering.
- Measurement based modelling including IP inter-domain delay analyser [4], ARIMA forecasting models [9] and Class Based Traffic models [20].
- Inter-domain simulation toolkit integrating measurement based traffic and inter-domain performance models in hybrid packet [21], fluid [22] and time series data [23] simulation technologies.

INTERMON integrates tools and technologies with common data base and user interfaces, using mechanisms for policy control interaction, to provide specific tasks in the area of inter-domain QoS analysis, planning and traffic engineering.

The spatio-temporal QoS pattern analyser is included in INTERMON as visual data mining application on top of CM Toolset QoS monitoring and topology data base. Scenarios using QoS patterns are based on particular interaction of INTERMON tools:

- Performance anomaly detection based on combining QoS pattern analysis with inter-domain path connectivity monitoring (QoS pattern detection in case of topology changes and inter-domain routing instability).
- QoS and capacity planning based on QoS pattern analysis and IPFIX traffic flow measurements for specific inter-domain connections.
- Inter-domain traffic engineering based on combined traffic flow matrix exploration and detection of QoS patterns of aggregated flows.
- QoS pattern detection of inter-domain connections included in multi-homing topologies for selection of “optimal” inter-domain paths for end users.
- Inter-domain route quality analysis using InterRoute Explorer combined with evaluation of results obtained from spatio-temporal QoS pattern analyser.

5 Conclusion

This paper was focussed on the automated spatio-temporal QoS analysis based on patterns. This is a new technology integrated in the INTERMON toolkit for systematic detection, exploration and dependency analysis of QoS patterns in large scale networking environment.

Current usage of the described technology in the framework of INTERMON project is aimed at collection of patterns obtained from real world QoS monitoring experiments to detect patterns of interest to be specified and interpreted by operators and network analysts. Using the patterns collected and related in the pattern analysis data base, the further focus is aimed at development of heuristics for solving of specific tasks in the area of network management, QoS planning and network anomaly detection.

Enhancing of the pattern analysis data base with new kinds of patterns such as inter-domain routing and traffic patterns will provide the possibility for integrated data mining in telecommunication networks considering spatio-temporal relationships of different kind of parameters. The application of the proposed technology for pattern based interpretation and deriving of data mining models using network engineer experience is further topic of research.

References

1. P.J. Brockwell, R.A. Davis, Introduction to Time Series and Forecasting, Springer Verlag, 2002.
2. E.Keogh, S.Lonardi, B.Chiu, Finding Surprising Patterns in a Time Series Database in Linear Time and Space, SIGKDD 02, Edmonton, Canada, July, 2002.
3. C.Estan, S. Savage, G.Varghese, Automatically Inferring Patterns of Resource Consumption in Network Traffic. SIGCOMM 2003 Conference, August, 2003.
4. Advanced architecture for INTER-domain quality of service MONitoring, modelling and visualisation, INTERMON project, http://www.ist-intermon.org.
5. E. Keogh, H. Hochheiser, B. Shneiderman, An Augmented Visual Query Mechanism for Finding Patterns in Time Series Data, University of Maryland, Techical Reports, CS-TR-4398, http://www.cs.umd.edu/Library/TRs/.
6. J.J. van Wijk, E.R. van Selow, Cluster and Calendar based Visualization of Time Series Data, IEEE Symposium on Information Visualization, San Francisco, 1999.
7. QueryScetch - http://www.bewitched.com/projects/querysketch.
8. K. Papagiannali, S. Moon, C. Fraleigh, P. Thiran, F. Tobagi, C. Diod, Analysis of measured single-hop delay from an operational backbone, INFOCOM, 2002.
9. I. Miloucheva, A. Anzaloni, E. Müller, A practical approach for QoS forecasting considering outliers, IPS, http://www.ist-intermon.org, Salzburg, 2003.
10. John F. Roddick, M. Spiliopoulou, A Bibliography of Temporal, Spatial and Spatio-Temporal Data Mining Research, ACM SIGKDD, June 1999.
11. NLANR Surveyor, http://dast.nlanr.net/Articles /measurements/surveyor.html .
12. S. Salza, M. Draoli, C. Gaibisso, A. Laureti Palma, R. Puccinelli, Methods and Tools for the Objective Evaluation of Voice-over-IP communications, INET, 2000.
13. L. Cortell, Logg, Throughput Time Series Patterns (Diurnal and Step Functions), http://www.slac.stanford.edu/comp/net/pattern/diurnal.html.
14. R. Agrawal, G. Psaila, E. L. Wimmers, M. Zait, Quering Shapes of Histories, Proc. 21th International Conference on Very Large Data Bases (VLDB'95), 1995.
15. P.A.A. Gutierrez, I. Miloucheva, Analysis of end-to-end QoS behaviour in inter-domain environment, IPS, http://www.ist-intermon.org ,Salzburg, 2003.
16. T. Pfeiffenberger, I. Miloucheva, U. Hofmann, A. Nassri, Inferencing of inter-domain path characteristics, IPS, http://www.ist-intermon.org, Salzburg, 2003.
17. F. Raspall, S.Tartarelli, M.Molina, J. Quittek, Implementing an IETF IPFIX meter, IPS, http://www.ist-intermon.org, Salzburg, 2003.
18. S. Michaelis, J. Seger, Concept of configurable filters for Visual Data Mining System, IPS, http://www.ist-intermon.org, Salzburg, 2003.
19. A.Kock, Flexibke Traffic Matrix Analyser for Inter-domain Network Operation and Planning, IPS, http://www.ist-intermon.org, Salzburg, 2003.
20. U. Hofmann, Trace based traffic modelling, IPS, http://www.ist-intermon.org,2003.
21. F. Baumgartner, M. Scheidegger, T. Braun, Enhancing Discrete Event Network Simulatord with Analytical Network Cloud Models, IPS, http://www.ist-intermon.org, Salzburg, 2003.
22. P. Haber, G.Bergholz, U.Hofmann, I.Miloucheva, Multi-class signal flow model for inter-domain traffic flow simulation, IPS, http://www.ist-intermon.org, 2003.
23. T. Mahr, T. Dreillinger, A. Vidacs, Time Series Based Simulation Architecture, IPS, http://www.ist-intermon.org, Salzburg, 2003.