ISSN 2229-5518

Resource Management for Online Demand Services

Aware Sachin B, Pansare Rushikesh B, Shevkar Abhishek C, Kulkarni Sumant S.

Abstract— Resource management offers Quality-of-Service reliability for time-critical continuousmedia applications. Currently, existing resource management systems in the Internet and ATM domain only provide means to reserve resources starting with the reservation attempt and lasting for an unspecified duration. However, for several applications such as video conferencing, the ability to reserve the required resources in advance is of great advantage. This paper outlines a new model for resource reservation in advance. We identify and discuss issues to be resolved for allowing resource reservation in advance. We show how the resource reservation in advance scheme can be embedded in a general architecture and describe the design and implementation of a resource management system providing reservation in advance functionality.

Index Terms— Resource Management for Online Demand Services, Resource Management,QoS For Online Demand Services.

---

1 Introduction

Computer systems used for continuous media processing must cope with streams having data rates of several Mbits/s and must provide timely processing guarantees. For instance, an endsystem shall synchronize audio and video streams up to a granularity of about 80 ms . Since available system resources are not abundant, applications have to be 'protected' such that they have access to the required resources in time. Otherwise the user will notice a glitch or drop in the presentation quality. Hence, means to manage the available system resources are necessary.

Resource management provides a way to offer applications reliability with respect to *Quality-of-Service* (QoS) . A resource management system controls the access to scarce system resources needed for audio and video data processing. It checks whether additional service requests can be satisfied, and if yes, the required resources are reserved for that application, else, the request is rejected. Sophisticated systems will allow for a negotiation according to the available capacities and constraints (e.g., by tariffs).

Requirements of Application Scenarios

Today existing resource management systems, for instance, HeiRAT , QoS Broker , Tenet , offer functions which only allow to reserve resources for a time interval which starts with the reservation attempt and which lasts for an unspecified time. For several application scenarios this model of immediate reservations is not appropriate. Consider, for instance, a virtual meeting room (conferencing) scenario supported by multimedia systems. Traditionally, a meeting will be scheduled for a specific time at a well defined location (room). To be sure that the respective room will be available at the scheduled time, a reservation entry, in some form of a meeting room calendar, is written *before* the meeting starts. The time between the reservation and the meeting itself can vary from short intervals,

e.g., half an hour or a few hours, to very long periods, e.g., months. In addition to 'one time events', meetings such as project meetings occur periodically. To support these 'virtual meeting room' scenarios the resource reservation system must offer mechanisms to reserve in advance the resources needed for the conference, i.e., certain capacities of networks, routers, and end-system resources.

Resource Reservation in Advance (ReRA) is not only needed for conferencing but for other scenarios such as video-on-demand as well. This resembles a video rental scenario where a user 'orders' a video for a specific time: for the video-on-demand system it means that the resources necessary to retrieve, transfer and present the video have to be reserved in advance,

i.e., video server, network, router, and end-system resources. Further application areas can also be found outside of typical multimedia applications, e.g., within manufacturing process control systems (where time-critical data must be processed and transmitted) or any kind of remote surgery in medicine.

Where resources are plentiful, not even immediate reservations may be necessary,but where resources are scarce enough to justify reservations at all, it makes sense to be able to make them in advance."

Classification of Reservation Types

To distinguish ReRA schemes from other reservation schemes, e.g., existing reservation techniques, we classify reservations based on two key factors:

- whether the resources are exploited at reservation time, and
- whether the reservation duration is known at reservation time.

The most stringent use of resource management is in the domain of process and control systems including embedded real-time systems. There, resources are reserved for the whole active phase of such systems,

i.e. for the lifetime. Changes can only be done at the initialization phase (and not at the actual run-time phase). Therefore we characterize such approaches as "static" opposed to the dynamic approaches. Traditional resource

ISSN 2229-5518

management systems (non-ReRA) assume that the resources are immediately used after they have been successfully reserved and no assumptions are made on the duration of the reservations. A ReRA scheme, on the contrary, is characterized by deferred resource usage and reservations of known duration (which might possibly be enlarged). In case of immediate usage and known duration, both schemes can be realized.

Architecture for resource reservation
Management of reservation

To allow for reservations in advance, the time axis is divided into slices. Within each slice a certain set of reservations exists and there is no change of this set or of the QoS parameters of these reservations, i.e., the reservation state is stable within each slice and changes only at the boundaries. Thus, the resource management system has a similar view as before: at a certain point in time (in a time slice) a fixed set of reservations with fixed QoS exist corresponding to a fixed resource utilization and free resource capacity. This view changes only if new reservations are established or existing ones end. Therefore, the following compo- nents of the resource management system need modification:

- The interface of the resource management system needs in addition to the QoS parameters now also specifications of the time parameters (begin and duration).
- These time values must also be contained in the flow specification distributed via the resource reservation protocols to all affected network nodes.
- The database of existing reservations must represent the time slices. For each time the set of existing resp. reserved streams with their QoS parameters and the free resources must be known.
- The resource management algorithms must take the time parameters into account.
- Additional failure handling mechanisms and means to save state information in permanent storage are necessary.

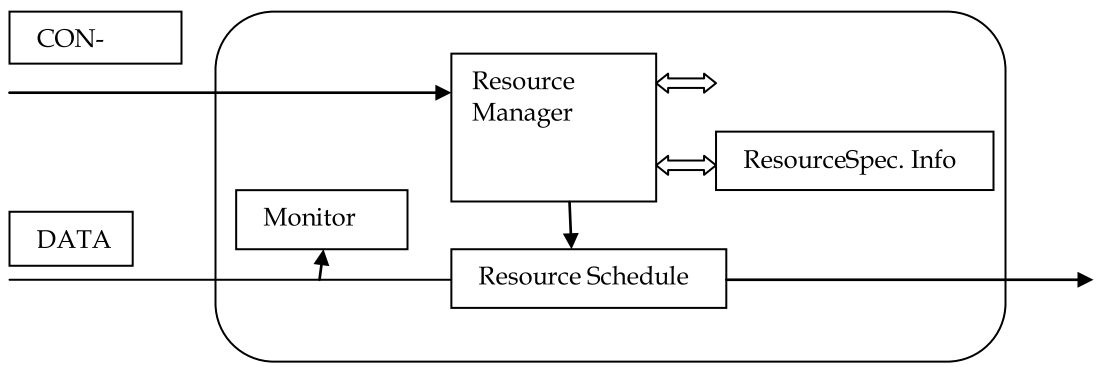

Figure 1:Component Of Resource Management System

MANAGEMENT OF RESOURCES CHARACTERISTICS

The usable capacity of a resource can vary within a long time interval, for instance, due to neccessary maintenance work only parts of the full capacity, e.g., in a network, might be available. Therefore, a system component independent of the reservation management should exist which keeps track of the capacities and characteristics of the managed resources.

The time of the reservation of resources does not necessarily coincide with the beginning of the usage phase, hence, the reserving application is in the mean time usually not active and reachable. Thus, in case of changes, another instance must be available which can implement corresponding reactions. This part can be taken over by the reservation management – it is informed about resource capacity changes and checks then whether all active and reserved streams can still be served. If the available resources are not sufficient to serve all these streams, some of the streams must be modified. For active streams, the application can be informed, whereas for reserved but not yet active streams, the application might not be reachable now. It will be informed about the changed situation when it contacts the reservation management, i.e., when it wants to use the reserved resources.

ALGORITHMS FOR ADVANCE RESERVATION

Different bandwidth allocation schedules could be obtained 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.

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, rein-

ISSN 2229-5518

forcement 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.

The Decomposition Algorithm

We represent the set accept by labelling each order (or order) with "accept" or "reject"2. Our first solver uses branching on accept/reject alternatives combined with a decomposition strategy which breaks the original problem in independent subproblems. It applies various operations to perform its decomposition.

Figure presents these operations. This example comprises 5 orders competing to access a uniform resource capacity of 10 units. The x-axis displays time slots.

The first operation that we can present is the RemoveTime, which is performed each time the capacity exceeds the demand. After this operation, the problem is bounded to time slots t3, t4 and t16.

The second operation is the ForcedAccept which can force the acceptance of the third order. We apply it each time the required QoS is compatible with the resource capacity. We can easily present ForcedReject as the inverse operation, i.e., when the demand is higher than capacity.

Conclusions

While current resource management systems provide mechanisms which offer reliability with respect to QoS, this is not sufficient since many well established application scenarios

We presented an architecture which addresses such issues and offers suitable ReRA functionality. Our implementation shows that it is possible to provide ReRA capabilities to time constrained multimedia applications.

References

[1] A. Banerjea, D. Ferrari, B.A. Mark, M. Moran: "The Tenet Real-Time Protocol Suite: Design,

Implementation, and Experiences", Technical Report TR-94-059, International Computer

Science Institute, Berkeley, CA, USA, November 1994.

[2] A. Campbell, G. Coulson, D. Hutchinson: "A Quality of Service Architecture", ACM Computer

Communication Review, Vol. 24, No. 2, April 1994, pp.6-27.

[3] Y.-H. Chang: "Network Support for a Multimedia Conference Scheduling Service", Proceedings

of SPIE, Vol. 2188, 1994, pp. 109-119.

[4] D. Clark, S. Shenker, L. Zhang: "Supporting Real-Time Applications in an Integrated Packet

Services Network: Architecture and Mechanisms", SIGCOMM 1992.

[5] M. Degermark, T. Köhler, S. Pink, O. Schelén: "Advance Reservation for Predicted Service",

Fifth International Workshop on Network and Operating System Support for Digital Audio and

Video, Durham, NH, USA, April 19-21, 1995.

[6] L. Delgrossi, L. Berger (Ed.): "Internet STream Protocol Version 2 (ST2) – Protocol

Specification – Version ST2+", Internet RFC 1819, August 1995.