Adaptable bandwidth planning using reinforcement learning

Dirk Hetzer
T-Systems International
dirk.hetzer@t-systems.com
Berlin, Germany

ABSTRACT

In order to improve the bandwidth allocation considering feedback of operational environment, adaptable bandwidth planning based on reinforcement learning is proposed.

The approach is based on new constrained scheduling algorithms controlled by reinforcement learning techniques.

Different constrained scheduling algorithms,, such as “conflict free scheduling with minimum duration”, “partial displacement” and “pattern oriented scheduling” are defined and implemented.

The scheduling algorithms are integrated into reinforcement learning strategies. These strategies include:

- Q-learning for selection of optimal planning schedule using Q-values;
- Informed Q-learning for exploitation and handling of prior-knowledge (patterns) of network behaviour;
- Relational Q-learning for improving of bandwidth allocation policies dynamically in operational networks considering actual network performance data.

Scenarios based on integration of the scheduling algorithms and reinforcement learning techniques in the experimental monitoring and bandwidth planning system called QORE (QoS and resource optimisation) are given.

The proposed adaptable bandwidth planning is required for more efficient usage of network resources.

1. INTRODUCTION

In order to support QoS based applications with guaranteed bandwidth requests, there is a need of tools planning the allocation of the available resources on efficient way to the applications. Where for long term bandwidth planning, there are forecasting methods such as ARIMA proposed for planning of resources [1], [2], [3], it is currently a lack of technologies and modelling approaches considering the dynamic of the environment for planning in short and mean term periods.

For efficient resource planning, it is a need of optimization techniques (operation research methods) using learning approaches (reinforcement, supervised learning) to consider automatically monitoring feedback. Such combination of optimization and learning is used to adapt the bandwidth planning strategies to the requirements of the traffic flows and measured performance parameters.

A lot of QoS monitoring and performance measurement architectures was proposed and used in Internet, such as AQUILA [4] and INTERMON [5]. However, currently the monitoring data bases of such architectures are not linked to bandwidth planning tools to provide bandwidth allocation dependent on the measurement data.

Adaptable planning using reinforcement learning, which is proposed in this paper, discusses new technology and tools for automated bandwidth planning based on reinforcement learning of performance feedback.

The goal is the optimal allocation of network bandwidth considering bandwidth reservation requests of QoS oriented applications, as well as performance data of measured traffic for the network connection.

The bandwidth planning tool for QoS and resource optimization (QORE) is developed for adaptable bandwidth planning enhancing the INTERMON QoS monitoring architecture [6]. The goal of QORE is optimization of the resource allocation based on interaction with monitoring technologies and performance databases. The tool uses performance patterns and reinforcement learning techniques to support optimal scheduling of resources considering monitoring data.

The next section is aimed to discuss the reinforcement learning architecture for adaptive bandwidth planning. In section 4, different scheduling algorithms are proposed, which are integrated in the reinforcement learning strategies.

Section 5 presents scenarios for bandwidth planning using QORE. Section 6 concludes this paper.

2. BANDWIDTH PLANNING APPROACH USING REINFORCEMENT LEARNING

The bandwidth planning approach using Reinforcement Learning (RL) is aimed to find optimal policy for bandwidth allocation in advance for QoS oriented applications using rewards from environment, such as QoS parameters. The rewards allow the computation of an optimal schedule, which considers measured QoS parameters, such as delay of the best effort traffic

2.1. Optimization using resource requests for reservation in advance

The application of RL is possible, because the different kinds of traffic (best effort and QoS oriented applications with advance resource requests) are using the same dynamic resource pool.

Optimization of resource allocation based on advance resource reservation is studied for different applications. In [7], advance reservation for Grid is proposed. “Alternative” calls allowing variable reservations are considered in [8].

The concept of adaptive bandwidth planning includes more flexible framework for specification of advance resource reservation based on measurement feedback [9], [10}.

Especially, the “interval” based reservation requests allow flexibility of allocation in specific interval, in order to optimize costs and provide better resource utilization and QoS provision in Internet [11].

The RL approach for bandwidth planning is based on an intelligent agent, which interacts with operational environment to:

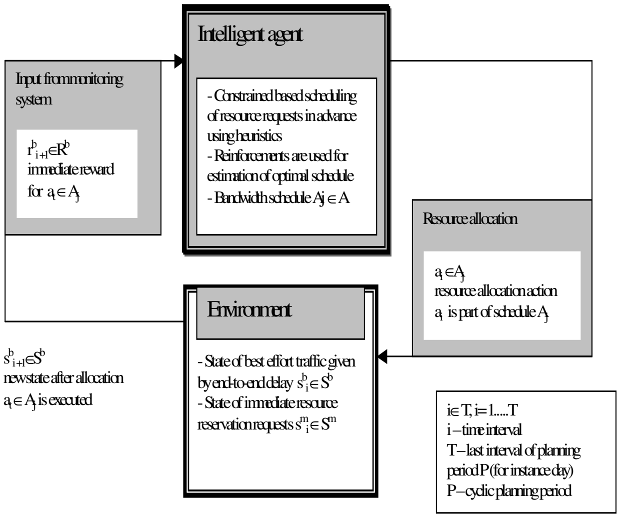
- Estimate rewards and environment states using input from monitoring system. Such inputs are monitored QoS parameters of best effort traffic.
- Perform bandwidth allocation actions for resource reservation in advance, which are part of bandwidth schedules. As result of the actions, the states of environment are updated.

The role of the reward function is to provide feedback to the RL algorithm about the effect of the resource allocation action of the particular bandwidth schedule. Based on this feedback, the RL algorithm could search for optimal scheduling policy.

2.2. Markov Decision Process

To describe the RL problem for bandwidth planning, the Markov Decision Process (MDP) could be used as formal framework.

The planning period P is divided in fixed time intervals 1, 2,…I,… ., T. At each time interval, the state transition is done, based on resource allocation actions, using rewards from environment.


Intelligent agent
- Constrained based scheduling of resource requests in advance using heuristics
- Reinforcements are used for estimation of optimal schedule
- Bandwidth schedule Aj ∈ A
Input from monitoring system
$r^b_{i+1} \in R^b$ immediate reward for $a_i \in A_j$
Resource allocation
$a_i \in A_j$ resource allocation action a_i is part of schedule A_j
Environment
- State of best effort traffic given by end-to-end delay $s^b_i \in S^b$
- State of immediate resource reservation requests $s^m_i \in S^m$
$s^b_{i+1} \in S^b$ new state after allocation $a_i \in A_j$ is executed
i∈ T, i= 1.....T
i – time interval
T – last interval of planning period P (for instance day)
P – cyclic planning period

Fig, 1: MDP for bandwidth planning

The MDP for bandwidth allocation planning is a finite horizon process, with a goal state S_T. The formulation is based on time divided into intervals, where a time interval t starts at time t -1 and ends at time t.

Similar approach of scheduling problems for other applications is proposed in [12]. The scheduling combined with RL have been studied extensively in [13] and [14]. Immediate rewards from the environment r_i, i= 1…T, are used to calculate the cumulative sum at the goal state S_T. The cumulative sum is used for the bandwidth allocation.

2.3. Bandwidth allocation schedule

Consider the bandwidth allocation problem at time t in state s_t. The agent selects an allocation action a_t belonging to bandwidth allocation policies (schedules) $A(R)_i$, i = 1…k, i.e. $a_t \in A_i$ and in turn receives a reward r_t. As a result of this action, the system moves to a new state s_{t+1}.

From this new state, the agent selects another action a_{t+1} according to A_i. Consequently, another reward r_{t+1} is received. This process continues and results in a particular sequence of future rewards, which have been generated by a bandwidth scheduling policy.

An empty action is applied, when no allocation is to be done at some time interval.

The bandwidth schedule A(R), constrained by resources R, is a composite action for resource allocation in a given planning period P with time steps 1…T. It is defined by A(R) = [a $_1$, a $_2$,…, a $_n$], where $a_{i\,,}$ i= 1, n are actions ordered in time 1,…, T.

2.4. RL problem formulation

The RL problem for optimal bandwidth allocation is intended to maximize the cumulative reward function for the planning period P, considering rewards r $_{I+1}$ from environment, based on QoS parameters of best effort traffic and accepted immediate resource reservation requests. The immediate rewards are results of actions for resource allocation, a_i, ∈ A(R) , for advance resource requests, at each time interval.:

Given

- A set of states s ∈ S describing environment with a goal state s_T ∈ S.
- A set of bandwidth allocation schedules A with A(R) ∈ A, A(R) = [a $_1$, a $_2$, ….a $_n$], and actions for resource allocation a $_j$ are executed at time interval t, where t = 0…T
- An unknown transition function δ: S × A -> S,
- An unknown real-valued reward function r: S × A -> R.
- Find a bandwidth schedule A(R) *: S -> A that maximizes a value function V *(S_t) for the goal state S_T ∈ S, where $V^* = E(\sum_{t=1}^{T} r_t \ \ )$.

  The value function V* is based on the cumulative sum of rewards r for a bandwidth schedule A(R) , obtained for the goal state S_T.

The RL problem formulation is based on the goal-oriented planning concept.

This ensures that once the learning agent reaches the goal state S_T ∈ S, it remains in that goal state. The cumulative sum of rewards is delayed, and calculated at the end of each planning period S_T.

2.5. Using patterns for state description

The states s ∈ S describe the QoS monitoring values at the specific time interval.

The question is how to present the state space based on the QoS monitoring values in order to design useful RL strategy.

The state space of the values of QoS parameters is large. To avoid large computations, abstraction of state space, based on patterns, is used.

The "threshold exceeding" QoS parameter pattern classifies the measured QoS parameter values in exceeding some maximum or minimum.

Considering "threshold" patterns, each state could be shown to be either in threshold overload, underutilization, or normal state. Similar way considering highly and low overload, as well as highly and low under-load state spaces, is used in a RL problem specified for storage assignment to applications [15].

2.6. Reward function

The reward function describes the specific planning utility, which is used for optimization considering the transition from state s_{i-1} , to the next s_i using selected scheduling action. Its calculation is based on the monitred QoS parameters, such as end-to-end delay of aggregated best effort traffic, collected during the state transitions.

The simple method for reward function is to define it based on the value of the QoS parameter $D(s_i)$ at the goal state s_i, and describe on this way the impact of the action.

Considering thresholds for acceptable QoS parameter values of the best effort traffic, the following reward function using values of –1, 0, 1, could be defined:

$$r_i = \begin{cases} 1, \text{ if } s_i = \text{normal } (L_{th} < D(s_i) < O_{th}) \\ 0, \text{ if } s_i = \text{underutilisation } (L_{th} > D(s_i)) \\ -1, \text{ if } s_i = \text{overload\_state } (O_{th} < D(s_i)) \end{cases}$$

where

L_{th} is the low end-to-end delay threshold, below which the resources are underutilised. In this case, is assumed 0 value for the reward.

O_{th} is the maximum end-to-end threshold characterizing QoS. Exceeding it, the QoS level of the best effort traffic is unacceptable for the user. The reward is negative, when the delay exceeds the overload threshold O_{th}.

A positive reward (1) is supplied for a state s_i , when the delay of the best effort traffic $D(s_i)$ is in normal state.

The reward function could be defined in more sophisticated way considering the structure of the QoS parameter behaviour, i.e. threshold overload patterns, of the best effort traffic.

3. REINFORCEMENT LEARNING ALGORITHMS FOR OPTIMAL PLANNING

Algorithms are aimed to find the most appropriate bandwidth schedules for proactive and reactive bandwidth planning.

The algorithms are designed using following approaches:

- Model-free RL estimating Q-value of the optimal bandwidth schedule. In first step, the whole set A (R) of conflict-free bandwidth schedules, which is possible to obtain for given restrictions R, is calculated. Because no knowledge of the environment is used, the search of the bandwidth schedule with maximal Q-values could be slow.

- Informed RL for handling of prior-knowledge of pattern classes and corresponding bandwidth allocation policies. Pattern based schedules are considered, which are derived based on the predicted QoS parameter patterns of the best effort traffic. Using of predictions makes this approach well suited for the proactive planning

- Relational RL based on run-time planning and reasoning of bandwidth allocation policies considering specific state structures (patterns) for improved value function approximation. It allows dynamically in on-line manner to enhance the set of bandwidth schedule and to find the schedule, which is more appropriate for the actual states patterns. It evaluates the Q-value of the on-line learned schedule and stores this schedule in the knowledge database. The approach is useful for reactive bandwidth planning. Similar to informed Q-learning, it uses a priori knowledge of environment patterns, as for instance outlier, to detect patterns for bandwidth scheduling.

Important for all these solutions is the calculation of initial conflict-free bandwidth allocation with minimum duration using appropriate heuristics. Using patterns and reinforcements from environment, following conflict-free schedules are obtained and used for optimal planning:

- Pattern based schedules, which allocate at each time scale the resources based on connection resource restrictions, considering in addition patterns for QoS parameter behaviour of the best effort traffic.
- Partial displacement schedule. Given an initial conflict-free schedule with minimum duration A_{min}, a partial displacement schedule A' is built by displacement of all resource allocations in intervals $t_k \ldots t_s$ to the later starting time.

In a partial displacement schedule A', no allocation actions are assigned to the repair interval $t_k \ldots t_s$

A' is a partial displacement schedule defined in respect to an initial conflict free schedule with minimum duration A_{min} , if for at least one time interval $t_{i\ldots}$ t_k of the schedule A_{min} hold that all allocation actions a_j , j = 1,…,s in this interval are re-allocated to an interval with starting time $t_i' > t_i$ in A'.

Let $r_k \in$ R, i=1,…l, is resource request in advance, which is rescheduled to later allocation begin, then all dependent resource requests of r_k are also rescheduled using the given constraints.

It should be noted that also other concepts for definition of partial displacement scheduling are discussed. [16] and [17] define partial scheduling for the case of failure, when jobs of the initial minimum schedule (critical path schedule), must be reallocated.

The learning approaches for optimal bandwidth scheduling differentiate in the set of bandwidth schedules, which are used for learning and for which the system computes Q-values.

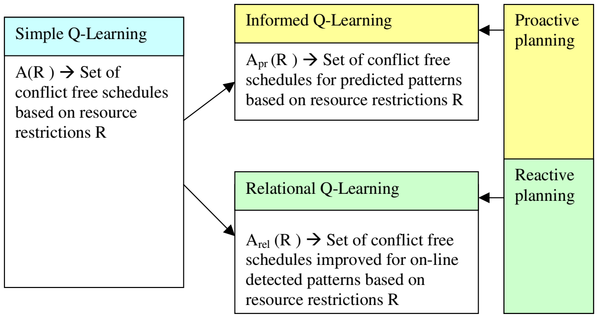

Simple Q-Learning
A(R) → Set of conflict free schedules based on resource restrictions R
Informed Q-Learning
A_{pr} (R) → Set of conflict free schedules for predicted patterns based on resource restrictions R
Relational Q-Learning
A_{rel} (R) → Set of conflict free schedules improved for on-line detected patterns based on resource restrictions R
Proactive planning
Reactive planning

Fig. 2: Summary of reinforcement learning algorithms for bandwidth planning

4. QORE SYSTEM FOR QOS AND RESOURCE OPTIMISATION

QORE is a tool for adaptable QoS-oriented proactive and reactive bandwidth planning. It is designed and implemented based on reinforcement learning concepts.

The main focus of the QORE architecture is to automate the calculation of resource allocation strategies in advance considering feedback from QoS parameter behaviour of best effort traffic. Patterns describing QoS parameter behaviour are stored and classified, based on which scheduling algorithms could be improved to consider the changes in the environment.

QORE system includes components and knowledge databases, which support functions for bandwidth allocation planning, such as:

- Specification of advance resource reservation requirements, such as bandwidth resource requests, dependency of applications, allocation times.
- QoS parameter monitoring based on distributed measurement agents for obtaining of effective bandwidth for applications and scenarios, as well as for QoS parameter pattern analysis of aggregated best effort traffic.
- Scheduling algorithms for bandwidth allocation based on heuristics.
- Resource constraints simulator.
- Effective bandwidth estimator for applications and scenarios.
- Pattern analyser for obtaining patterns of QoS parameters (delay, delay jitter and packet loss) of best effort traffic.
- Detection of outliers and their filtering.
- Visual data mining of optimal resource schedules with feedback from environment.

Common Graphical User Interface accesses the functions of QORE components and the database repository.

The QORE components are interacting using information repository, i.e. knowledge database, which store information for measurement and bandwidth allocation scenarios together with their results.

QORE components and their interaction with the database are shown in Figure 3:

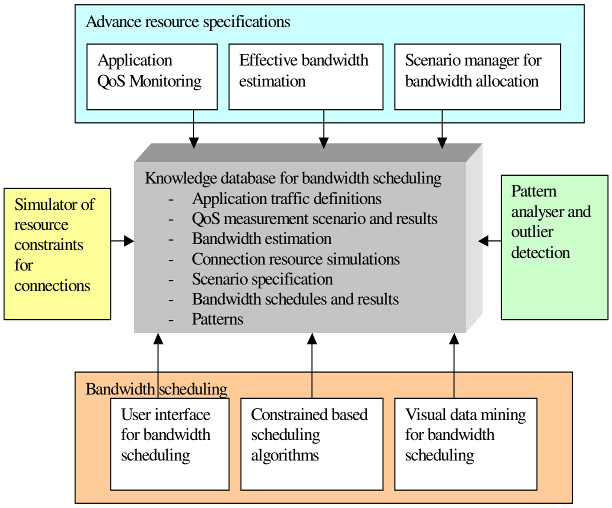
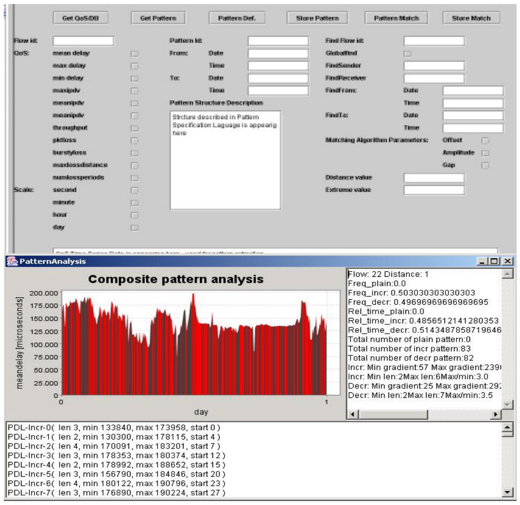

Advance resource specifications
Application QoS Monitoring
Effective bandwidth estimation
Scenario manager for bandwidth allocation
Knowledge database for bandwidth scheduling
- Application traffic definitions
- QoS measurement scenario and results
- Bandwidth estimation
- Connection resource simulations
- Scenario specification
- Bandwidth schedules and results
- Patterns
Simulator of resource constraints for connections
Pattern analyser and outlier detection
Bandwidth scheduling
User interface for bandwidth scheduling
Constrained based scheduling algorithms
Visual data mining for bandwidth scheduling

Fig. 3: QORE components and interaction with database

QoS parameter patterns for the network connections are processed by:

- Outlier detection tool used to eliminate abnormal behaviour from the QoS parameter structure
- Pattern analyser for detection of "threshold overload" patterns, which could be used to adapt the bandwidth schedule [18].

Figure 4 shows the GUI of the QORE pattern analyser for QoS parameter and "threshold overload" patterns. In addition, an example for extraction of end-to-end delay pattern of the best effort traffic is given.



Strcture described in Pattern Specification Laguage is appearig here

```
Flow: 22 Distance: 1
Freq_plain:0.0
Freq_incr: 0.503030303030303
Freq_decr: 0.49696969696969695
Rel_time_plain:0.0
Rel_time_incr: 0.4856512141280353
Rel_time_decr: 0.5143487858719646
Total number of plain pattern:0
Total number of incr pattern:83
Total number of decr pattern:82
Incr: Min gradient:57 Max gradient:239
Incr: Min len:2Max len:6Max/min:3.0
Decr: Min gradient:25 Max gradient:29
Decr: Min len:2Max len:7Max/min:3.5
```

```
PDL-Incr-0( len 3, min 133840, max 173958, start 0 )
PDL-Incr-1( len 2, min 130300, max 178115, start 4 )
PDL-Incr-2( len 4, min 170091, max 183201, start 7 )
PDL-Incr-3( len 3, min 178353, max 180374, start 12 )
PDL-Incr-4( len 2, min 178992, max 188652, start 15 )
PDL-Incr-5( len 3, min 156790, max 184846, start 20 )
PDL-Incr-6( len 4, min 180122, max 190796, start 23 )
PDL-Incr-7( len 3, min 176890, max 190224, start 27 )
```

Fig. 4: GUI for pattern analyser and daily pattern example

The optimal conflict free bandwidth schedule is obtained using different methods

- Conflict free schedule with minimum duration using heuristics;
- Partial displacement schedules based on extracted patterns, which are aimed to derive optimal allocation schedules keeping the QoS parameters of best effort traffic at acceptable level defined by thresholds.

5. SCENARIOS FOR ADVANCE RESERVATION USING QORE

Different bandwidth allocation schedules could be obtained in QORE using heuristics and learning of performance data. The conflict free bandwidth schedule satisfies all user service requests in advance and the set of policies required by the provider. Such advance bandwidth reservation schedules are obtained considering the duration of the bandwidth reservation of applications, the bandwidth requests in advance and the possible times for allocations of the requests [10].

For provision of more dynamical and flexible planning, as well as to consider the impact of the environment, reinforcement learning techniques are used. In particular, reinforcement learning strategies are applied to find automatically the most appropriate conflict-free bandwidth plan based on learning of performance and traffic behavior. The learning algorithms are based on the calculation of initial conflict-free bandwidth schedule with minimum duration.

A Q-value evaluating the selected schedule is obtained using the performance data of the network connection (QoS parameters), when the selected bandwidth allocation is applied. Based on learning of the Q-values of different bandwidth allocations, the best one is selected for the network connection.

Different reinforcement learning strategies are used in QORE.

Model-free Q-learning is based on evaluation of the Q-value of the possible conflict-free bandwidth schedules, which satisfy the requests in advance by the applications. This approach uses no prior knowledge for selecting possible schedules. An example for a daily bandwidth planning is given in figure 5.

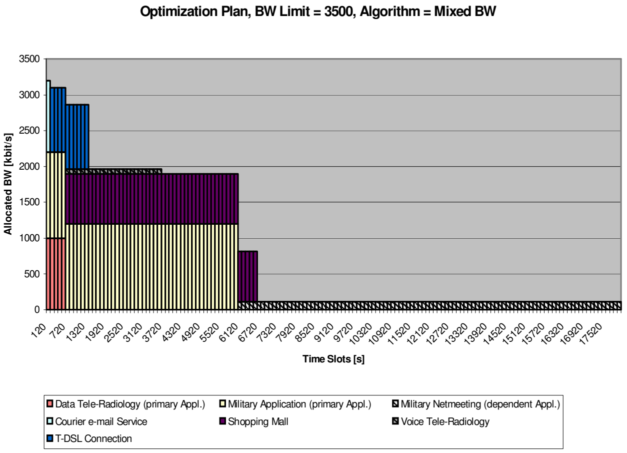


Fig. 5: Bandwidth plan based on Q-learning

Informed Q-Learning restricts the sets of the possible advance resource reservation schedules using the predicted performance patterns of the operational traffic, which is sharing the total resources of the connection with the advance reservation applications (see figure 6).

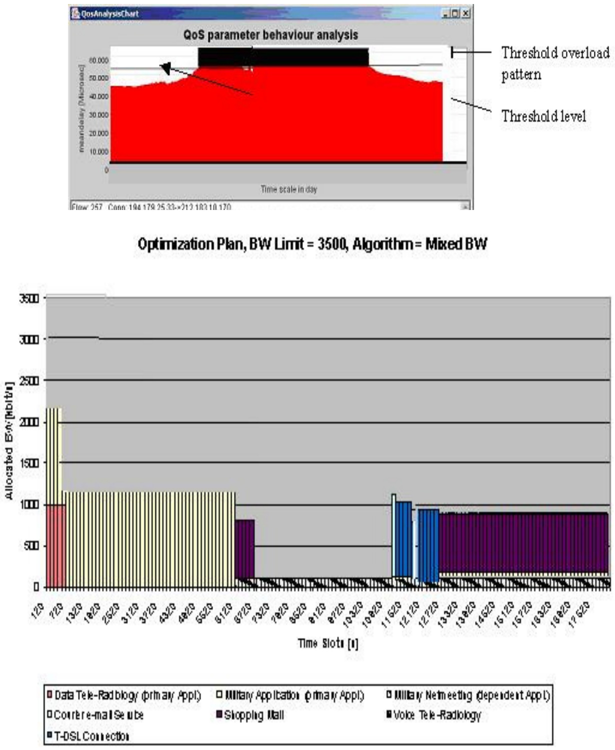


Fig. 6: Bandwidth plan using informed learning

The forecasting of performance (resource) patterns makes this approach well suited for the proactive planning. Figure 6 shows the change of the model-free Q-learning schedule considering predictions of resources of operational traffic.

Relational Q-learning allows dynamically based on operational traffic load to update the advance bandwidth reservation schedules to consider more efficiently the actual situation of the network. The method allows reactive planning, i.e. change of the bandwidth plan allocations taking into account the load in the operational networks (see figure 7).

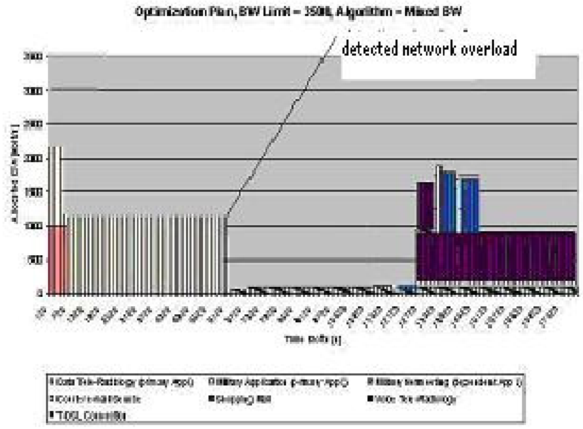


Fig. 7: Updated bandwidth plan using relational learning

6. CONCLUSIONS AND FURTHER WORK

This paper discusses the theoretical background of adaptive bandwidth planning using reinforcement learning for optimal scheduling considering QoS parameter patterns as feedback from environment. Based on RL approach, a practical usable system called QORE for automated bandwidth planning was developed and discussed.

The QORE system could be enhanced in different directions for designing of more powerful bandwidth planning architecture improving the current state-of-the art of Internet management concepts for QoS and SLA support.

Particular goals for further usage of adaptive bandwidth planning technology and QORE tool is to enhance the current QoS broker concepts. Currently, the DiffServ bandwidth brokers control the resource reservation without to c onsider advance reservation and optimal planning. The focus of further research is integrated bandwidth brokerage architecture including advance reservation for QoS oriented applications.

The business aspects of learning for QoS provision and efficient resource utilisation are important issue for bandwidth allocation planning. QORE architecture has a great potential for enhancement of Management Information Bases (MIBs) concerning allocation of bandwidth for Telecom users. New services could be integrated based on concepts for resource reservation in advance in the Telecom infrastructure. Especially, customers of multimedia data transfer, real time embedded services and Grid applications could take benefit for the integration of QORE system in the practical management.

7. References

[1] 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.

[2] N.K. Groschwitz, G.C. Polyzos, "A Time-Series Model of Long-term Traffic on the NSFNET Backbone", **IEEE International Conference on Communications (ICC'94)**, New Orleans, LA, page 1400-1404, May 1994

[3] J. Ilow, “Forecasting Network Traffic Using FARIMA Models with Heavy Tailed Innovations", **ICASSP**, Istanbul, Turkey, June 2000.

[4] U. Hofmann, I. Miloucheva, “Distributed Measurement and Monitoring in IP Networks (CMToolset for AQUILA DiffServ)”, **IEEE Networking Conference**, Orlando, June 2001.

[5] D.Hetzer, U. Hofmann, I.Miloucheva, J. Quittek, F.Saluta, “ INTERMON Integrated Information System for Inter-domain QoS Monitoring, Modelling and Verification”, **EURESCOM Summit 2002 , Powerful Networks for Profitable Services,** Heidelberg/ Germany, 21 - 24 October 2002.

[6] I. Miloucheva, P.A. Gutierrez, D. Hetzer, M. Beoni, “INTERMON architecture for complex QoS analysis in inter-domain environment based on discovery of topology and traffic impact,” 2nd International Workshop on **Inter-domain Performance and Simulation,** Budpaest, Hungary, February, 2004.

[7] A. Galstyan, K. Czajkowski, K. Lerman, “Resource Allocation in the Grid Using Reinforcement Learning”, **in Third International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 3 (AAMAS'04),** New York City, New York, USA, 19-23 July 2004.

[8] T. Erlebach, “Call Admission Control for Advance Reservation Requests with Alternatives**”, Eidgenössische Technische Hochschule Zürich,** Swiss Federal Institute of Technology Zurich, TIK-Report Nr. 142, July 2002

[9] D. Hetzer, I. Miloucheva, “Adaptable bandwidth planning for enhanced QoS support in user-centric broadband architectures”, **World Telecommunications Congress**, Mai 2006.

[10] D. Hetzer, I. Miloucheva, K. Jonas, “Resource Reservation in Advance for content on-demand services, **IEEE Conference Networks 2006**, New Delhi, India, November 6 - 9, 2006.

[11] D. Hetzer, K. Rebensburg: “Advance resource reservation for adaptable bandwidth planning”, **IEEE SOFTCOM Conference**, Split, Croatia, Sept. 2005.

[12] M.E. Dyer, L.A. Wolsey, “Formulating the single machine sequencing problem with release dates as a mixed integer program”, **Discrete Applied Mathematics**, 26, page 255-270, 1990

[13] J.P. De Sousa, L.A. Wolsey, “A time-indexed formulation of non-preemptive single-machine scheduling problems”, **Mathematical Programming,** 54, page 353-367, 1992

[14] M. Van den Akker, C. P. M. Van Hoesel, M. W. P. Savelsbergh, “A polyhedral approach to single machine scheduling”, **Mathematical Programming,** 1997.

[15] V. Sundaram, P. Shenoy, A Practical Learning-Based Approach for Dynamic Storage Bandwidth Allocation, K. Jeffay, I. Stoica, and K.Wehrle (Eds.): **IWQoS 2003, LNCS 2707**, pp. 479–497, 2003.

[16] W. Zhang, T. Dietterich, ”A reinforcement learning approach to job-shop scheduling” in **Proceedings of the 14th International Joint Conference on Artificial Intelligence,** 1995.

[17] M. Zweben, B. Daun, M. Deale, “Scheduling and rescheduling with iterative repair”, in **“Intelligent Scheduling”** M. Zweben and M. S. Fox (eds), Chapter 8, page 241-255, 1994.

[18I. Miloucheva, D. Hetzer, A. Naasr,, “Data mining approach to study Quality of Voice over IP Applications**”, Data Mining Conference**, Malaga, Spain, 2004.