Real-time Architecture for Networked Multimedia Streaming systems

Larisa Rizvanovic, student
Mälardalen University
Department of Computer Science and Electronics
larisa.rizvanovic@mdh.se

Gerhar Fohler, supervisor
Kaiserslautern University of Technology
Department of Electrical and Computer Engineering
fohler@eit.uni-kl.de

Abstract

The work presented here has started as part of a FABRIC EU IST project. The aim of the FABRIC project was to develop an architecture in which several interoperability standards and technologies in the home networking context can be integrated. In addition, the FABRIC aimed to handle the complete network to satisfy End-to-End Quality of service (QoS) requirements. In this paper we propose an adaptive QoS framework for efficient resource management, called the Matrix approach. The Matrix is a concept to abstract from having detailed technical data at the middleware interface. In stead of having technical data referring to QoS parameters like: bandwidth, latency and delay we only have discrete portions that refer to levels of quality. The underlying middleware must interpret these values and map them on technical relevant QoS parameters.

1 Problem definition

The project deals with real-time architectures for networked multimedia streaming systems. Key challenges to be addressed include specification of stream and resource characteristics, high demands on processing and timely delivery of multimedia streams, wireless communication between devices and transmission of streams, and architectures for the integration of numbers of devices from various manufactures with diverse demands and capabilities. The project in this area involves both work on developing new algorithms which can match the varying multimedia streams with the varying network and CPU resources, and implementation work, developing and implementing new architectures for system capable of handle such networked streaming issues.

In the study project presented here we want to focus on resource management in a home network. A home network brings some challenges for efficient resource management, as we have:

- Heterogeneous system. We have different local schedulers for CPUs and networks on diverse devices.
- Limited resources. Monitoring and management orders have to be transported over the same resources as streams, potentially incurring high delays and bandwidth cost at streams' expense.
- Highly fluctuating resources. A media application is highly dynamic due to the dynamic nature of the audio/video media content. When processing a video stream there are two types of load fluctuation due to data dependency; temporal and structural [2]. A temporal load is also called a stochastic load and it can be caused by different frame size of stream. A structural or systematic load is often due to a scene change. On the other hand on a wireless networks we talk about long- and short-term bandwidth variations. The long-termed variations can be caused by another application in the system. The short-term changes are often present in the system due to radio frequency interference, like microwave ovens or cordless phones.

Efficient transport of streams with acceptable playout quality requires management of both networks and CPUs. As resources will be limited in the home environment, guarantee mechanisms for continuous stream transport are demanded, e.g., it will not be acceptable to interrupt a football game for the start up of another stream. One of the key issues for resource management is an efficient representation of the fluctuating system state and resource allocation decisions, to provide a small interface to decouple device scheduling and system resource allocation.

2 Matrix approach

We propose a global abstraction of device states as representation of the system state for resource management and as interface to decouple device scheduling and system resource allocation. This global abstraction, called Matrix contains information about device states in a format appropriate for a resource manager. The accuracy of the information represented is suitable for resource management, abstracting over fluctuations or changes, which will overload scheduling: the very fine grain resolution of values is mapped into a very small number or discrete values,e.g. High, Medium and Low (see Figure 1).

The overhead to keep system wide state information fresh is dramatically reduced and no explicit communication or synchronization between resource manager and local schedulers is needed. Orders for resource allocation, made by the resource manager, on individual devices is put in high level abstractions of limited value ranges into the Matrix as well, from where the devices pick up orders to translate them into local scheduling policies or parameters. Thus without the need for explicit costly communication and negotiation between devices, decision about which version of streams to transport or the decomposition of end-to-end delays can be performed by global resource management, reducing overheads and resource waste due to limited local device knowledge. The Matrix provides a logical abstraction of the view of the system state, not an actual centralized implementation requirement. Rather, the Matrix is represented in a distributed way. Not necessarily do all devices in the system have to use the Matrix, but a mix of direct explicit communication and Matrix abstraction is

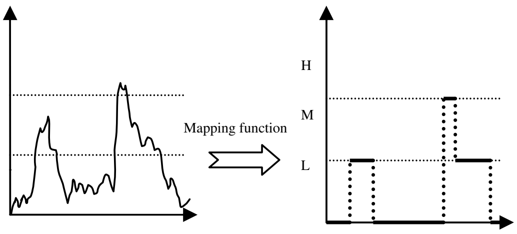

Figure 1: Mapping various kind of traffic specification to few QoS levels

conceivable for resource management, although benefits of the data abstraction would obviously be reduced.

2.1 Architectural design aspects

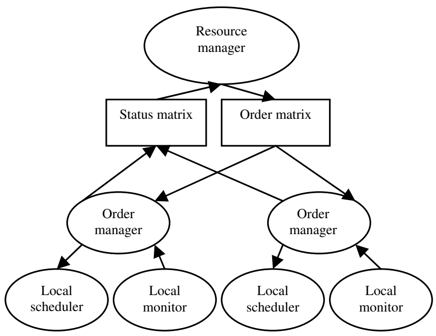

Figure 2: Information flow between Matrix's components

Our framework is composed of several entities that constitute an effective mechanism for monitoring and scheduling available resources in the system. Figure 2 shows the data flow (information flow) between the Matrix components. The functions of these components are further discussed in the following subsections.

The resource manager will be used to schedule and reserve resources, with the system. One part of stream scheduling is providing end-to-end timing constraints. So, the resource manager will provide sub deadlines (sub delays) for each device in the system by real-time methods. In order to deal with resource reservation, the resource manager has to have knowledge about currently available resources in the system. This information is obtained from the Status Matrix. If there are enough resources to support the requested connection, the resource manager puts orders for resource reservation into the Order Matrix. Orders can be seen as an interface between the resource manager's global view of resources and set of entities (order manager, local scheduler and local monitor), which we call "local enforcement mechanism".

The Status Matrix contains information about available resources in the system (provided by the devices). Each resource is represented by its

- current value (out of the limited number range)
- current granularity, i.e., the time interval until which the current value is likely to not change, and
- likelihood that 2) holds.

The Order Matrix contains directions for resources reservation on the devices, made by the resource manager. Each device is presented by one element in the Order Matrix, from where the device picks its order in form of

- delay (sub-delay)
- value (out of the limited number range, QoS performance levels)

An order manager is responsible for determining the available resources at the devices, transforming various kind of traffic specification into a few QoS levels, allocating resources at devices, and providing parameters to local schedulers.

A local scheduler is responsible for scheduling of device's local CPU resource or outgoing network packets.

A local monitor will perceive changes in the bandwidth availability class with a given granularity. Received and specified performances will be compared by the local monitor and the outcome of the comparison will be sent as feedback information to the local monitor on sender device.

3 Research questions

QoS characterization and specification. Mapping of QoS, i.e. translation of streams demands to specific resource requirements (CPU, bandwidth, memory). We believe that we have to find the tradeoff between accuracy of streams demands presentation and efforts to transport and process the same information. Thus, the question is what is a minimum relevant information as needed for an efficient resource management?

Different kind of loads in a system. Given that we have a structural load or long-term bandwidth variation, a resource management has to make an adjustment of streams and resources. The adjustment of streams often involves a switch from one QoS level to another one. Concurrently we want to consider how the change in QoS levels influence perceived quality. Since, a video stream with a constant low quality can be perceived as a better quality than a video stream that has a higher average quality, but with more quality fluctuations [1].

Thus, some of question that we have here are:

- How do we define a change in structural load?
- How is a change between two different QoS levels performed? How we go from one steady state to another one? Two known approaches: " go from one state to another in a controlled way, by passing different under steps, and having controlled situation all the time " or we just go from one steady state to another one, in the fastest possibly way without bothering about having chaos in between.
- Relation between switching/transition and a perceived quality?

4 Summary

Key challenges for a real-time architecture for networked multimedia streaming systems are specification of streams and resources, dealing with high demands on processing, providing timely delivery of multimedia streams, wireless communication between devices and transmission of streams. In this paper we propose an adaptive QoS framework for efficient resource management, called Matrix. The Matrix is a concept to abstract from having detailed technical data at the middleware interface. Instead of having technical data referring to QoS parameters like: bandwidth, latency and delay, we only have discrete portions that refer to levels of quality. Thus, we propose a global abstraction of device states as representation of the system state for resource management and as interface to decouple device scheduling and system resource allocation. The accuracy of the information represented will be suitable for resource management, abstracting over fluctuations or changes that will overload scheduling: the very fine grain resolution of values is mapped into a very small number or discrete values.

References

[1] Reinder J. Brill. *Real-time scheduling for media processing using conditionally guaranteed budgets*. Ph.D Thesis, Technical University Eindhoven, 2004.

[2] Peter van der Stok Sjir van Loo Alejandro Alonso José F. Ruíz Reinder J. Bril Clara M. Otero Pérez, Liesbeth Steffens and Marisol García Valls. QoS-based resource management for ambient intelligence. In *Ambient intelligence: impact on embedded system design*, pages 159–182. Kluwer Academic Publishers Norwell, MA, USA, 2003.