Operation research techniques of QORE system for inter-domain capacity engineering

Dirk Hetzer
T-Systems Nova Deutsche Telekom
Goslarer Ufer 35
D-10589 Berlin, Germany
dirk.hetzer@t-systems.com

Abstract

This paper is focused on the Quality of Service (QoS) and Resource Optimization (QORE) toolkit for capacity planning of routers and edge systems in an Internet environment. QORE toolkit is aimed at efficient allocation of resources for a wide range of QoS enabled applications and traffic flows in order to meet QoS requirements and Service Level Agreements (SLA).
QORE concept integrates measurement based traffic and application performance modeling with operations research methods and algorithms.
Special focus of the paper are operations research algorithms for QoS and resource optimization using QORE common measurement and modeling data base for application QoS and traffic flows.
Scenarios for capacity planning and resource allocation with different optimization constraints and criteria based on QORE toolkit are discussed.

1. Introduction

With the wide deployment of QoS provisioning technologies (DiffServ, IntServ, MPLS) and QoS based networks and services (ATM, UMTS) in today advanced telecommunication infrastructures, as well as with the introduction of emerging multimedia and mobile QoS enabled distributed services and applications, the need of tools and concepts for QoS and resource optimization planning is becoming an important aspect of capacity management of network domains of Internet Service Providers (ISP).

Capacity management is dealing with network design and provisioning to meet performance objectives for traffic demands at minimum costs considering SLA and QoS requirements of traffic classes and applications as well as effective bandwidth estimations.

Optimization concepts being part of ISP capacity planning are specified with different goals, such as:

- core router resource assignment to traffic flows in order to increase performance for different applications and traffic classes
- inter-domain traffic allocation to border router interfaces at minimum costs
- distributed content server resource selection in order to optimize performance for users.

The usage of measurement based models describing performance characteristics of traffic and applications is the focus of different research activities addressing QoS and resource optimization, such as:

- QoS tuning and adaptation for application and traffic classes in specific networking environments [18]. Enhanced QoS measurement and monitoring techniques (active, passive, event based) as well as distributed measurement databases are used for the evaluation and validation of user network contracts (SLAs) and corresponding application or traffic QoS parameters.

- QoS based adaptive resource control for different traffic classes using Bandwidth Broker (BB) strategies for DiffServ traffic classes and corresponding SLAs and QoS requirements. A description of an example architecture facilitating QoS based adaptive resource control can be found in [1] and the references therein.
- QoS based traffic aggregation using special techniques such as Measurement Based Admission Control (MBAC) techniques for different kinds of QoS demanding applications.
- Inter-domain QoS analysis and traffic engineering [19]
- Backbone traffic engineering and optimization of router resources. For an example see [10].

Recently, the concept of measurement based modeling and optimization of inter-domain QoS interfaces in large scale telecommunication infrastructures using an integrated monitoring, modeling, and simulation architecture is proposed within the framework of the INTERMON project [12].

This paper discusses the QoS and Resource Optimization (QORE) concept and toolkit for capacity planning of routers and edge systems in an Internet environment. In QORE the measurement based modeling is used in two directions:

- to provide mapping between performance and resource utilization models (effective bandwidth).
- to supply input data for operations research algorithms aimed at the computation of variants for resource allocation plans considering performance and QoS optimization issues.

The main focus of QORE are novel operations research algorithms for capacity planning in telecommunications networks aimed at traffic flow assignment to router resources at minimum costs considering different synchronization, multiplexing and priority restrictions of applications as well as timing dependencies and requirements for resource usage.

The constraints for the operations research algorithms are derived from the demands of actual communication applications and the practical needs for resource allocation dependent on the characteristics of these applications.

A common QORE information data base designed to integrate different kinds of traffic engineering information (e.g. measurements, models) with optimization data for capacity planning (operations research algorithms and evaluated capacity plan variants) is the basis for the QORE optimization.

In particular, for the purpose of QoS and resource optimization, the design of the data base facilitates:

- mapping of traffic demands to performance models (QoS evaluations) and resource utilization estimates, i.e. effective bandwidths
- mapping of admission control strategies to traffic demands, performance and resource utilization models with special focus on MBAC
- mapping of performance and resource models to operations research constraints and planning variants, i.e. QORE capacity plans
- mapping of performance, resource, and admission control models as well as QORE capacity plans to Internet router and edge system resources.

This allows an efficient capacity planning technology based on

- comparison of planning variants for different optimization goals and constraints
- simulation of planning variants for different admission control strategies, traffic demands, QoS requirements and resource utilization estimates

- consideration of measurement and modeling information as feedback for enhancements of planning variants.

The paper is structured as follows. Section 2 describes requirements for QoS and resource optimization which are addressed by the QORE architecture. The architecture of QORE can be found in section 3 and section 4 gives an overview on the algorithms which are integrated into QORE. The sections 5 and 6 describe test scenario sets used for QORE optimization and show the achieved results. Section 7 concludes the paper.

2. Requirements for QoS and Resource Optimization Addressed by the QORE Architecture

There can be different issues for QoS and resource optimization in the Internet, depending on the scope chosen for optimization. The optimization architecture must address the specific router resources which are the object of optimization and to which the QoS requirements of application flows are mapped.

Demands exist for QoS support for specific multimedia applications within specific domains and among multiple internet domains. In order to cover the intra- and inter-domain issues for QoS and resource optimization, the scope of the QORE architecture is focused on the access or border routers (*access channels*) of core networks. The core networks could be Autonomous Systems (AS) in the Internet context, but also multiple interconnected network domains in the context of telecommunication providers or Virtual Private Networks (VPN).

Intra-domain QoS and resource optimization is aimed at optimizing the QoS of application scenarios within a specific domain focusing on the optimal selection of ingress/egress border router resources and their optimal interworking with the core-routing system. The objective of this optimization is:

- optimal assignment of QoS based applications to the resources of the ingress and egress domain access routers
- selection of optimal routing paths between the ingress and egress routing system
- optimization of the core router resource configurations.

Inter-domain QoS and resource optimization considers interconnected domains which can be based on different SLAs, QoS provisioning mechanisms and policies. The objective of inter-domain QoS monitoring is a macroscopic optimization of the application assignments to the resources of ingress/egress border routers of different interconnected domains in order to provide enhanced QoS and minimize costs for resource utilization. The inter-domain optimization is emphasizing on:

- optimal assignment of QoS based applications to the resources of the ingress and egress access routers considering different interconnections of sending and receiving domains
- selection of optimal routing paths between sending and receiving domains in cases when there is more than only one possibility to interconnect the domains
- optimization of ingress/egress border router interconnections.

Generally, a traffic flow could be characterized by a number of QoS metrics, e.g. bandwidth, delay, delay variation and others. However, using more than only one metric for characterizing a traffic flow may lead to a significant computation complexity, e.g. finding a route with two constraints is usually a np-hard problem [4]. Hence, for the sake of simplicity, only one metric – the bandwidth metric – is used within QORE for QoS and resource optimization.

Costs are assigned to resources at access channels dependent on economical criteria such as:

- resource type
- duration of resource usage compared with thresholds for expected resource utilization

- time intervals of resource usage, for instance night or day time usage of resource.

The efficiency of resource utilization depends on which interval the application is assigned to the resource. The cost could be minimized if the time interval for application allocation to the resources is selected comparing the cost factors.

The QORE concept is aimed at supporting QoS and optimization decisions for different application scenarios emerging from new multimedia applications in the Internet. It is designed to combine different requirements for QoS and resource optimization such as:

- bundling (multiplexing) of connections
- synchronization of multimedia streams
- multiple path selection
- multiple destination selection.

QoS and resource optimization algorithms have to obtain the appropriate number of multiplexed connections of the given traffic type. The optimal number of multiplexed connections has different aspects in network planning. Multiplexing can be of different types:

- multiplexing of connections of the same traffic type, e.g. Voice over IP (VoIP) connections
- multiplexing of different types of multimedia traffic
- multiplexing of connections of different types of traffic characteristics, e.g. adaptive and constant traffic type.

Timing requirements are derived from reservation timing restrictions and application precedence criteria. Application precedence criteria restrict the resource commitment dependent on the priority of applications, execution timing relationships of application, and application scenarios (parallel, precedence relations).

3. QORE Toolkit

Admission control and network dimensioning for cases of guaranteed QoS have attracted the attention of researchers. Successful approaches are closely related to the ability of quantifying the usage of resources on the basis of traffic modeling and measurements [8]. This approach is chosen by the QORE toolkit. The goal of QORE is to provide enhanced end-to-end QoS while optimizing resource usage policies, by making use of traffic modeling and distributed measurements as well as operations research algorithms and application requirements.

All relevant data, for instance resource type specification, resource usage, application specification, traffic models, etc. is stored in a database, the QORE QoS-DB.

3.1 Introduction to QORE toolkit

The QORE toolkit is designed for measurement based and application service oriented QoS and resource optimization in the Internet [11], [9]. QORE uses optimization algorithms based on operations research[1] and has been designed to fulfill requirements for optimization in the context of

- operational management and capacity planning of Internet resources (bandwidth) at access points (border routers, exchange points, peering points and edge routers)
- simulation and modeling of resources and QoS in Internet networks and Autonomous Systems (domains).

As QoS based routing is aimed at finding the optimal path for various QoS based applications, QORE is aimed at finding an optimal assignment of resources to paths, i.e. the assignment of application scenarios to paths and the optimal

[1] Operations Research is the science of optimal decision and planning using mathematical methods.

dimensioning of network elements belonging to paths.

QORE is intended to be used by ISPs for capacity planning in order to optimize the resource usage at edge systems and border routers regarding the provision of QoS for applications, especially when considering requirements of multimedia applications and resource hungry applications. Another operational area could be the reduction of costs for Virtual Private Networks (VPNs) in enterprises by optimizing with respect to resource assignment to access points, transmission cost and duration.

3.2 Basic Concepts of QORE

The main focus of the QORE toolkit is enhanced end-to-end QoS and resource optimization considering different application types and scenarios. It is based on distributed measurement agents interacting with the QoS measurement database for storage of QoS evaluations per communication application and scenario. Using this database, the QoS measurements are mapped to network resource reservations dependent on the network specifics (e.g. usage of RSVP, IP DiffServ, IP TOS field or IPv6 flow label). The “optimal” QoS to resource mappings are selected for different networking technologies and used as input for optimizations for scenarios in Internet core networks.

The QORE system is based on an object-oriented specification of :

- *applications* with their traffic and protocol configuration parameters, as well as requirements for QoS parameter evaluation
- *application scenarios* for measurement of application QoS and resource utilization defining requirements for synchronization and multiplexing of applications as well as for resource reservations (effective bandwidth)
- *scenario sets* describing relationships of application scenarios (priorities and timing considerations and application execution dependencies)
- *resource* and *effective bandwidth specifications* for application scenarios
- *communication channels* (i.e. access networking elements like border routers, exchange points, edge routers, peering points) with resource (i.e. bandwidth) characteristics
- *QoS* and *resource optimization* strategies (i.e. algorithms) applied to scenario sets.

A graphical user interface (GUI), written in Java, is used for the specification of the above items and provides access to the optimization functions for optimal resource assignment.

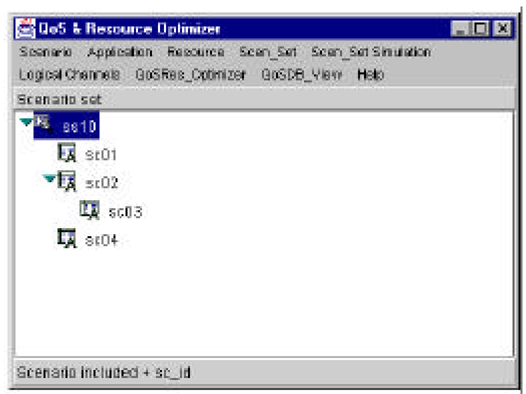

Figure 1: QORE GUI showing a scenario set comprising four scenarios

Figure 1 depicts the main window of the QORE GUI. It shows a scenario set, *ss10*, which consists of four scenarios, *sc01* to *sc04*. The indentation of *sc03* represents the temporal dependence of *sc03* on *sc02*, i.e. *sc02* has to be executed prior to *sc03*. Each of the scenarios can comprise several different applications, with resources assigned per scenario. The resource usage of the applications is characterized by effective bandwidths. This value can be assumptive and can later be substituted by the correct value retrieved by simulation of the scenario.

The scenario set can now be optimized with respect to:

- least administrative cost
- best QoS
- shortest duration (fastest execution) of the given scenario set
- most efficient resource utilization.

The new approach of QORE compared with existing measurement tools for Internet communication protocols and applications is the QoS and resource optimization.

3.3 Components of QORE

Figure 2 shows an overview of the different components QORE consists of. As can be easily seen, the central component is the QoS measurement and optimizations database, as it stores all the relevant information contained in the “logical” components, or blocks, of QORE, which are the following:

- **Application and QoS measurement scenario**
  This component is used to specify applications which should be measured together with their QoS and resource requirements dependent on the application types and protocol tuning parameters (e.g. TCP window size).

- **Resource assignment**
  To every scenario, a certain resource of specific type (ATM, UMTS, DifServ, MPLS LSP, etc.) and bandwidth limitation can be assigned and specified with QORE.

- **Access channel configuration**
  A logical “access channel” (i.e. border router, edge router, exchange point, peering point) can be specified, which is used to carry the traffic generated by the previously specified application scenarios.

- **Scenario set configuration**
  A scenario set is used to group a number of application (measurement) scenarios which may be dependent on each other, which may have different priorities or which have to be executed at a specific time.

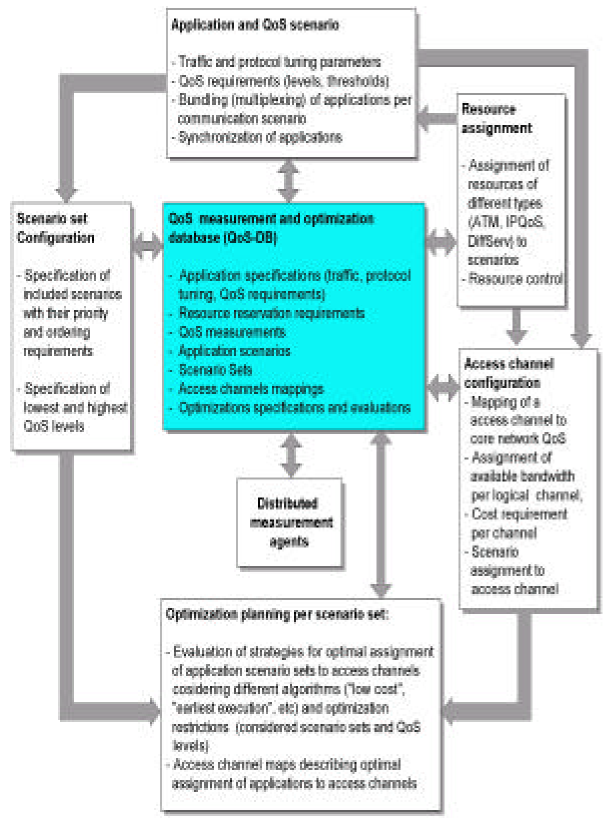

Figure 2: Components of QORE

- **Optimization Planning**
  In this component, a certain bandwidth limitation of the access channel with respect to a specific scenario set together with one of several optimization algorithms can be defined. This is used as input to the QORE optimization function based on operations research methods, which tries to find the best temporal

ordering of application scenarios, leading to efficient resource utilization of the access channel (the bandwidth limit must not be exceeded) and a short execution duration of the complete scenario set.

- **Distributed measurement agents**
  Distributed measurement agents are used to collect data regarding the QoS level, which the traffic flows generated by the applications have experienced.

Other ideas on which QORE is based on, such as the scalability of measurements, measurement data bases, and distributed active and passive measurements, can be found in actual measurement architectures as shown in Table 1.

QORE Concepts	Related Work
Active measurement	CAIDA Skitter and related work (see [7], [3], [5], [6]) Surveyor (see [14])
Performance metrics and their measurement	Surveyor IPPM performance metrics and related work (see [13], [2], [17])
Performance monitoring methodology using active and passive measurements	NLANR Network Analysis Infrastructure (see [15])
Scalable and distributed measurement architecture	NIMI (see [16]) AQUILA (see [1])

Table 1: QORE concepts compared to current research

4. QORE Optimization as an Operations Research Problem

In QORE, heuristic methods derived from operations research are used to obtain minimum costs and/or earliest execution for the different optimization tasks. Their goal is to find the optimal ordering of scenario executions with respect to least administrative cost, best QoS, minimum bandwidth usage and shortest execution time. These algorithms are more efficient in terms of computational overhead and consideration of relevant restrictions compared to linear programming approaches.

4.1 Simple "core" Algorithm

The concept the QORE algorithmic solution is based on a main "core" algorithm which can be flexibly adapted using heuristics for different kinds of restrictions and goal function parameters. Communication scenarios in QORE's algorithmic approach are considered as jobs. Their duration, priority, requirements of execution order, QoS and resource mapping, as well as resource usage, is obtained from the description of the scenario set stored in the QoS-DB, to which the optimization plan, i.e. the optimization specification, is applied.

Following minimum restrictions are assumed in the "simple" core algorithm:

- simple QoS and resource mapping
- assumption of one channel
- temporal dependence of applications (precedence relation).

Based on this, the Operations Research problem formulated by

$$\max E = \max \sum_{1}^{N} E_i ,$$

where

$$E_i = \frac{D_i \cdot Q_i}{S_i \cdot R_i} ,$$

with

E_i	–	efficiency metric of job *i*
D_i	–	duration of job *i*
Q_i	–	QoS level of job *i*
S_i	–	Starting time of job *i*
R_i	–	Resource usage of job *i*
N	–	Number of jobs.

Where the global resource restriction GR is fulfilled:

$$R = \sum (R_i \cdot D_i) \leq GR.$$

The time restrictions (precedence relations) of communication jobs can be formulated by

$$S_k > E_l,\ k = 1\ldots n,\quad l = 1\ldots n,\quad n \leq N,$$

where S_k is the start time of job k and E_l represents the end time of job l. According to this inequality, job k can only be executed after job l, because its starting time is greater, i.e. later in time, as the end time of job l.

The idea behind the simple "core" algorithm is to determine the "bundle" of applications (i.e. multiplexed application set) satisfying the time precedence restrictions on one access channel at each iteration step of the algorithm. One iteration step covers the time interval determined by the least common denominator of all application durations. In each iteration step the algorithm tries to calculate the bundle of applications using the channel resources which is expected as "optimal cutting" of the channel resources. This algorithm is illustrated in Figure 3.
The simple "core" algorithm is based on two techniques used in operations research, namely *optimal cutting* and *critical path*, with consideration of special QORE restrictions.

4.2 Critical Path Method

The Critical Path Method (CPM) is focused on finding the ordering of jobs restricted by their duration and ordering relationships.

The earliest execution time in terms of CPM is specified as follows:

$$FEZ_j = FAZ_j + D_j$$

$$FAZ_j = \max_{i=1\ldots n}(FEZ_i)$$

FAZ_j is the earliest start time of application j, FEZ_j is the earliest end time of application j and FEZ_i is the earliest end time of applications i= 1...n preceding j.

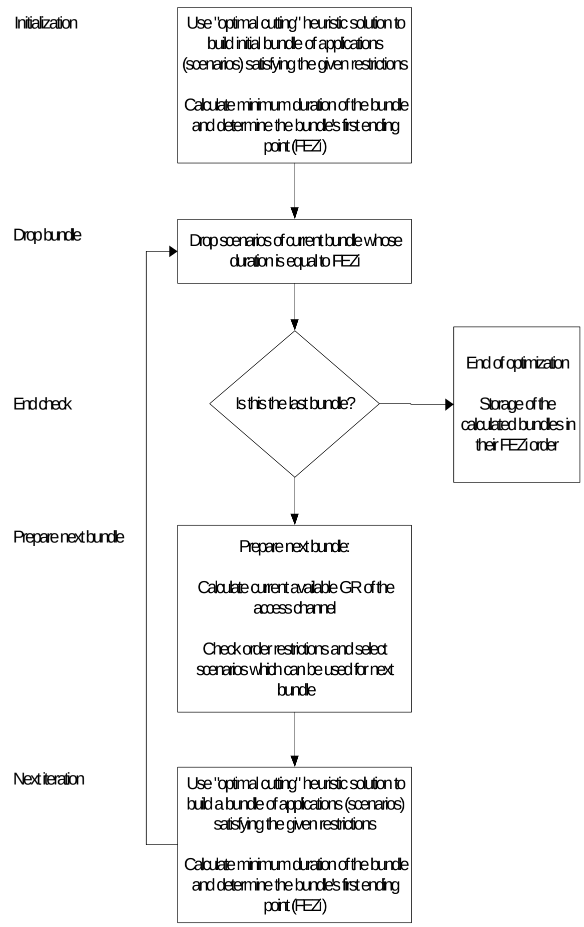

Figure 3: Simple "Core" optimisation algorithm

The algorithm used in QORE is a modification of the original CPM, because the optimal duration must be calculated considering the required restriction parameters in each iteration step. Such restrictions are used resource, QoS level, application and scenario priorities, cost and timing requirements.

4.3 Optimal Cutting

The optimal cutting (assignment) method is used on each iteration step to select the set of applications and scenarios, so as the sum of required resources optimally “fills” the pipe of available resources i.e. causes only a minimum of wasted resource.
Following optimal cutting algorithms can be used for optimal multiplexing of applications under resource restrictions, i.e. a fixed bandwidth limit:

- **Small Bandwidth**
  Resource assignment for the applications and flows is based on their ordering by increasing resource requirements starting with the applications with minimum bandwidth demands. This bundling strategy is suited for optimization of applications with low resource requirements, for instance, Voice over IP.

- **Large Bandwidth**
  Allocation of resources for applications and flows is based on their ordering by decreasing resource requirements starting with the application with a maximum resource requirement. Such bundling is used for giving priority to resource intensive applications (such as bulk data transfer).

- **Mixed Bandwidth**
  For allocations, applications and flows with increasing and decreasing resource requirements are selected in exchange. Used for building of heterogeneous bundles of applications.

- **Small Bandwidth Preferred**
  Preferring applications which use a small amount of the available bandwidth and trying to allocate in appropriate intervals also applications and flows with larger bandwidth.

- **Large Bandwidth Preferred**
  Preferring applications which use a large amount of the available bandwidth and trying to allocate in appropriate intervals also small bandwidth applications.

One of these five algorithms can be selected per optimization plan to be used in optimizing the scheduling of traffic flows.

Furthermore, it can be selected, whether a certain application and traffic flow scenario has to be assigned to resources at a specific start time or during a specific time range (indicated by earliest possible start time or latest possible end time) or whether the scheduling does not depend upon time. The first two cases are called “fixed” scheduling dependence, the latter one is called “simple” scheduling dependence.

Figure 4 shows different kinds of combinations of the “core” optimization strategy with different algorithms (heuristics).

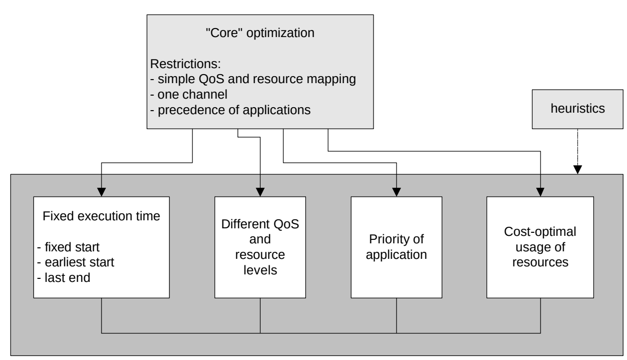

Figure 4: Concept of QORE optimization algorithms

5. QORE Optimization Plans

Using the QORE database, scenarios for obtaining optimization plans for capacity planning are discussed.

Scenarios are described with MBAC characteristics such as traffic models and effective bandwidths for aggregated traffic. The effective bandwidth values for these scenarios were not based on calculations using the respective traffic model, but were simply assumed at this time. This has no effect on the optimal cutting algorithms, which are assumed here. QoS levels, priorities and administrative costs have not been defined in the scenarios, thus the optimization is only focused on the optimal temporal scheduling order.

5.1 Obtaining Optimization Plans for Traffic Aggregates

Table 2 lists the scenario set used to test the optimization algorithms, comprising seven scenarios. For each scenario an appropriate traffic model, its effective bandwidth and its duration have been specified. These parameters are the input for the QORE optimization. The optimization algorithms are using the effective bandwidth and the duration of each scenario to obtain an optimal resource assignment.

The application scenario pairs *Data tele-radiology / Voice tele-radiology* and *eB Video streaming / eB Netmeeting* are assumed to be dependent, which means that those scenarios marked with *(d)* have to executed after their corresponding *(p)* scenarios have been finished.

Aggregated Traffic (QORE Scenario)	Traffic Model	Effective BW [kbps]	Duration [s]
Data tele-radiology (p)	Empirical	1000	600
Voice tele-radiology(d)	On-off-Fluid	110	3600
eB Video streaming (p)	Periodic	1200	7200
eB Netmeeting (d)	On-off-Fluid	115	10800
Courier e-mail service	Empirical	1000	120
Shopping mall data	Empirical	700	7200
ADSL WWW	FBM	900	1200

Table 2: QORE test scenario set

5.1.1 Optimization Plan Visualization

Figure 5 shows an example of the visualization of an optimization plan in QORE. Each "Duration" represents the time interval, in which the mix of scenarios (bundle) for applications and traffic flows remains constant. Also, the bandwidth consumption of each single scenario, *kz01* to *kz07* is shown.

Although this visualization includes all relevant data, it is not suited to get an overall picture. Thus, for the purpose of better understanding the output of the QORE optimizations, visualizations were made using Microsoft Excel.

The relevant data was imported from the QORE DB via Excel's ODBC data import feature. Future work might include the implementation of a VBA (Visual Basic for Applications) macro in MS Excel to simplify visualization tasks, or to implement a visualization feature directly in the QORE GUI.

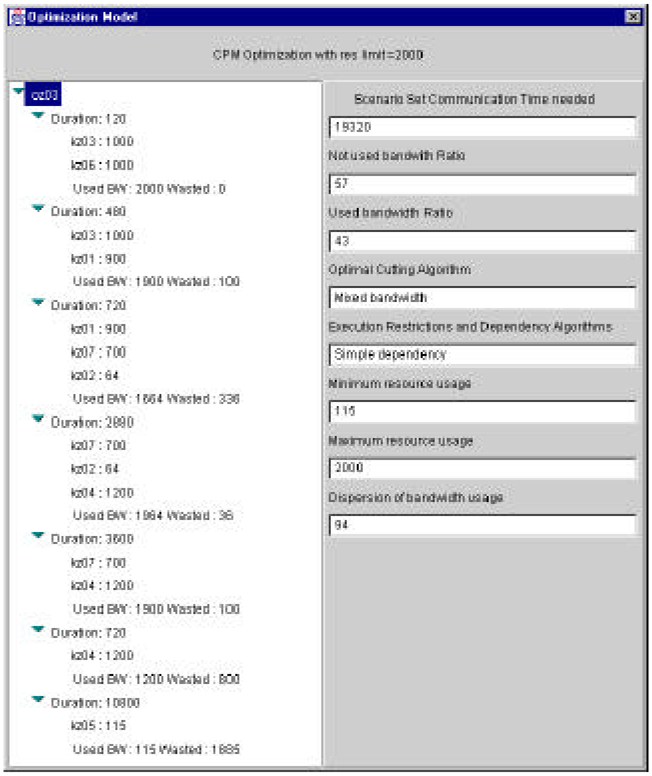

Figure 5: Example of a QORE optimization plan evaluation

Figure 6 depicts the QORE visualization of resource allocation scenario.

It shows the bundling of applications and traffic flows together with the consumed bandwidth vs. time. The bandwidth limit for the whole traffic scenario was set to 3500 kbps and the "Mixed Bandwidth" optimal cutting algorithm was flowschosen.

It can be seen that the dependent s are correctly scheduled after the primary scenarios. These dependence constraints worsen the bandwidth usage ratio. If, e.g. the Netmeeting traffic flow was not dependent on the video streaming, the execution time of the complete scenario would have been reduced by one third, approximately.

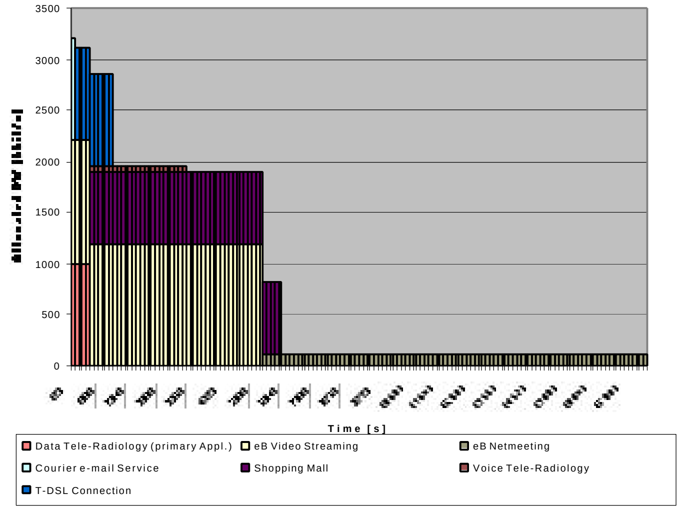

Figure 6: QORE optimization using the „Mixed Bandwidth" algorithm

5.1.2 Comparison of the Different QORE Optimization Algorithms

For the scenario set listed in Table 2, the optimization plans were compared for the five different optimal cutting algorithms and for five different total bandwidth limitations ranging from 1500 kbps to 3500 kbps.

In this section no restrictions for the start time of scenarios for the capacity planning are required. In section 6, further optimization strategies are analyzed based on the setting of fixed starting times for application scheduling. The fixed starting time of application and traffic flow assignment describes the requirements for start relative to the begin of the resource allocation plan.

The optimization plans were examined with respect to following characteristics:

- bandwidth usage ratio
- bandwidth limit of access channel
- duration.

The bars in Figure 7 represent the bandwidth usage in percent, large values represent better utilizations. The x-axis shows the complete scenario set duration for each algorithm, values on the left of each algorithm represent shorter duration times, i.e. are better.

As can be seen, the different algorithms produce similar results. The "Large Bandwidth" and "Large Bandwidth Preferred" algorithms show the best resource utilization for small bandwidth limitations (1500 kbps) while still having a short execution duration. It is obvious that the bandwidth usage ratio is increasing with decreasing bandwidth limit.
Dependent on the capacity planning goals at the access channels (border and edge routers, exchange and peering points), the best fitting QORE optimization algorithm (i.e. plan derived from the algorithm) could be selected by the ISP operator. Table 3 gives an overview of different possible optimization goals for capacity planning and the appropriate optimization algorithms.

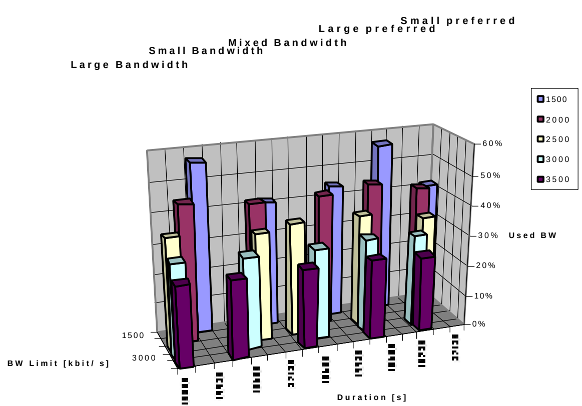

Figure 7: Comparison of the different optimization algorithms

Optimization Goals for Capacity Planning	Appropriate QORE Optimization Algorithm
Small bandwidth dispersion in order to better support inclusion of new applications with long durations (or to free access resources for enough long time).	Small Bandwidth Preferred algorithm
Best resource utilization	All algorithms have similar behavior with the smallest resource limit of 1500 Kbps
Best application QoS delivery	All algorithms except lowest bandwidth algorithm at resource limit 2500 Kbps
Optimal resource usage and QoS delivery	All algorithms at 2000 Kbps have similar behavior and show a similar level of resource utilization and QoS delivery

Table 3: Finding the optimal algorithm

6. Comparing Resource Allocation for Applications and Traffic Flows with Fixed Starting Times

As mentioned in chapter 4.3, it can be selected, whether or not the optimization procedure considers the optionally specified execution times of applications and traffic flows.
The following analysis is aimed to study the impact of aggregated traffic flows with fixed starting times. The analysis is performed for different bandwidth restrictions of the access channel:

- 2100 kbps
- 2500 kbps
- 3000 kpbs

The different kinds of optimization algorithms discussed in chapter 4.3 are compared for the following two cases of scheduling dependencies within the scenario set:

- dependencies based on simple precedence of scenarios

- dependencies based on simple precedence of scenarios and fixed start restrictions.

Table 4 lists the ten applications / traffic flows together with their dependencies, bandwidth requirements, durations and starting times.

Applic. / Flow name	Dependent on	Effective bandwidth [kbps]	Duration [s]	Start time relative to begin [s]
ks00		200	90	
ks01		1000	120	
ks02		200	360	0
ks03		700	60	90
ks04		1200	120	120
ks05		500	90	
ks06		200	30	
ks07	ks 01	100	40	
ks08	ks 05	1000	60	
ks09		500	240	

Table 4: Scenario set for comparing simple and fixed start optimization

In Figure 8 optimization plans, which were determined by the optimization of the above scenario set, are shown with regard to the following parameters:

- duration of the optimization plan
- used bandwidth of the access channel
- bandwidth dispersion.

The bracket labeled "simple" indicates the optimization plans which were obtained without taking into account the fixed starting times. On the other hand, the bracket labeled "fixed" indicates the optimization plans which were produced by additionally taking into account the fixed starting times of applications and flows.

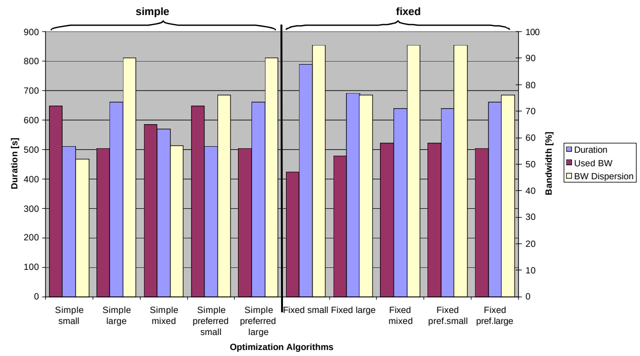

Figure 8: Comparison of simple/fixed optimization, BW limit = 2100kbps

It can be observed, that the bandwidth usage ratio is smaller for optimizations including scenarios with fixed start, which was expected, as the fixed algorithm adds constraints to the scheduling of applications and traffic flows. Furthermore, overall duration is longer and bandwidth dispersion is larger for optimizations taking into account applications and flows with fixed start.

Figure 9 shows the comparison for a bandwidth limit of 2500 kbps. The observed characteristics are very similar to the chart for 2100 kbps.

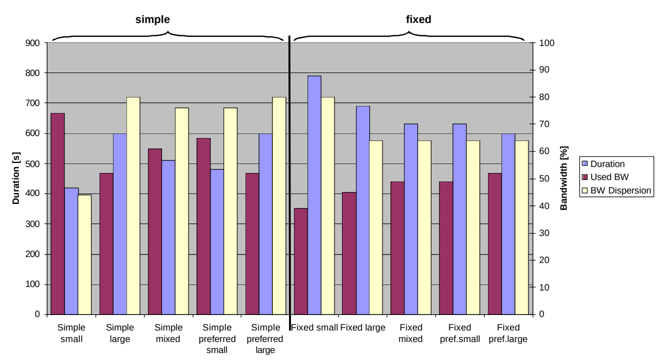

Figure 9: Comparison of simple/fixed optimization, BW limit = 2500kbps

And last, Figure 10 illustrates the optimization result for a bandwidth limit of 3000 kbps. Here, the "simple small" algorithm results in a very large bandwidth dispersion compared to the previous charts. As before, overall duration is longer for fixed than for simple optimization, as expected.

The optimization plans using "fixed start" strategies are very similar because of the restrictions given with the fixed start.

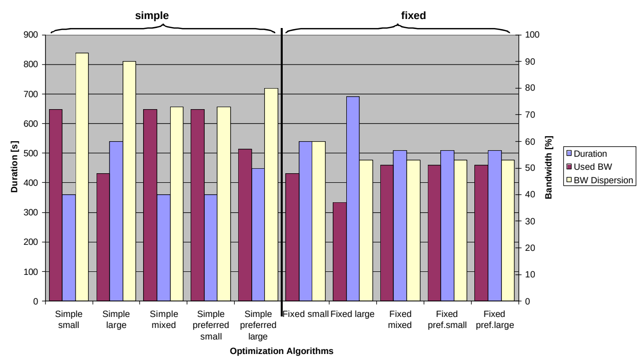

Figure 10: Comparison of simple/fixed optimization, BW limit = 3000kbps

7. Conclusions

In this paper, the QORE concept using operations research algorithms for the optimization of resource assignments of access channels (access, border and edge routers) under several constraints is proposed. Different algorithms for obtaining QoS and resource optimization are discussed.

The following conclusions considering the selection of algorithms for optimization in the described scenarios can be drawn:

1. As expected, the optimization plans based on fixed start usually show a larger duration for some scenarios, a smaller percentage of bandwidth usage and smaller bandwidth dispersion. This is due to the fact, that "fixed start" restricts the point of allocation of bandwidth to the scenario and this way reduces the possible optimization variants.

2. The fixed start of specific applications and traffic flows has also impact on the total bandwidth requirement.

   In the case of simple precedence of applications and traffic flows, it is required that $G_b > G_{max}$, where G_b is the total required bandwidth and G_{max} is the maximum bandwidth required for a scenario *i* belonging to the scenario set.

   In the case of fixed start of applications and traffic flows, it is required that $G_b > \mathrm{Sum}(G_{fi})$, where $\mathrm{Sum}(G_{fi})$ is the sum of bandwidth allocated for the maximum number of scenarios with fixed start that are executed at the same time.

   In the case discussed here, G_b for the optimization plans with fixed start is 2100 kbps, and G_b for optimization plans without fixed start is 2000 kbps.

3. Optimization plans with fixed scheduling strategies are similar and their dependence on the used algorithm (*Small Bandwidth*, *Large Bandwidth*, *Small Bandwidth Preferred*, *Large Bandwidth Preferred* and *Mixed Bandwidth*) is not so obvious, because of the restrictions given with fixed start.

Further research is the usage of the QORE system for border router resource modeling and simulation in the framework of INTERMON project [19], [12].

8. References

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

[2] Borella, M. S., D. Swider, S. Uludag, and G. B. Brewster, "Internet Packet Loss: Measurement and Implications for End-to-End QoS", In Proceedings of the International Conference on Parallel Processing (1998).

[3] CAIDA's Skitter project web page. http://www.caida.org/tools/measurement/skitt

er/

[4] Chen, S. and K. Nahrstedt, “An Overview of Quality of Service Routing for Next Generation High Speed Networks: Problems and Solutions”, *IEEE Networks*, vol. 12, no. 6, 64-79 (1998).

[5] Claffy, K., “Internet measurement and data analysis: Topology, workload, performance and routing statistics”, NAE 99 Workshop (1999).

[6] Claffy, K., G. Miller and K. Thompson, “The Nature of the Beast: Recent Traffic Measurements from an Internet Backbone”, INET'98 (1998).

[7] Cooperative Association for Internet Data Analysis (CAIDA), Homepage: http://www.caida.org

[8] Courcoubetis, Costas et al., “Application of the many sources asymptotic and effective bandwidths to traffic engineering”, *Telecommunication Systems*, 12, 167-191 (1999).

[9] Documentation and User’s Guide - QoS and Resource Optimization in the Internet (QORE System), T-Systems Nova (2001).

[10] Feldmann, A. et al., “NetScope: Traffic Engineering for IP Networks”, IEEE Network (2000).

[11] Hetzer, D., “QoS and Resource Optimization in the Internet (QORE System)”, Technical Report (2002).

[12] INTERMON consortium, “Advanced architecture for INTER-domain quality of service MONitoring, modelling and visualisation”, INTERMON homepage: http://www.ist-intermon.org

[13] IP Performance Metrics (IPPM) Working Group der IETF, Homepage: http://www.ietf.org/html.charters/ippm-charter.html

[14] Kalidindi, S. and M.J. Zekausus, “Surveyor: An Infrastructure For Internet Performance Measurements” (1999)

[15] McGregor, A., H-W.Braun, and J. Brown, “The NLANR Network Analysis Infrastructure”, *IEEE Communications Magazine*, vol. 38 (5), 122-128 (2000).

[16] Paxson, V., A. Adams and M. Mathis, “Experiences with NIMI”, In Proceedings of PAM 2000 (2000).

[17] Tobe, Y., Y. Tamura and H. Aida, “Detection of Change in One-Way Delay for Analyzing the Path Status”, In Proceedings of PAM 2000 (2000).

[18] M.Schmidauer, U. Hofmann, L.Villasenor, L.Lamont, I. Miloucheva: “Global Measurement Infrastructure of Application Classes and Their Bundling over ATM”, 4th European Conference on Multimedia Applications, Services and Techniques ECMAST´99 Madrid (Spain) 26-28 MAY 1999

[19] 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