

Pattern based spatio-temporal Quality of Service analysis for capacity planning

Natalie Toumba
Department for Computer Science
University of Salzburg
ntoumbas@salzburgresearch.at

Ilka Miloucheva
Jakob Haringer Str. 5/III
Salzburg Research
ilka.miloucheva@salzburgresearch.at

Abstract

Analysis of QoS behaviour for different periods (hourly, daily, weekly) is fundamental for capacity planning and enhanced QoS provision. Before QoS forecasting techniques are applied, preliminary spatio-temporal analysis of QoS behaviour should be done, in order to detect bottlenecks in resources and QoS outliers due to congestion and anomaly events.

This paper is focussed on a practical approach for spatio-temporal analysis of QoS time series data based on QoS pattern detection and analysis. Statistical methods are applied to study QoS patterns in different periods (hourly, daily, weekly). We argue the usage of bucket histograms based on QoS thresholds (levels) for pattern oriented QoS analysis useful for network capacity planning.

The proposed spatio-temporal QoS analysis includes two components:

- *temporal pattern detection and analysis of time series data of QoS parameters obtained for network connections*
- *spatial dependency analysis of patterns based on time series data of QoS parameters describing different network connections.*

Practically, we discuss pattern oriented QoS analysis for different periods (daily, weekly) based on QoS time series data obtained from real measurement scenarios.

Keywords: QoS, pattern, spatio-temporal analysis, histogram, bucket, capacity planning,

1. Introduction

QoS data mining is important for short term capacity planning and QoS optimisation decisions. Patterns describe some specific behaviour of QoS time series data (univariate, multivariate) for a given time period. In the known research more focus was put until now on to study traffic patterns, as for instance [GP 94], [MC 00], used to improve traffic engineering and traffic resource allocation.

Pattern based QoS analysis is useful for studies and comparison of the QoS provision for different periods in network links. Pattern analysis of univariate and multivariate time series data describing QoS parameter values could significantly improve capacity planning for short and medium term periods (hourly daily, weekly), network dimensioning and load balancing in inter- and intradomain environment. For instance, based on patterns analysis, “congested” periods could be identified and additional resource allocation for these periods could be provided. Pattern characterisation as periodically, partially periodical and casual patterns could provide insight in QoS provisioning and resource scheduling.

A practical approach for spatio-temporal QoS analysis based on usage of statistical methods and especially QoS threshold histogram techniques for pattern description is proposed in this paper. The paper is focussing on the efficient QoS based pattern analysis technology useful for short term capacity planning and QoS operation.

An experimental environment based on QoS monitoring equipment (CM Toolset) is built for studies of hourly, daily and weekly QoS patterns in spatio-temporal context of real world networks. Preliminary results are discussed in this paper.

This paper is organised as follows. Section 2 describes the framework for pattern based spatio-temporal analysis. The experimental testbed for pattern based QoS analysis is shown in section 3. Studies on hourly, daily and weekly QoS patterns of connections between access and inter domain provider networks are discussed in section 4. Section 5 concludes this paper with the further work.

2. Framework for pattern based spatio-temporal QoS analysis

The contribution of this paper is the development of methods to characterise QoS behaviour as patterns useful to study capacity and QoS provisioning in spatio-temporal network context.

For this purpose we first study QoS pattern characterisation based on statistical techniques and QoS thresholds. After this we consider the integration of pattern characterisation techniques for spatio-temporal QoS analysis.

2.1. QoS pattern characterisation based on statistical techniques and QoS threshold histograms

Specific behaviour of the time series data (QoS parameter) for some time interval (time window) is called pattern. QoS behaviour described by patterns could be characterised with different statistical techniques (distributions, histograms, etc), intervals for time series data aggregation and algorithms for evaluation and approximation of the data. Pattern analysis is based on pattern characterisation and comparison for a given time window.

Research efforts to study QoS patterns such as “outliers” are discussed in [AMM 03], outliers due to route failure are focussed in [PMFT 02]. Delay “congestion” pattern are addressed in [NLANRSurveyor] and “delay break” patterns in [SDG 00]. [CL 03] describes diurnal and step function patterns. Oscillations in the throughput on a daily basis, e.g. high throughput at night or weekends when there is lower utilization and congestion, and higher performance at other periods are referred in as diurnal variations. Sudden step

changes in throughput are usually associated with a network change, e.g. a new route, or a link upgrade.

For QoS pattern characterisation and comparison, statistical methods such as mean, minimum, maximum values, variance and distribution could be used.

Histograms are well known means for approximation of distributions, therefore they could be used for pattern descriptions and analysis.

Histograms capture data distribution in a space efficient fashion. They have been designed to work well with numeric value domains and have been used to support data mining, query optimisation and statistical studies [GKS 01] [DIR 99] [JPKS 98].

The proposed QoS-threshold based histogram approach considers that the QoS for different kinds of applications could be described with different thresholds which are significant for the behaviour of the applications and resource reservation.

The QoS threshold histogram splits the time series data in buckets, defined by lower and higher QoS thresholds.

This approach is usable for QoS and capacity planning.

Let $\{X_t\}$ be time series data consisting of QoS parameter values. With $q_1, \dots, q_i, \dots, q_n$ the QoS thresholds values are given so that $0 = q_0 < q_1 < q_2 < \dots < q_n = q_{\max}$ where q_{\max} is the maximum possible QoS threshold value.

A bucket b_i is defined by $[q_i, q_{i+1}]$ with q_i as low boundary and q_{i+1} as high boundary.

A QoS parameter value x_t observed at time t belongs to bucket b_i if $q_i < x_t \leq q_{i+1}$

Thus, the QoS threshold histogram is based on obtaining the amount of the QoS parameter values belonging to buckets, defined by lower and higher QoS thresholds.

In addition to building the frequency bucket histogram based on QoS thresholds, the bucket based technique could be used to characterise temporal patterns based on the assignment of the QoS parameter values of $\{X_t\}$ to the buckets b_i when certain thresholds are exceeded.

2.2. Spatio-temporal pattern analysis across network connections

The spatio-temporal QoS pattern analysis across network connections as well as spatial inferencing and composition of QoS is a new research topic arising from the need to study performance and resource dependencies in macroscopic and inter-domain Internet environment.

General spatio-temporal visual data mining concepts are addressed by a lot of international activities and research [RS 99] [AR 97], [AR 98]. The data mining community has developed a wide variety of innovative techniques for algorithmically extracting interesting patterns from time series data [BC 96], [FRM 94] in combination with spatial queries. The most contributions are however aimed at geographical, geophysical, medicine and other areas and cannot directly be used for telecommunication networks.

The spatio-temporal pattern analysis proposed in this paper is based on the characterisation, comparison and inferencing of QoS patterns observed for traffic flows of different network connections.

Network connections are constituted of spatial elements (border router, end-systems, ISPs) and their relationships. Relationships to be considered in the spatio-temporal framework are:

- Type of spatial elements (border, core, egress router, Autonomous systems, countries...)
- Relations between spatial elements (i.e. spatial relation) describing inter-domain topology and other dependencies of interest (example for dependencies between spatial elements are inter-domain connection, end-to-end connection). Examples for such relations are
 - “An inter-domain connection is part of end-to-end connection”,
 - “Border router R is element of inter-domain connection”,
 - “Border router R belongs to Autonomous System AS”, etc.
- Spatial Clusters. Clustering is the task of grouping of connections for specific purposes, for instance cluster is defined to identify connections belonging to a Virtual Private Network (VPN) or to characterise a set of

autonomous systems included in some region or country).

- Spatial temporal relationships defining dependencies between spatial elements, traffic and QoS time series data (for instance, traffic data and border router QoS data and connections (inter-domain, end-to-end).

The spatio-temporal pattern analysis framework based on the QoS time series data is illustrated in the following figure:

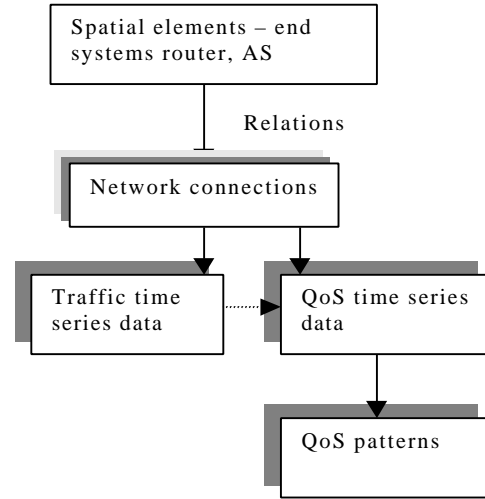


Figure 1: Spatio-temporal QoS analysis framework

A concrete data base design useful for inter-domain spatio-temporal analysis of QoS data is the INTERMON data base [HMH 02] which relates inter-domain topologies and traffic with QoS time series data. Pattern based spatio-temporal QoS analysis includes pattern comparison using statistical and histogram techniques.

Pattern dependencies could be used to study spatial composition of the end-to-end QoS.

If the time series data $Q_{\text{end_conn}, \Delta}$ is describing the QoS of the end-to-end connection C_{e2e} for some interval Δ , then the spatial QoS composition of $Q_{\text{end_conn}, \Delta}$ is defined by the sum of the QoS time series data – $Q_{C_i, \Delta}$ describing the QoS of connections C_i , $i = 1 \dots n$, which are part of the end-to-end connection, in interval Δ .

$$C_{e2e} = C_1, \dots, C_i, \dots, C_n$$

$$Q_{\text{end_conn}, \Delta} = Q_{C_1, \Delta} + Q_{C_i, \Delta} + Q_{C_n, \Delta}$$

QoS parameter values for a connection could be obtained for aggregated or application traffic, in such cases the spatial composition will have additional characteristics dependent on the QoS of the observed traffic.

Spatio-temporal QoS inference is another kind of spatio-temporal QoS analysis which could be viewed as special network tomography problem.

The network tomography is focussed on inferring QoS and traffic in large inter-domain infrastructures [CHNY], [MNH 02], [TTO 02] considering in specific cases QoS and pattern dependencies.

3. Experimental testbed for QoS pattern studies

Time series data of QoS parameter values was obtained based on the CM Toolset monitoring of end-to-end connection between access network (Salzburg Research) and ISP service provider (Austria Telekom). CM Toolset is agent based monitoring tool of QoS parameters of application flows [AQUILA], [MH 01], [SDH 01], [HMSP 02].

The following picture describes the networking scenario – delay monitoring of end-to-end connection between access network and ISP service provider.

Such networking scenario is typical for the today providers and customers of networking services and applications, for instance VPNs, which are interested in capacity planning and QoS provision considering access routers.

We have two routers between the server and receiver, the one of them is the access router of the customer network to the ISP.

The delay measurement approach is based on GPS clock synchronisation.

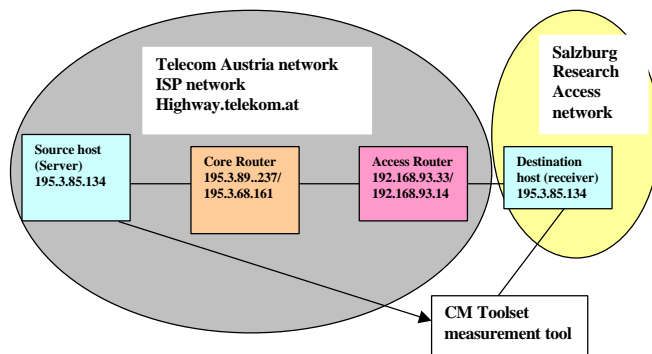


Figure 2: End-to-end QoS measurement scenario

The end-to-end delay and loss monitoring information was filtered on second, minute, hourly, daily and weekly base, i.e. QoS parameters are obtained for each of these periods, in order to detect QoS patterns with different expressive power considering QoS planning and provision. Mean, minimum and maximum time series data describing different QoS parameters were also obtained and used for QoS pattern characterisation.

The scenario does not include spatial time series data sources, such analysis will be studied in future work.

4. Temporal pattern studies for capacity planning

QoS patterns are studied during 2 weeks to provide information for congestion, load balancing and capacity planning especially considering QoS buckets obtained from threshold boundaries.

In this studies, the patterns discussed are based on delay QoS analysis – the end-to-end delay QoS monitoring values are considered – mean, maximum and minimum obtained for different time scales, primary per hour.

4.1. Daily QoS patterns

Study of daily QoS patterns is important for short term capacity planning and load balancing. In this case, a daily QoS pattern is defined based on the hourly minimum, maximum and mean QoS value behaviour.

The usage of these values is useful to identify the QoS “rush hour” periods of the day.

As seen in the Figure 3, the mean delay is greater than the minimum delay from 6:00 until 18:00. In the “not rush hour” period, the mean delay is nearly equal to the minimum.

The maximum delay values are not so high as in the “rush hour”. The same pattern with a clear “rush hour” QoS pattern period is obtained for the other days of the observed two weeks.

For all “working” days, a similar daily delay QoS pattern with “rush hour” effect is observed. The patterns of two days are shown.

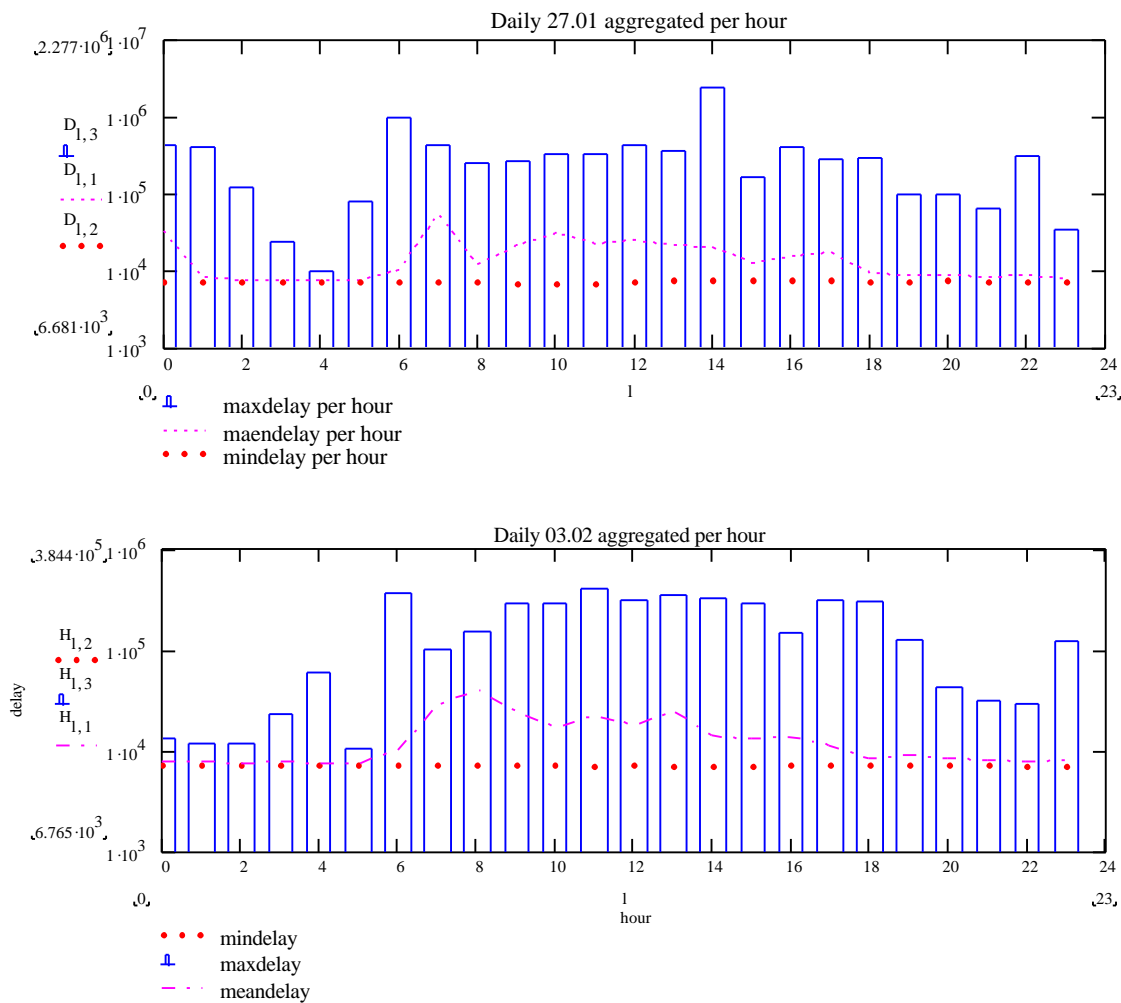


Figure 3: Daily QoS patterns described by minimum, maximum and mean QoS parameter values of the end-to-end delay obtained per hour

Daily QoS pattern studies are aimed to detect performance bottlenecks, “QoS outliers”, QoS behaviour due to congestion, peaks and QoS stability considering the hourly QoS behaviour.

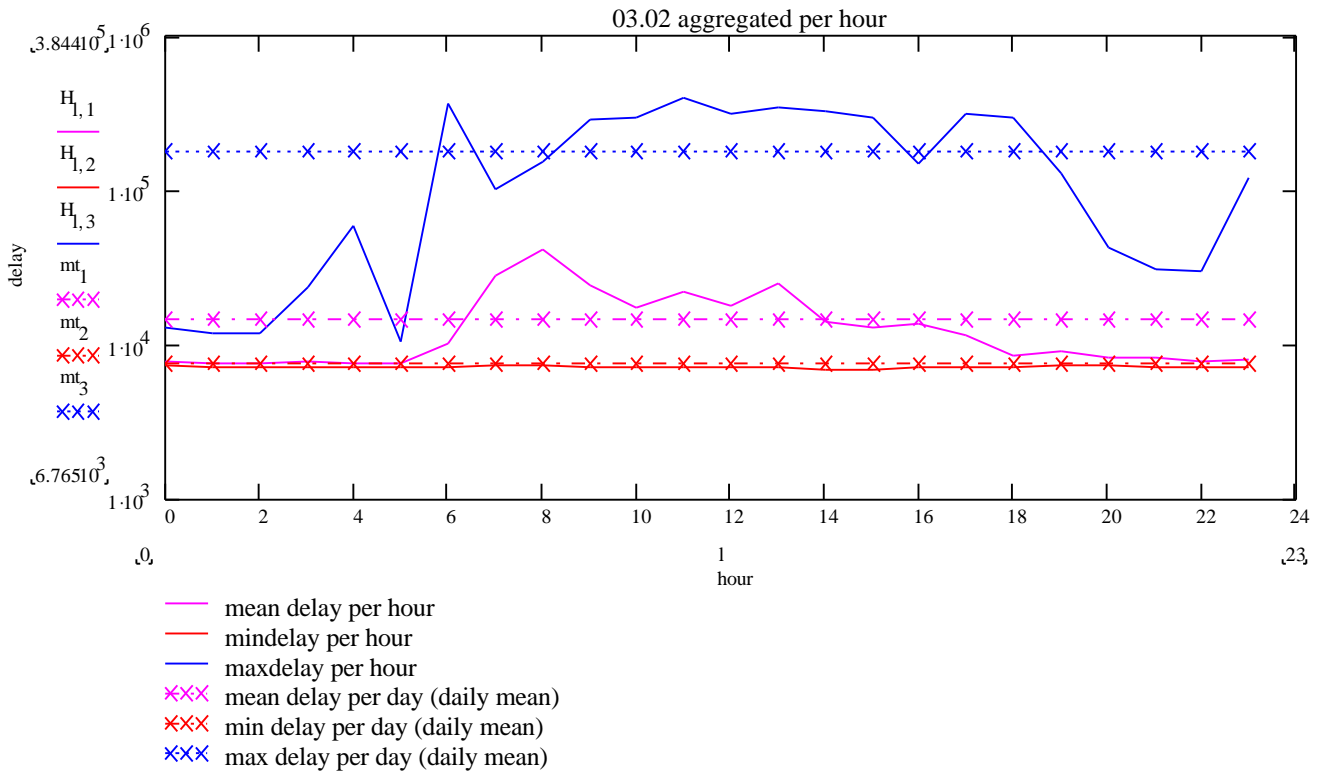


Figure 4: Daily QoS pattern compared with daily threshold values (mean, maximum and minimum delay per day)

Therefore daily pattern analysis could be extended considering comparison of the daily patterns with some thresholds such as daily mean QoS values and the standard deviations

In Figure 4 the minimum, mean and maximum delay QoS behaviour characterising the daily pattern are visualised and compared with the mean, minimum and maximum delay QoS value of the day. In the period between 6:00-14:00 it is shown that the mean value of the day is exceeded.

Figure 5 shows again a daily pattern based on mean, maximum and minimum delay per hour compared with the daily mean value of the delay.

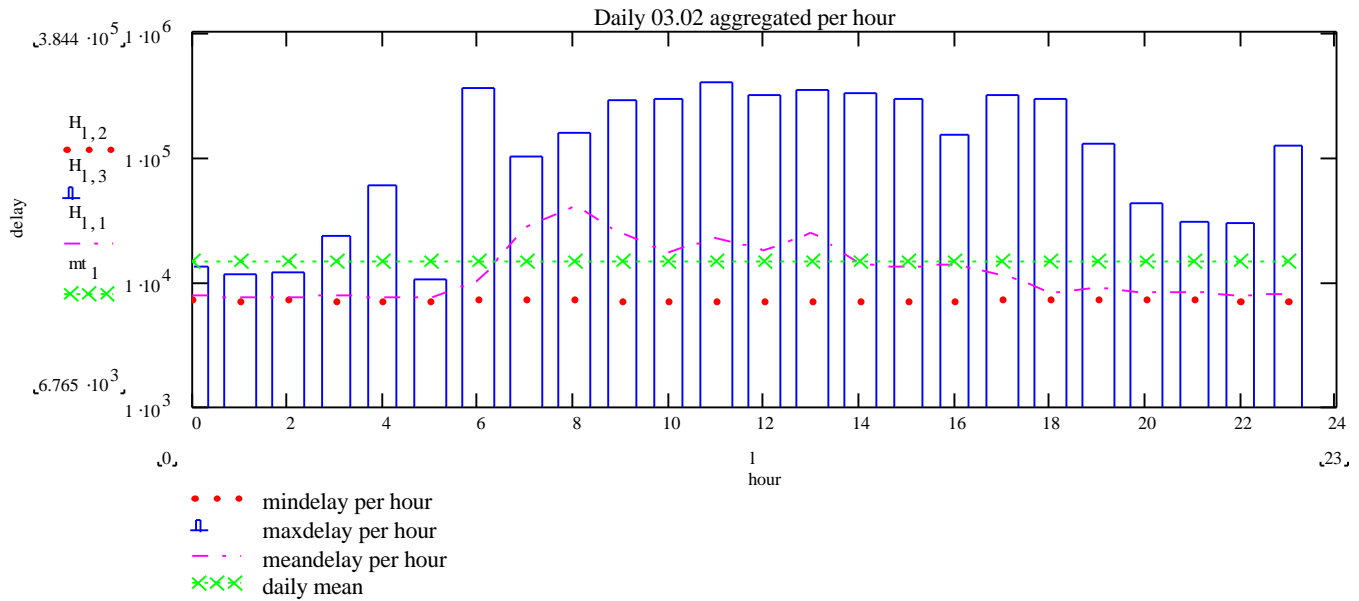


Figure 5: Daily end-to-end delay QoS pattern compared with the daily delay QoS mean value

The daily standard deviation compared to the hour standard deviation could also provide insight in the “rush hour” pattern behaviour as it is seen in the Figure 6. The hour standard deviation is greater than the daily standard deviation for the “rush hour” period 6:00-18:00.

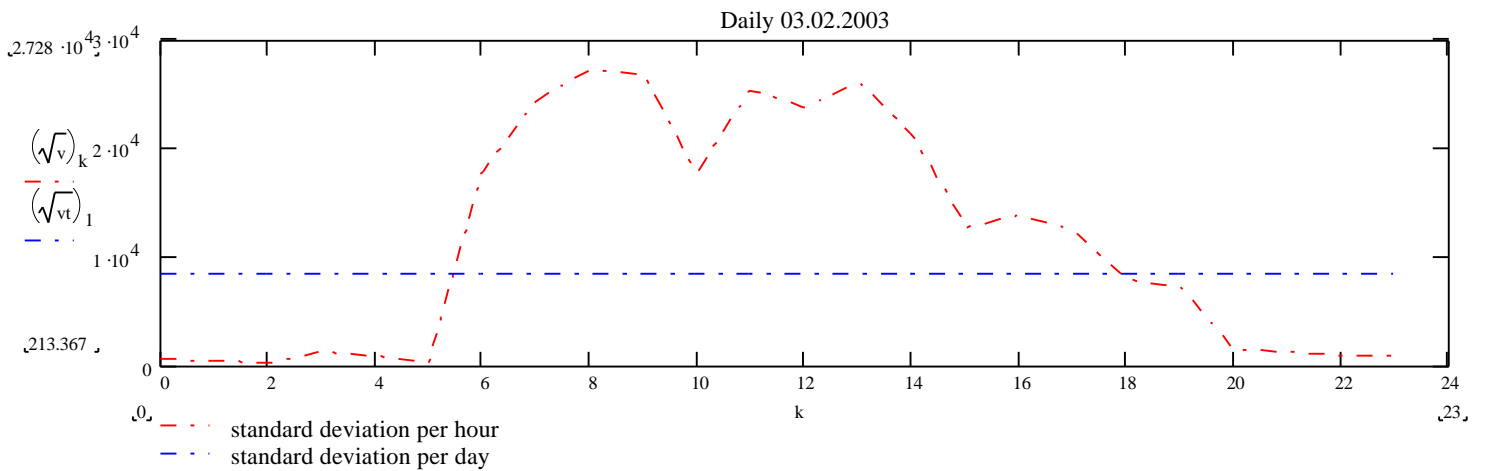


Figure 6: Daily pattern based on end-to-end delay standard deviation per hour compared with the standard deviation per day

4.2. Patterns based on QoS threshold histograms

Considering QoS thresholds, different patterns could be defined. Figure 7 shows a temporal QoS pattern based on assignment of end-to-end delay QoS values to buckets built from delay QoS threshold ranges. At each interval, the behaviour is given by a QoS threshold value which is not exceeded by the mean QoS value of this interval.

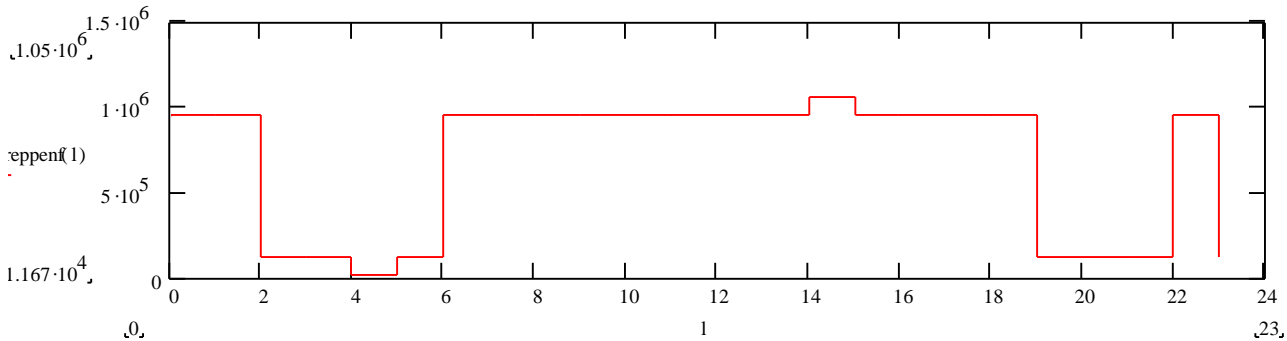
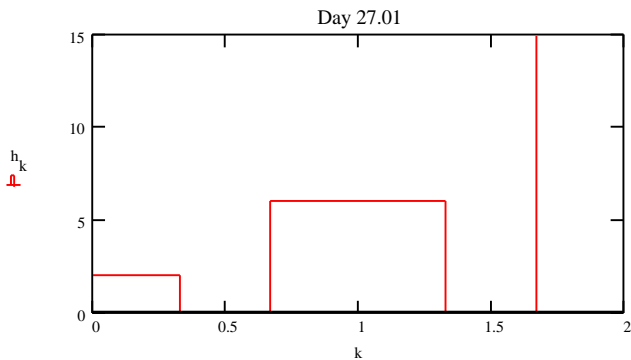
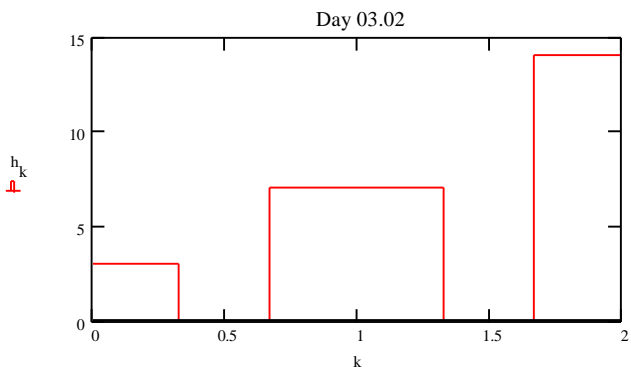


Figure 7: Temporal pattern based on QoS thresholds
In figure 8, the patterns for two days are based on the QoS threshold histogram (section 2). It relates the QoS values to specific thresholds buckets.



2 Measures lie within Interval 0 and thresh0
6 Measures lie within Interval thresh0 and thresh1
15 Measures lie within Interval thresh1 and thresh2



3 Measures lie within Interval 0 - thresh0
7 Measures lie within Interval thresh0 and thresh1
14 Measures lie within Interval thresh1 and thresh2

Figure 8: Patterns based on QoS threshold histogram

4.3. Weekly QoS patterns

The weekly QoS patterns is described by the integration of the daily QoS patterns of the week. The analysis is aimed to compare the daily QoS patterns in order to study performance bottlenecks and resource scheduling strategies for the different days of the week. We have analysed the QoS patterns of two weeks and have identified that there are similarities in the daily pattern behaviour considering specific days of the week.

Especially the calendar effect which is addressed in econometrics is identified in the two weeks – i.e. sunday and saturday. Further the patterns of some specific days are much more similar than the patterns of other days. At the end of the second week there is a QoS outlier describing the minimum delay. Normally minimum delay per time scale is stable and reflects the propagation delay, but in cases of routing and system troubles the minimum delay could take some unexpected peak value like in this case.

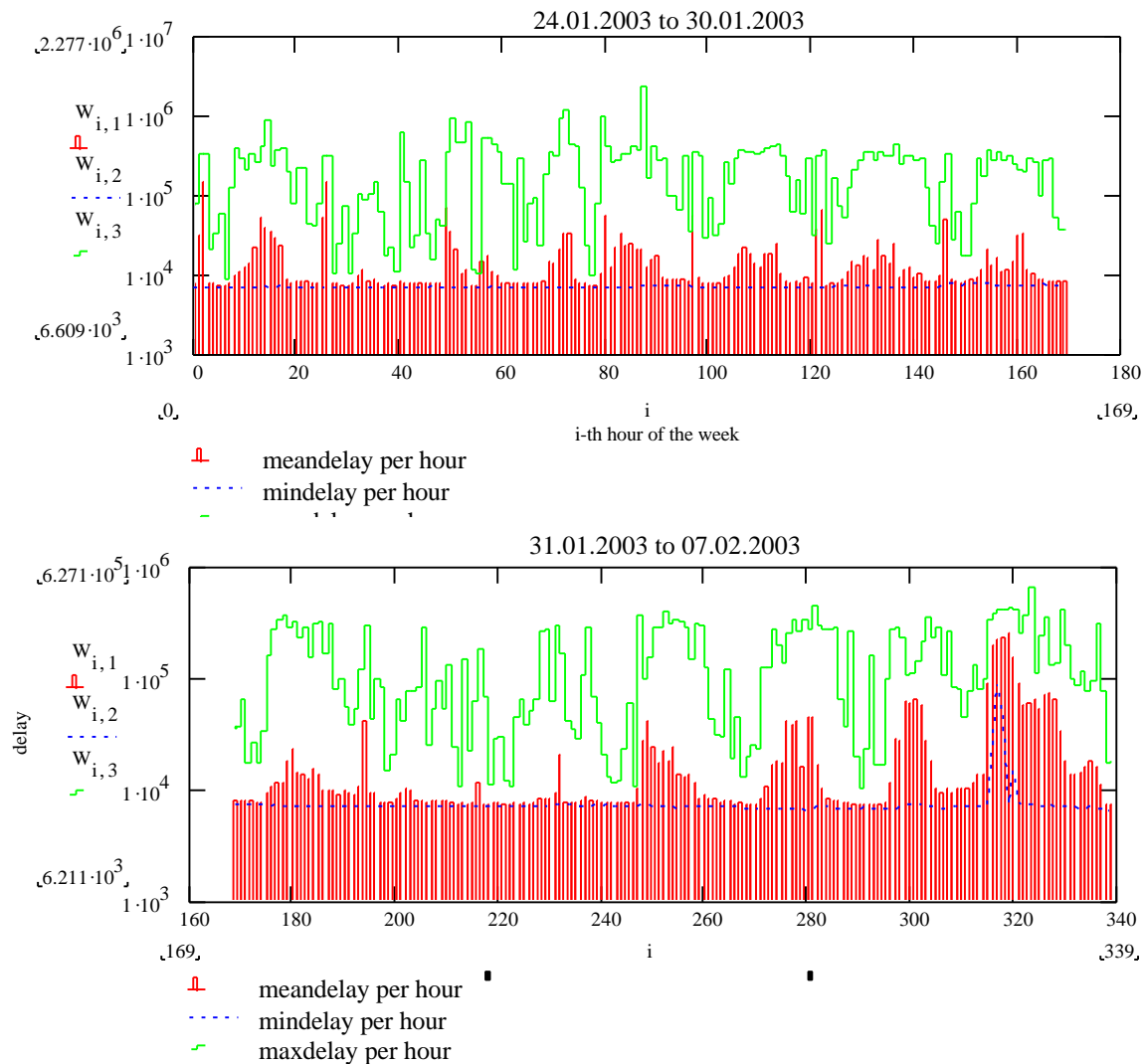


Figure 9: Weekly QoS pattern for minimum, mean and maximum delay

4.4. Detailed QoS pattern analysis

It is useful to compare patterns describing QoS behaviour of different time intervals in order to characterise in detailed manner congestion or network anomalies.

In the following figure, the pattern analysis is based on the mean end-to-end delay QoS data obtained per second.

This pattern shows the time exactly when a QoS peak value happens and could be used for dependency analysis on a fine grained time scale.

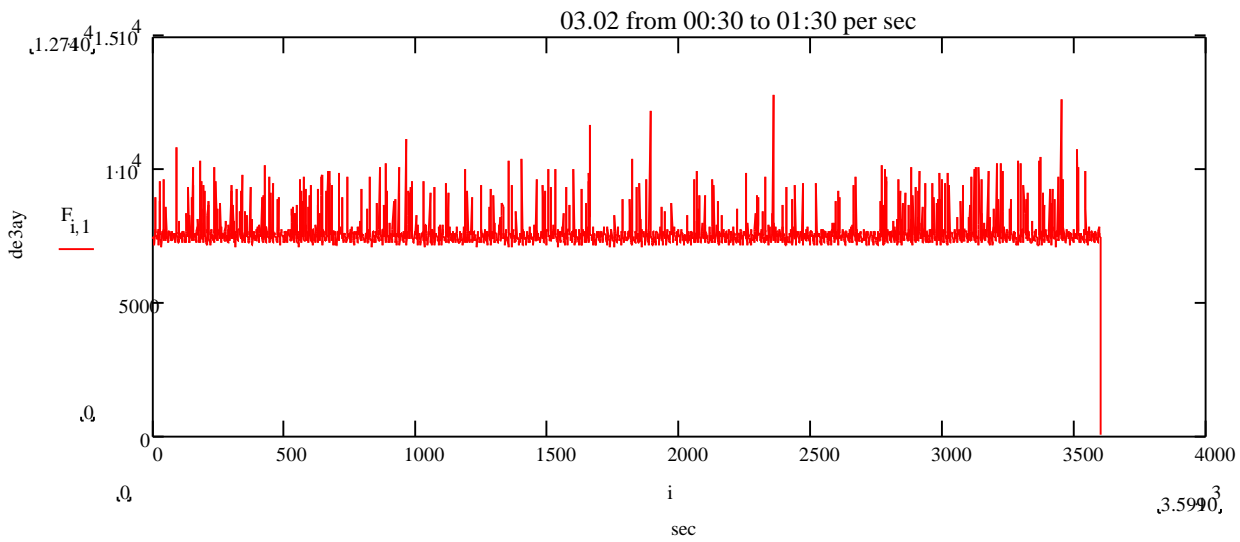


Figure 10: Detailed end-to-end mean delay QoS pattern (time series data behaviour in second intervals)

5. Conclusions and further work

A methodology to use QoS patterns for short term (daily and weekly) planning, load balancing and outlier detection considering spatio-temporal relationships of QoS time series was discussed. Pattern comparison and analysis using statistical techniques and QoS threshold histograms was proposed.

The feasibility of the QoS analysis using patterns was shown in a real world experiment. The studies are based on a real QoS measurement data obtained during more weeks for an end-to-end connection between access network and ISP service provider. [BC 96]

Further work is aimed at:

- Spatio-temporal and spatial composition studies in inter-domain environment considering patterns.
- Combination of spatio-temporal QoS pattern analysis with forecasting techniques based on learning historic QoS patterns and fuzzy classification.
- Integration of the spatio-temporal QoS analysis in the INTERMON [INTERMON] toolkit for capacity and network planning.

Pattern detection could be integrated in forecasting scenarios. Current work on traffic prediction applying pattern matching and fuzzy classification could be considered [HC 99].

Generally, the QoS pattern analysis technology could be integrated as preliminary step in QoS forecasting scenarios, because it is aimed to identify patterns such as “outlier” which are due to abnormal behaviour or to identify “congestion” patterns which could be improved by better capacity planning and network dimensioning (including of additional resources) and need not to be considered for forecasting.

6. References

[AMM 03] A. Anzaloni, I. Miloucheva, E. Müller, A practical approach for QoS forecasting considering outliers, Inter-domain performance and simulation Workshop, February, 2003

[AR 97] Abraham, T. and Roddick, J.F. ‘Research issues in spatio-temporal knowledge discovery’. In Proc. SIGMOD’97 Workshop on Data Mining, Arizona, USA, 1997

[AR98] Abraham, T. and Roddick, J.F. ‘Opportunities for knowledge discovery in spatio-temporal information systems’. Aust. J. Inf. Syst. 5(2):3-12

[AQUILA] AQUILA consortium, “Adaptive Resource Control for QoS Using an IP-based Layered Architecture”, AQUILA homepage: <http://www.ist-aquila.org>

[BC 96] Berndt, D.J. and Clifford, J. Finding patterns in time series: A dynamic programming approach. In Advances in Knowledge Discovery and Data Mining, Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., and Uthuramy, R., Eds., Menlo Park, CA: AAAI Press, 1996, 229-248.

[CKT 02] Chen-Mou Cheng, H.T.Kung, Koan-Sin Tan, Use of Spectral Analysis in Defense Against DoS Attacks, GLOBECOM 2002

[CL 03] Les Cortell, Connie Logg, Throughput Time Series Patterns (Diurnal and Step Functions), <http://www.slac.stanford.edu/comp/net/pattern/diurnal.html>

[CHNY 02] M.J.Coates, A.O.Nero, R. Nowak, B.Yu, Internet tomography, review article, IEEE Signal Processing Magazine, May, 2002

[DIR 99] D. Donjerkovic, Y. Ioannidis, and R. Ramakrishnan. Dynamic histograms: Capturing evolving data sets. Technical report, Department of Computer Science. University of Wisconsin-Madison, 1999.

[FRM 94] Faloutsos, C., Ranganathan, M., Manolopoulos, Y. Fast Subsequence Matching in Time-Series Databases. In Proceedings of SIGMOD ’94 (Minneapolis MN, May 1994), ACM Press, 419-429.

- [GP 94] N.K. Groschwitz and G.C. Polyzos, "A Time-Series Model of Long-term Traffic on the NSFNET Backbone," Proc. IEEE International Conference on Communications (ICC'94), New Orleans, LA, pp. 1400-1404, May 1994.
- [GKS 01] S.Guha, N.Koudas,K.Shim.Data Streams and Histograms. ACM Symposium on Theory of Computing,2001
- [HC 99] Heribert Kirschfink, Claire Chadenas, Traffic Situation Prediction Applying Pattern Matching and Fuzzy Classification, European Symposium on Intelligent Techniques, June 3-4, 1999 Orthodox Academy of Crete, Greece
- [HMH 02] D. Hetzer, I. Miloucheva, U. Hofmann, J. Quittek, F. Saluta, Integrated Information System for Inter-Domain QoS Monitoring, Modelling and Verification, EURESCOM Konferenz , Oct. Heidelberg, 2002
- [HMSP 02] U. Hofmann, I. Miloucheva, F. Strohmeier, T. Pfeiffenberger, Evaluation of architectures for QoS analysis of applications in Internet environment, The 10th International Conference on Telecommunication Systems, Modelling and Analysis Monterey, CA, USA, October 3-6, 2002
- [INTERMON] Advanced architecture for INTER-domain quality of service MONitoring, modelling and visualisation,INTERMON-IST-2001-3412 <http://www.ist-intermon.org>
- [JPKS 98] H.V. Jagadish, V.Poosala, N.Koudas, K.Sevcik, S. Muthukrishnan, T.Suel.Optimal Histograms with Quality Guarantees.Proc. 24th Int. Conf. Very Large Data Bases, VLDB,1998
- [MC 00] S. McCreary and K. Claffy, "Trends in Wide Area IP Traffic Patterns --- A View from Ames Internet Exchange", Proceedings of ITC'2000.
- [Pow 99]R. S. Povinelli. The time series data mining: identifying temporal patterns for characterisation and prediction of time series events, Dissertation, Wisconsin 1999
- [OSGC 95] T. Oates, M.D. Schmill. D.E. Gregory, P.R. Cohen. Detecting complex dependencies in categorical data. In Finding Structure in Data : Artificial intelligence and Statistics, p.185-195, 1995
- [OC 96] T. Oates, P.R. Cohen, Searching for structure in multiple streams of data. In Proceedings of the Thirteenth International Conference on Machine Learning, 346-354, 1996
- [NLANRSurveyor]<http://dast.nlanr.net/Articles/measurements/surveyor.html>
- [MH 01] I. Miloucheva, U. Hofmann, Distributed Measurement and Monitoring in IP Networks (CMToolset for AQUILA DiffServ), SCI 2001 /ISAS 2001 Orlando 7/2001
- [MNH 02]I. Miloucheva (ed.), Traffic Measurement Roadmap, NGNI Deliverable, October 2002
- [PMHN 03] T. Pfeiffenberger, I. Miloucheva, U. Hofmann, A. Nassri, Inferencing of inter-domain path characteristic based on active end-to-end QoS monitoring - scenario studies in emulation environment, Inter-domain Performance and Simulation Workshop, Salzburg, 2003
- [PMFT 02] K.Papagiannali, S.Moon, C. Fraleigh, P. Thiran, F. Tobagi, C. Diod, Analysis of Measured Single-Hop Delay from an Operational Backbone Network, in Proceedings of INFOCOM 2002, June 2002, New York
- [RS 99]John F. Roddick, Myra Spiliopoulou, A Bibliography of Temporal, Spatial and Spatio-Temporal Data Mining Research *SIGKDD* Explorations. Copyright 1999 ACM SIGKDD, June 1999. Volume 1, Issue 1 – page 34
- [SDH 01] F. Strohmeier, H. Dörken, B. Hechenleitner. AQUILA Distributed QoS Measurement. In Proc. of COMOCON8

Conference, Crete, Greece, pages 177-185, June 2001.

[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 , INET Proceedings, 2000, http://www.isoc.org/isoc/conferences/inet/00/cdproceedgs/li/li_2.htm

[TTO 02] M. Tsuru, T.Takine, Y.Oie, Inferring arrival rate statistics of each flow from aggregated flow rate measurements, 2002 <http://www.genesis.tao.go.jp/pub/TR/measure.html>

[YSJ 00] B.-K. Yi, N.D. Sidiropoulos, T. Johnson, H.V. Jagadish. C. Faloutsos. A. Biliris, Online Data Mining for Co-evolving Time Series, ICDE 2000