A Component-based Approach to SLA Monitoring in Premium IP Networks

S. D’Antonio², M. Esposito², M. Gargiulo², S. P. Romano¹ and G. Ventre¹,²
{salvatore.dantonio, marcello.esposito, mauro.gargiulo}@napoli.consorzio-cini.it
{spromano, giorgio}@unina.it

¹: Dipartimento di Informatica e Sistemistica - Università degli Studi di Napoli “Federico II”
– Via Claudio, 21 – 80125 Napoli – ITALY
²: Lab. ITEM - Consorzio Interuniversitario Nazionale per l’Informatica (C.I.N.I.) – Via Diocleziano, 328 – 80124 Napoli – ITALY

Abstract

Premium IP networks provide users with a portfolio of services, thanks to their intrinsic capability to perform a service creation process while relying on a QoS-enabled infrastructure. The most natural interface between users and Premium IP networks is represented by the Service Level Agreement (SLA). In order to allow an effective exploitation of such networks, it is of primary importance the availability of mechanisms for the monitoring of performance parameters related to a specified service instance. This capability, named SLA monitoring, is of interest for both the end user, as the entity that uses the service, and the service provider, as the entity that creates, configures and delivers the service.

In SLA monitoring several components are involved, such as network operators, content providers, and end-users, often spread over multiple domains. Involving all these entities has a cost: it is due to the usage of computational, storage and network resources, as needed for information retrieving, analysis and distribution.

In this paper, we illustrate a proposal for a monitoring framework for the collection and distribution of performance data. The idea our proposal is mainly based on, makes reference to the definition of an information document named Service Level Indication (SLI). Monitoring information, related to a specific Service Level Agreement, are so retrieved from the network and provided to the user application, by collecting and appropriately combining performance metrics and measures in an SLI document instance. Such a document is the interface allowing to exchange monitoring information between peer components located in different domains as well as between components at different levels of abstraction in the same domain.

Keywords: Quality of Service, Performance Monitoring, Service Level Agreements, Service Level Indication, Inter-domain Monitoring

1. Introduction

In the last few years, a new approach to Quality of Service (QoS) issues has aimed at the definition of new network architectures, in which nodes and elements are capable to provide deterministic or statistical guarantees for data transmission. Such an approach is based on the definition of mechanisms and algorithms for admission control and resource management (Pro-active QoS) and, albeit very complex, it is producing interesting results in terms of definition of realistic, deployable models of novel network architectures.

By relaying on a QoS-enabled infrastructure, Premium IP networks [5] can deploy and offer a wide portfolio of services, like Video on Demand, VoIP and VPNs, thanks to their intrinsic capability to perform a service creation process.

In such a scenario, users can choose among different levels of Quality of Service in order to best meet their application and pricing constraints. The service and its delivery quality are negotiated through a contract, named Service Level Agreement (SLA), between a user and a service provider. Such a service provider is intended as an entity capable to assemble service contents and both network and server side resources. These resources could be located in different domains and/or belong to different providers [6].

The subscription of a Service Level Agreement implies two aspects, only apparently un-related: first, the auditing of the actual satisfaction of current SLA; second, the dynamic re-negotiation of the service level agreements themselves. As far as the first aspect, we can reasonably forecast that as soon as users start to pay for communication services with QoS-related guarantees (e.g. service availability), it will be required to verify whether or not the provider actually meets the conditions
specified in the SLA. With reference to the second aspect, indeed, re-negotiation of QoS has been always accepted as an important option in performance guaranteed communications. The re-negotiation of an SLA can be triggered by several events linked to critical problems such as the efficiency of network resource allocation, the performance of user’s application, and the reduction of communication costs.

A typical example we might consider refers to a scenario where the received Quality of Service can be seen as influenced by several factors: the network performance, the server load and the client computational capability. Since those factors can be varying in time, it is logical to allow modifying Service Level Agreements on the basis of the QoS actually achievable and perceivable at the application layer.

From this point of view, results provided by traffic measurement activities represent an important element on the basis of which users might request an SLA re-negotiation. We therefore believe that the possibility to modify the existing QoS-based agreements between a service provider and an end-user will assume an important role in the management of Premium IP networks.

In this document, we illustrate a proposal for a novel monitoring and control framework for the collection and distribution of performance data in a multi-domain scenario.

When designing such a framework, we have identified the main actors playing a role in the context of SLA-based services. Users, service providers and network operators are interested in QoS monitoring activities since results provided by such activities can be exploited for several goals.

Our proposal makes reference to the definition of an information document, named Service Level Indication (SLI), which acts as an interface among several entities involved in the SLA monitoring task.

As this document will disclose, Service Level Indication is to be produced with the cooperation of all of distributed components in order to obtain a detailed picture of the level of service that is currently offered. The SLI-based monitoring architecture is quite simple in its formulation; nonetheless it brings in a number of issues, related to its practical implementation, to its deployment in real-life scenarios, and to its scalability in complex and heterogeneous networked infrastructures. Some of these issues will be highlighted in the following, where we will also sketch some possible guidelines for deployment, together with some pointers to potential innovative approaches to this complex task.

The paper is organized as follows. In section 2 we illustrate motivation and requirements related to the SLA monitoring service. A document for the exchange of monitoring information is introduced in section 3 by describing its exploitation within a framework for service provisioning over Premium IP networks. In section 4 we explain the data export process. Finally, section 5 provides some concluding remarks to the presented work.

2. Motivation

Computer networks are evolving to support services with diverse performance requirements. To provide Quality of Service (QoS) guarantees to these services and assure that the agreed QoS is sustained, it is not sufficient to just commit resources since QoS degradation is often unavoidable. Any fault or weakening of the performance of a network element may result in the degradation of the contracted QoS. Hence, QoS monitoring is required to track the ongoing QoS, compare the monitored QoS against the expected performance, detect possible QoS degradation, and then tune network resources accordingly to sustain the agreed QoS [3].

The adoption of a pro-active behavior with reference to the provisioning of QoS communication, introduces several issues that we believe are worth of investigation for their impact on future SLA-based Premium IP networks. Among them, a major role is played by dynamic negotiation between applications and networks for the selection of an adequate Quality of Service. In fact, when a user contacts the provider of a video-delivery service, he will expect to negotiate the access to the service and its price. In Premium IP networks such price will be influenced not only by the multimedia content of interest to the user but also by the QoS level that will be required in the delivery of the content itself across the network. In this context, we have a negotiation for a service where the user will ask for contents and for a certain quality of their delivery, and the service provider will answer with a price. Actually, we might also expect that such initial negotiation will be performed automatically between the client application on one side and the provider on the other one. Furthermore, such a negotiation might happen at service subscription time rather than upon service invocation.

From the user’s point of view, SLA monitoring represents an important functionality as it allows getting information about the fulfillment of negotiated service parameters.

By exploiting such a functionality, a user can identify possible violations of his SLA or trigger an SLA re-negotiation if the current level of QoS is not satisfactory. Similarly, the service provider can ask for SLA re-negotiation in order to optimize resource utilization.

Therefore, monitoring seems to offer excellent novel opportunities to service providers who do have the possibility to effectively engineer their network infrastructures. By this way, they can exploit at its best the unprecedented potential disclosed by dynamic service
creation and delivery. In this case, “SLA monitoring” may be seen in the light of a more general activity related to “Network monitoring”, and this is mainly due to the following objectives:

- usage-based accounting
- traffic profiling
- traffic engineering
- attack/intrusion detection

### 2.1. Monitoring as a service

In a Premium IP network, the service provisioning is the result of an agreement between the user and the service provider, and it is regulated by a contract. The SLA is the document resulting from the negotiation process and establishes the kind of service and its delivery quality. The service definition stated in the SLA is understood from both the user and the service provider, and it represents the service expectation, which the user can refer to. Such SLA is not useful to give a technical description of the service, functional to its deployment. Therefore, a new and more technical document is needed. The Service Level Specification document, as described in [1], derives from the SLA and provides a set of technical parameters with the corresponding semantics, so that the service may be appropriately modeled and processed, possibly in an automated fashion.

The SLS can also be used by different providers, in order to cooperate in the service fulfillment: this issue, which is mainly related to the inter-domain scenario, requires that a thorough definition of the protocols and mechanisms involved in the exchanging of information between each pair of peering entities along the service delivery chain is provided.

As a service, the monitoring activity might be required by a user to a provider, which is responsible for the service itself. Such a provider is in charge of both retrieving monitoring data across involved domains and forwarding results to the user after a proper computation.

Like other services, we can forecast that the kind of provided information will depend on the user’s profile, whereas the accuracy of such an information will be proportional to the service cost. From this point of view, ad hoc defined pricing policies have to be specified and instantiated.

More precisely, drawing inspiration from the concept of *metadata*, we might define the monitoring as a *metaservice*, i.e. a ‘service about a service’. This definition is mainly due to the fact that it is hardly conceivable a monitoring service per se: monitoring is strictly linked to a pre-existing service category, for which it provides some value-added information.

Therefore we won’t consider a standalone monitoring service, but we will rather look at it as an optional clause of a traditional service, thus taking it into account in the SLA negotiation phase.

### 2.2. Requirements

In an architecture capable to dynamically negotiate a rich portfolio of services, it becomes of primary importance the availability of mechanisms for the monitoring of service performance parameters related to a specified service instance. This capability is of interest both to the end-users, as the entities that ‘use’ the service, and to the service providers, as the entities that create, configure and deliver the service.

Furthermore, in an inter-domain scenario SLA monitoring can likely imply the cooperation of several providers scattered along different Autonomous Systems (ASs). In such a scenario it is extremely important to establish a set of rules for the interactions among involved entities. We can identify two kinds of interactions:

- **vertical**: among entities at different abstraction levels within the same AS or domain;
- **horizontal**: among peer entities belonging to different ASs.

The former is useful to the aggregation and refinement of data collected at the device level. The latter is especially suitable to exchange high-level, concise information needed to build the end-to-end monitoring result.

Since intra-domain protocols exchange a huge amount of data, they should take care about communication performances. On the other hand, inter-domain protocols should mind the security of communications because they act in a wide-area scenario.

In the case of SLA-based Premium IP networks, QoS monitoring information should be provided by the network to the user application, by collecting and appropriately combining performance measures in a document which is linked to the SLA itself and which is conceived following the same philosophy that inspired the SLA design: i) clear differentiation of user-level, service-level and network-level issues; ii) definition of lean and mean interfaces between neighboring roles/components; iii) definition of rules/protocols to appropriately combine and export information available at different levels of the architecture.

### 3. Service Level Indication

Since the service and its quality are perceived in a different fashion depending on involved actors (end user, service provider, network operator), there is a need to define a number of documents, each pertaining to a specific layer of the architecture, suitable to report information about currently offered service level. As far
as data reports, we have defined a set of new documents, named Service Level Indication (SLI), aiming at indicating whether measured data, related to a specific service instance, are in accordance with the QoS level specified in the current SLA and SLS documents.

3.1. The reference framework

On the basis of the monitoring service definition, we aim at describing the roles needed to its implementation. In the CADENUS project [7], we find a possible architecture for service provisioning in a Premium IP network.

In the CADENUS architecture, there are three main functional components, which act as mediators, between the user and the resources involved in the service provisioning. These components are the following:

• **Resource Mediator(s):** they have to manage the available resources, by configuring the involved nodes. Each service can concern different domains and then different Resource Mediators. The Resource Mediator also has to gather raw monitoring information, and produce the Technical SLI document.

• **Service Mediator(s):** they are in charge of creating the service as required from the user, using the resources made available by one or more Resource Mediators. By exploiting information contained in the SLS and the Technical SLI, they have to evaluate SLA fulfillment and produce the User SLI.

• **Access Mediator(s):** they act as service brokers, so they have to provide the user with AAA functionality. The Access Mediator receives the User SLI from the Service Mediator and returns this information to the user according to his profile.

3.2. The SLI specification

By considering the different abstraction levels (user, service, and network), it is possible to identify three kinds of SLIs (see Figure 1):

• **Template SLI,** which provides a general template for the creation of the documents containing the monitoring data associated to a specific service;

• **Technical SLI,** which contains detailed information about the resource utilization and/or a technical report based on requirements stated in the SLS. This document, which pertains to the same level of abstraction as the SLS, is built by the Resource Mediator;

• **User SLI,** it contains, in a friendly fashion, information about the service conformance to the negotiated SLA. The User SLI is built by the Service Mediator on the basis of the SLS, the Template SLI and the Technical SLI. Such a document is sent to the Access Mediator which creates a formatted User SLI by taking into account user’s profile and terminal capabilities.

Figure 1. The three SLI documents and their scope

The service monitoring has to be finalized to the delivery of one or more SLI documents, User SLI or Technical SLI. In the SLI issue, multiple entities are involved, as network elements, content servers, and user terminals. Involving all these elements has a cost: it is due to the usage of computational, network and storage resources, needed for information collection, analysis and distribution. This cost depends on both the number of elements involved and the information granularity.

For example, we can consider a VPN service spanning dozens of hosts located all over the world. For such a service, the monitoring might be able to report detailed information about throughput, delay, packet loss, availability, etc.

By using the XML format, in Figure 2, we present a possible example of a Technical SLI document.

Obviously, information stated in Figure 2 can be further filtered and submitted to the user in a more friendly fashion (i.e. a formatted User SLI).
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE SLI SYSTEM "F:\Projects\Cadenus\XML\SLI.dtd">
<SLI>
  <ID>352189</ID>
  <SLAIID>726549</SLAIID>
  <FlowIdentification>
    <SourceAddress>217.9.64.200</SourceAddress>
    <DestinationAddress>217.9.64.210</DestinationAddress>
    <SourcePort>80</SourcePort>
    <DestinationPort>90</DestinationPort>
    <ProtocolNumber>10</ProtocolNumber>
  </FlowIdentification>
  <Performance>
    <One-Way-Delay>
      <Value>100ms</Value>
      <MeasurementPeriod>86400s</MeasurementPeriod>
      <ThresholdViolations>5</ThresholdViolations>
    </One-Way-Delay>
    <PacketLoss>
      <Value>0,001</Value>
      <MeasurementPeriod>86400s</MeasurementPeriod>
      <ThresholdViolations>8</ThresholdViolations>
    </PacketLoss>
    <Jitter>
      <Value>5ms</Value>
      <MeasurementPeriod>86400s</MeasurementPeriod>
      <ThresholdViolation>10</ThresholdViolation>
    </Jitter>
    <Throughput>
      <Value>1Mbs</Value>
      <MeasurementPeriod>86400s</MeasurementPeriod>
      <ThresholdViolations>4</ThresholdViolations>
    </Throughput>
  </Performance>
  <Reliability>
    <MeanDownTime>
      <Value>6000s</Value>
    </MeanDownTime>
    <TimeToRepair>
      <Value>3000s</Value>
    </TimeToRepair>
  </Reliability>
  <Sampling Type="Periodic"/>
</SLI>

Figure 2. An example of Technical SLI in XML format

4. Data Export

Till now we have made a bird’s eye view analysis of current definitions and issues related to QoS measurement and monitoring. However, in the context of SLA-based services the following innovative aspect has to be considered: in order to allow users, service providers and network operators to have information about QoS parameters and network performance the need arises to export data collected by measuring devices. For this reason, the concept of data model has to be introduced. Such model describes how information is represented in flow records. As stated in [4], the model used to export measurement data should be flexible with respect to the flow attributes contained inside reports. Such reports can be obtained in two possible ways: push mode and pull mode. In push mode, the measuring device decides autonomously when to send a report on the measured traffic. Furthermore, the measuring device should report collected data regularly according to a given time interval and when a specific event occurs. In pull mode, report sending is triggered by an explicit
request from a data collector or some other receiver of monitoring information.

Having in mind the CADENUS architecture, it is possible to identify the components responsible for the creation of each of the monitoring documents introduced in previous section. Such documents are then exchanged among the components as described in the following (see Figure 3), where we choose to adopt a bottom-up approach:

1. At the request of the Service Mediator, the Resource Mediator builds the Technical SLI document on the basis of data collected by measuring devices located in its own domain. The fields it contains are directly derived from those belonging to the SLS and are filled with the actual values reached by the running service. The resulting document is sent to the Service Mediator.

2. Thanks to the Technical SLI received from the Resource Mediator, the Service Mediator is capable to evaluate the service quality conformance with respect to the requests formulated through the related SLS. It can be interested in such an information both for its own business and in order to gather data for the creation of a complete report in case a user requests one.

3. At the user’s request, the Service Mediator, exploiting data contained in a Technical SLI, produces a further report indicating the QoS level, as it is perceived by the end user. The document it prepares is derived from a service specific template (the so-called SLI Template), which provides an abstraction for the measurement results in the same way as the SLA Template does with respect to the service parameters. Such a document, hereby called User SLI, is to be delivered to the end-user.

4. The Access Mediator receives the User SLI from the Service Mediator, puts it in a format that is compliant with both the user’s preferences and the user’s terminal capabilities and then forwards it to the end-user (Formatted User SLI).

Moreover, in a multi-domain scenario (see Figure 4) in order to obtain monitoring information between two end-points, named $H1$ and $H2$, the first Resource Mediator along the service chain ($RM_A$) receives the monitoring request from the Service Mediator and decomposes such a request into two fragments:

- the first one refers to the Autonomous System managed by $RM_A$ itself;
- the second one is under responsibility of the next Resource Mediator en route towards the $H2$ end-point ($RM_B$). Such a fragment implies making measures within other ASs. $RM_B$ will recursively perform the same actions, by splitting again the received request into two parts.

The fragment received by $RM_C$ completely refers to Autonomous System $C$, so it does not need further fragmentation. $RM_C$ builds its local Technical SLI and then forwards it to $RM_B$.

Each Resource Mediator along the monitoring chain (i.e. $RM_B$) builds a Technical SLI document, by assembling local data with results contained in the Technical SLI received from the adjacent Resource Mediator.

Finally, the Resource Mediator interacting with the Service Mediator (i.e. $RM_A$) is in charge of assembling the global Technical SLI containing the end-to-end performance. Such a document is then forwarded to the Service Mediator.

5. **Discussion and Conclusions**

The need for technologies and architectures for the provisioning of seamless monitoring and performance data in future QoS based networks is under everybody’s eyes. It is a need related to the concrete requirements of a wide number of different players: application designers, content providers, service providers, network operators, end users and last but not least, third party monitoring agencies. However, it is also clear that the provisioning of such feature is particularly complex and critical, since it involves the coordination and orchestrated operation of a large number of elements, separately owned and managed along what we have called the provisioning chain from the service location to the end user. We therefore foresee a number of issues to be faced before this problem can be solved. For some of them we provided a possible solution, while for the others the discussion is still wide open.

We briefly mention here the main facets of the general issue of QoS monitoring, focusing on the networking infrastructure.
Figure 3. Information passing in an Intra-domain scenario

Figure 4. Information passing in an Inter-domain scenario
First of all, the collection of monitoring data from the network elements. This issue is clearly related to both technical and business aspects.

As far as the first ones, the work ongoing in the area of policy based management of network elements is providing a technical framework in which the control and configuration of network nodes will be much more straightforward than that currently achievable through the traditional SNMP based approach. However, it is clear that for global communication infrastructures, involving large number of nodes with a huge number of active connections, we do have a problem of scalability with respect to the collection and delivery of performance data. In spite of this, we believe that there are features in the existing network architectures that might be exploited to reduce at least this problem. For example, in DiffServ based network architectures monitoring of Service Level Agreements can be performed usually per traffic class and not per single traffic flow, and could be normally limited to the ingress and egress points of a domain. More detailed performance data collections (in terms of specific flows or network elements) could be triggered only in the presence of specific demands from the involved parties or in the case of anomalies. We could therefore imagine a scenario where a network operator could regularly broadcast information (in our model a Technical SLI) related to the average performance behavior of its network infrastructure.

As far as the business aspects, i.e. those related to the business nature of the provisioning of communication services, we can mention here the one we believe is the most important: trust. In global networks, large scale infrastructures will be managed by a multitude of different operators, each managing a separate network domain. Quality of Service will therefore be an issue involving a number of parties, each responsible only for the service provided in the domain that it directly manages. Such parties will be obliged, at the same time, to compete and to cooperate with peering entities. Can we foresee a scenario where such performance data will be openly (albeit in a controlled way) available? We believe that rather than being an obstacle to the deployment of a common framework for SLA monitoring, trust will be an important trigger for it, if not a prerequisite. In fact, we can expect that no operator will start charging for premium services involving infrastructures owned by others without a formal, standardized way for exchanging performance data about the communication services offered to and received from other operators.

The proposed approach, based on a clear definition of roles and responsibilities of involved entities, addresses in a neat fashion the trust-related issue.

A further issue is related to the devising of a common Quality of Service measurement framework. It is clear that performance data should be provided in a way that is independent of both the network architecture offering the service and the application service demanding it. Our proposal is a first attempt in this direction.

6. Acknowledgments

This work has been performed under the partial financial support of the Italian Ministry of Education, University and Research (MIUR) under grants “LABNET2” and of IST project INTERMONT IST-2001-34123, http://www.ist-intermon.org. We also gratefully acknowledge access to the results of IST project CADENUS IST-1999-11017 “Creation and Deployment of End-User Services in Premium IP Networks”, http://www.cadenus.org.

7. References