Extension of Resource Management in SIP

Franco Callegati and Aldo Campi

University of Bologna, Italy
{franco.callegati,aldo.campi}@unibo.it

Abstract. In this work we discuss the issue of communication QoS management in a high performance network oriented to carry GRID application. Based on a set of previous works that successfully proved the feasibility of the concept, we propose to use sessions to logically identify and mange the communication between the applications. As a consequence the quality of service of the communication is mapped on reservation of network resources to support a given session. Starting from a framework defined to support such a task for VoIP applications we show here how this framework can be extended to match the need of GRID computing.

Keywords: Application oriented networking, QoS, SIP, Session layer.

1 Introduction

In a GRID computing environment the network becomes part of the computation resources and its performance becomes a major issue. Up to a certain extent the rather "trivial" issue of bandwidth availability can be solved by "brute force", enhancing the technology of the links and making them more powerful. Nonetheless other forms of more intelligent quality of service management as well as the signaling to provide it, requires some enhancement in the network infrastructure. In general terms the network will successfully satisfy the applications demands only if it will be able to provide the necessary grade of service. For instance the time needed to transfer a data object that is part of a more elaborated and distributed processing task will impact on the overall time needed to complete the processing task itself etc. The network must match such requirements and the applications must be able to make evident its requests at the session establishment phase.

This is a general issue that requires a general answer. The work presented in this paper refers to the study case of an optical network using a fast switching technology such as Optical Burst Switching to support short lived consumer grid applications [1][2]. For such a case study previous works [3][4] investigated the possibility to implement all the signaling needed to support the GRID services as part of the network by means of a network session management layer using the SIP protocol [5]. Tasks like resource publication, discovery and reservation may be accomplished by means of SIP message exchanges as part of communication sessions [4]. This solution appears appealing since it leverages a lot on existing building blocks, even though the conventional SIP server must be enhanced with modules that interact with the network and

P. Primet et al. (Eds.): GridNets 2008, LNICST 2, pp. 113 – 120, 2009.

with the application languages. This is also feasible and a full functional application working on top of an OBS data plane was reported in [6].

The concepts presented in this paper are an extension of this work to support QoS management. They can be easily applied to a general network environment but we will stick to the aforementioned scenario as a reference.

RFC 3312 [7] addresses the problem of QoS management in session oriented communications using SIP as signaling protocol. The RFC defines a framework that is based on the concept of “precondition” at session establishment as will be explained later in the paper. Now the question is: could this framework be used to support the needs of a QoS oriented reservation of network resources in an application oriented OBS network? We argue in the following that the answer to this question is no and provide some first insight into a possible solution by analyzing what is missing and proposing a generalization of the framework.

The paper is organized as follows. Section 2 reviews the use of the session layer to support application oriented networking. Then in section 3 are presented the problems to be addressed o support specific quality of service requirements by the applications. The applicability of the existing QoS management framework in SIP is analyzed and the weakness identified. In section 4 is proposed an extension for general application oriented networking. Finally in section 5 we draw some conclusions.

2 SIP Based Application Oriented Networking

A trend we can envisage in new IT services such as GRID computing is that they tend to be “state-full” rather than state-less. The more complex communication paradigms require a number of state variables to manage the information flows. Moreover the Quality of Service issue is more and more important when the communication paradigms become more complex. Multimedia streams have rather stringent QoS requirements and communication without QoS guarantees is rather meaningless in this context.

We believe that the state-full approach with QoS management capabilities has to be pursued for an “application aware” control plane that must manage “vertically” the communication. This is where the session concepts and the SIP protocol get into play:

- sessions are used to handle the communication requests and maintain their state by mapping it into session attributes;
- sessions are mapped into a set of networking resources with QoS guarantees;
- the SIP protocol is used to manage the sessions since it provides all the primitives for user authentication, session set-up, suspension and retrieval, as well modification of the service by adding or taking away resources or communication facilities according to the needs.

In [3] and [4] we have already described some different approaches to resource discovery and reservation exploiting the SIP protocol for GRID network. The resource reservation is part of the establishment of a session that is then used to carry the information flows. We considered that the SIP user agent (UA) already knows the location of the resources into the network, so direct reservation can be performed. In this scenario a SIP UA coupled with an application opens a dialog with a peer SIP UA. The dialog is the logical entity identifying the service relation between end-points (e.g. Video on

Demand dialog) and is made of one or many sessions. The sessions are the logical entities representing the communication services delivered between end-points, e.g. audio session, video session, chat session etc. Sessions exploit the network resources and may be associated with QoS requests. These concepts are in line with the IP Multimedia Subsystem (IMS) architecture [8].

Generally speaking the GRID application can be successfully served if both the application resources to execute the required processing and the network resources required to exchange the related data are available. Two models are possible to finalize the network resource reservation:

- *end-to-end model*, requiring that the peers have the capability to map the media streams of a session into network resource reservations, for example, IMS supports several end-to-end QoS models and terminals may use link-layer resource reservation protocols (PDP Context Activation), Resource ReSerVation Protocol (RSVP), or Differentiated Services (DiffServ) directly;
- *core oriented model,* where the transport network already provides QoS oriented service classes (for instance using DiffServ) and the sessions are mapped directly into this classes by the network itself.

In both models the sequence of actions to be performed in the application layer are:

- agree on the QoS characteristics for the communication paths of the session;
- check if the related network resources are available;
- reserve the resources for the desired session;
- set-up the application session and consequently start the data streams.

In the network layer the aforementioned actions are mapped into:

- check by means of the network Control Plane if enough resources are available;
- reserve the network resources providing the QoS required;

The main issue is how the search and reservation of application and network resources are combined in time during the session start-up phase. The importance of this is due to the fact that the completion of the session requires alerting the end-user to ask whether the session is acceptable or not (phone ringing in a VoIP call for instance). Whether this has to be done before, while or after the network resources required for a given QoS communication are reserved is a matter to be discussed and tailored according to the specific service.

It is worth mentioning that for communications spanning over multiple domains, QoS support is not straightforward. In the remainder of this work we do not consider the multi-domain issue, that will be subject of further investigation, and focus on QoS guarantee within a single network domain.

3 Session QoS Management with SIP

Given that a session set-up phase is always started with an INVITE message sent by the caller to the callee, several interaction models are possible to guarantee the establishment of a session with network QoS guarantee.

Network reservation during session set up: while the INVITE message goes through the network (e.g. with anycast procedures) a network path is reserved before the

INVITE message reach the resource destination. This method can be used for very fast provisioning purpose.

Network reservation before application reservation: as soon as the INVITE message arrives at the destination and application resources are checked, the network reservation starts. After the network resources are reserved the application resource is reserved and the session started. This is the case of VoIP calls.

Network reservation before session set up: as soon as the INVITE message arrives at the destination, and the application resources are found, the application and network resource reservations are started in parallel. Once both reservation are completed independently and both the application resource and the network path are available the session is started. This can be the case of standard GRID sessions.

Network reservation after application reservation: the INVITE message arrive at the destination and application resources are checked and reserved. Then, network reservation starts into the network and as soon as the path between GRID user and GRID resource is established, the application session is started. This can be used when there are few application resources available and the probability to find a free application resource is low.

Network reservation after session set up: the INVITE message arrive at the destination and application resources are checked and reserved, immediately the application session can start without having any network resource reserved. Then, network reservation starts into the network and as soon as the path between GRID user and GRID resource is established, the application session already started can take advantage moving the transmission state from best-effort to QoS enable. This can be used when the QoS is not crucial for the application (or at least not in the beginning part of the session).

The management of the QoS issue is not part of the standard SIP protocol but the issue is there and, not surprisingly, a Resource Management framework was defined for establishing SIP sessions with QoS [7]. The solution proposed exploits the concept of pre-condition. The pre-condition is a set of "desiderata" that are negotiated during the session set-up phase between the two SIP UAs involved in the communication (the application terminals). If the pre-conditions can be met by the underlining networking infrastructure then the session is set up, otherwise the set-up phase fails and the session is not established.

This scheme is mandatory because the reservation of network resources frequently requires learning the IP address, port, and session parameters of the callee. Moreover in a bidirectional communications the QoS parameters must be agreed upon between caller and callee. Therefore the reservation of network resources can not be done before an exchange of information between caller and callee has been finalized. This exchange of information is the result of the initial offer/answer message exchange at the beginning of the session start up. The information exchange sets the pre-conditions to the session that is established if and only if they can be met.

3.1 Weakness of the Current Resource Management Framework

The current framework says that the QoS pre-conditions are included into the SDP [9] message, in two state variables: current status and desired status. Consequently, the

SIP UA treats these variables as all other SDP media attributes. The current and desired status variables are exchanged between the UAs using offers and answers in order to have a shared view of the status of the session. The framework has been proposed for VoIP applications and is tailored to this specific case, but has the following limitations when considering a generalized application aware environment.

1. Both session and quality of service parameters are carried by the same SDP document. Thus, a session established with another session protocol (i.e. Job Submission Description Language (JSDL)) is not possible.
2. The pre-condition framework imposes a strict “modus operandi”, since pre-conditions must be met before alerting the user. Therefore network reservation must always be completed before service reservation.
3. Since the network reservation is performed by the UA at the edge of the network, only a mechanism based on end-to-end reservation is applicable.
4. QoS is performed on each single connection, without the ability of grouping connections resulting from sessions established by others.
5. Since SDP is a protocol based on lines description, it has a reduced enquiring capability. Moreover, only one set of pre-conditions can be espressed by SDP. Multiple negotiations of pre-conditions are not possible.
6. The SDP semantic is rather limited. It is not possible to specify detailed QoS parameters since the SDP lines are marked with “QoS” parameter without specifying any additional detail (i.e. bandwidth, delay, jitter, etc...).

Given these issues we believe that an extended resource management framework is needed in order to achieve a general framework for QoS for application oriented network supporting GRID applications. In the next section we propose an extension to the framework to generalize its applicability to the case of the *Network reservation before application reservation* scenario that we believe to be the more likely to be used in a GRID network. Extensions to other scenarios are possible and will be presented in future works.

4 The Extended Resource Management Framework

The extension can be implemented exploiting the following ideas:

1. do not limit the framework to the use of the SDP protocol but allow more general protocols in the SIP payload at session start-up;
2. separate the information about the request and requirement related to the application layer from the information about request and requirements at the network layer, using two different protocols to carry them;
3. allow as many re-negotiation as possible of the pre-conditions both at session start-up and while the session is running.

Regarding the protocol to declare the pre-conditions we propose to use two protocols.

- Application Description Protocol (ADP): is used to describe application requirements, for instance JSDL which is a consolidate language for describing Job submission for GRID networks.

- Network Resource Description Protocol (NRDP): is used to describe the network requirements and QoS requests. The use of a complete different document descriptor for QoS allows the use of many ADPs for describing application requirements, and not only the SDP. In this work we assume Resource Description Framework (RDF) as the candidate protocol for this task.

The Resource Description Framework [10] is a general method of modeling information through a variety of syntax formats. RDF has the capability to describe both resources and the resources state. For this reason RDF can be used instead of SDP as a more general description language for the network resources. Furthermore, because RDF is a structured protocol, it is possible to enrich its semantic at will with the detailed information about QoS (i.e. Bandwidth, jitter, delay, etc...)

The main advantage of this approach is that the network does not need the capability to understand the ADP since the QoS requirements are only described in the NRDP.

The main drawback is that a link between ADP and NRDP is needed; and therefore some sort of "interface" has to be implemented in the UA on top of a standard SIP UA. The extension to pre-condition negotiation is implemented exploiting two concepts already present is SIP:

- INVITE multi-part message: that is an INVITE that carries more than one protocol in the payload (multi-body), in this case for instance RDF and JSDL;
- NOTIFY message [11]: a message that allows the exchange of information related to a "relationship" between two UAs, for instance a session that is starting up or that is already running.

In Fig. 1 is presented an overview of the call flow in an end-to-end scenario where the network resource reservation is a pre-condition to the session set up.

User A sends an initial INVITE multi-part message including both the network QoS requests and an application protocol for the application requests.

The QoS requests are expressed by means of an RDF document while the application protocol depends on the type of application (i.e. SDP for VoIP, JSDL for GRID computing, etc...)

A does not want B to start providing the requested service until the network resources are reserved in both directions and end-to-end. B starts checking if the service is available, if so it agrees to reserve network resources for this session before starting the service. B will handle resource reservation in the B⇒A direction, but needs A to handle the A⇒B direction.

To indicate so, B returns a 183 (Session Progress) response with an RDF document describing the quality of service from its point of view.

This first phase goes on with the two peers exchanging their view of the desired QoS of the communication, thus setting the "pre-conditions" to the session in term of communications quality.

Then both A and B start the network resource reservation. When an UA finishes to reserve resources in one direction, it sends a NOTIFY message to the other UA to notify the current reservation status by a RDF document in the message body. The NOTIFY message is used to specify the notification of an event. Only when the network channel meets the pre-condition B starts the reservation of the desired resources in the application domain, and session establishment may complete as normally.

Once the service is ready to be provided B send back a 200 OK message to A. Data can be exchanged by the two end-point since both application and network resourced have been reserved. BYE message close the dialog.

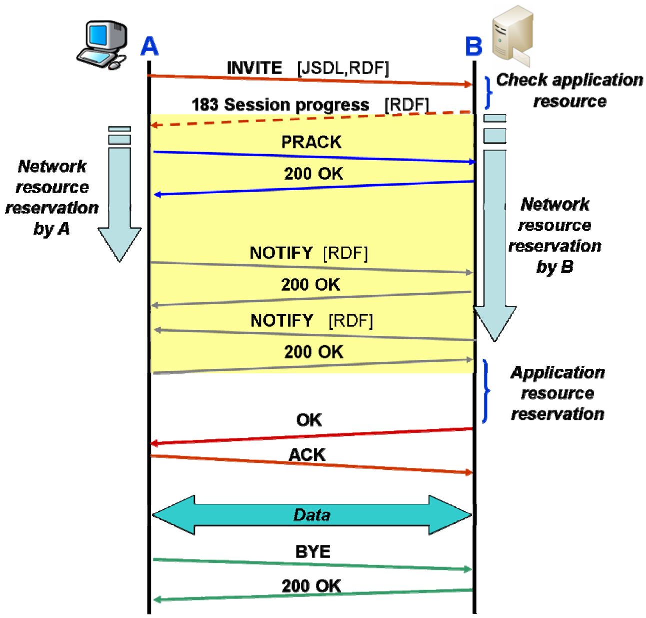

Fig. 1. Example of Call Flow at session set-up

The main differences with the current approach are:

- Use of a separated RDF document in an multi-part INVITE message. The INVITE message starts a SIP dialog between parties. The peer RDF is carried by a 183 response.
- Use of NOTIFY messages instead of UPDATE. Since ADP and NRDP are separated (even if carried by the same INVITE message) the UPDATE message can be sent only by UA A and can be used to update both ADP and NRDP. The NOTIFY can be sent by both UAs and is related to NRDP only and guarantees that the issues of network resource reservation is addressed independently from that of application resource reservation. The NOTIFY messages have Dialog-ID, From and To tag equal to the INVITE message.
- The INVITE message implies the opening of a logical relationship between the peers (a subscription) that ends with the dialog tear down. Each time that the network reservation state changed a NOTIFY message is used to notify the changes for a specific subscription.

The Resource Description Framework is used to generate an offer regarding the desired QoS into the network. The network resource reservation parameters can be many with different semantics. The propose of this paper is not to give an extensive and detailed list of reservation and QoS parameters but is to present the general solution that exploiting the flexibility of the RDF can be adapted by the end users or by the network provider according to the specific needs.

5 Conclusion

In this work we have presented a possible extension of the resource management framework for QoS in SIP. We have addressed the issues of application oriented networking presenting an approach based on communication service mapped into sessions and showing that different session requirements require different interaction between application and network layer. The current SIP Resource Management framework for QoS can cope only with VoIP application using a generic precondition mechanism. We have presented the first step for an extended Resource Management framework showing the call flow that can fit with GRID application. Extensions to other scenarios and details about the interaction between RDF and JSDL will be presented in future works.

Acknowledgments

The work described in this paper was carried out with the support of the BONE-project ("Building the Future Optical Network in Europe"'), a Network of Excellence funded by the European Commission through the 7th ICT-Framework Programme.

References

1. Simeonidou, D., et al.: Optical Network Infrastructure for Grid. Open Grid Forum document GFD. 36 (August 2004)
2. Simeonidou, D., et al.: Dynamic Optical-Network Architectures and Technologies for Existing and Emerging Grid Services. IEEE Journal on Lightwave Technology 23(10), 3347–3357 (2005)
3. Campi, A., et al.: SIP Based OBS networks for Grid Computing. In: Proceedings of ONDM 2007, Athens, Greece, May 29-31 (2007)
4. Zervas, G., et al.: SIP-enabled Optical Burst Switching architectures and protocols for application-aware optical networks. Computer Networks 52(10) (2008)
5. Rosenberg, J., et al.: SIP: Session Initiation Protocol. IETF RFC 3261 (June 2002)
6. Zervas, G., et al.: Demonstration of Application Layer Service Provisioning Integrated on Full-Duplex Optical Burst Switching Network Test-bed. In: OFC 2008, California, USA (June 2008)
7. Camarillo, G., et al.: Integration of Resource Management and Session Initiation Protocol (SIP). IETF RFC 3312 (October 2002)
8. Poikselka, M., et al.: The IMS: IP Multimedia Concepts and Services, 2nd edn. Wiley, Chichester (2006)
9. Rosenberg, J., Schulzrinne, H.: An offer/answer model with session description protocol (SDP), RFC 3264, Internet Engineering Task Force (June 2002)
10. Beckett, D. (ed.): RDF/XML Syntax Specification (Revised). W3C Recommendation, February 10 (2004)
11. Roach, A.B.: Session Initiation Protocol (SIP)-Specific Event Notification. IETF RFC 3265 (June 2002)