Semantic Description of Web Service QoS Profiles for Context-aware Web Service Provision

JOANNA ŚLIWA, MAREK AMANOWICZ
Military Communication Institute, ul. Warszawska 22A, 05-130 Zegrze, Poland

Abstract: The article briefly presents semantic description of web service QoS profiles that is part of the larger framework for context–aware service provision. It consists of upper-level ontology that defines basic concepts and their relationships and domain ontology that specifies user QoS profile, service QoS profile and network performance. Application of QoS semantic descriptions has been carried out as a case study that resulted in several conclusions important in terms of their application for web services provision in military SOA – based systems. The main focus of the article are conclusions from the case study.

Keywords: web services QoS, semantic description, ontology, context-aware services

1. Introduction

Network Enabled Capabilities Concept provided by NATO (NNEC) is based on seamless exchange of information among mission participants in a dynamic environment, where the forces of the alliance usually cooperate with unanticipated partners. On the basis of multiple capabilities, they can achieve information superiority by sharing reliable information collected from various sources, creating situational awareness and distributing it among mission participants, across domain, context and organizational boundaries.

One of the technologies that strongly support achieving those goals is Service Oriented Architecture (SOA), an architectural paradigm that promotes loose coupling among systems components, reusability of the components and is theoretically software – independent. On high command levels, it is the most often implemented using Web Services (WS), which provide the means to build a very flexible environment, able to dynamically link different system components to each other.

WSs are described by a wide range of standards that deal with different aspects of WS realization, transport, orchestration, semantics, etc. These standards are based on the eXtensible Markup Language (XML) and have been designed to operate

---

1 Work has been co-financed by the European Regional Development Fund under the Innovative Economy Operational Programme, INSIGMA project no. POIG.01.01.02-00-062/09.
in high bandwidth links. XML gained wide acceptance and became very popular for
the reason that it solves many interoperability problems, is human- and machine-
friendly and facilitates the development of frameworks for software integration,
independently of the programming language. Nevertheless, it undoubtedly adds
significant overhead, both in terms of computation and network resources while
being transported.

The utilization of SOA-based implementation of Web Services in NEC environ-
ment has been shown in many international experiments (e.g. CWID – Coalition
Warrior Interoperability Demonstration). They prove that WS technology improves
collaboration, interoperation and information sharing in the Federation of Systems
(FoS). However, in order to achieve efficient information exchange between the users,
SOA solutions need to work with different types of information and communica-
tion systems. Service interoperability needs to be provided among all command
levels on the end-to-end basis. The challenge is, therefore, to apply SOA and web
services also in low bandwidth disadvantaged tactical communications systems,
which usually cope with high error rates and frequent disruptions.

The effects that the web services’ users experience when they are connected
by low bandwidth, unstable link are high response times or lack of response from
the application server. The question is, therefore, how to achieve quality of ser-
service satisfying the user. However, this question does not only relate to the aspect
of QoS mechanisms that must be triggered in the TCP/IP and lower layers of the
protocol stack (which is also related to the cross-layer cooperation), but also
to the problem of stating the user requirements and service offers in terms of QoS,
and then processing them at the application layer, e.g. by the service discovery
and matching algorithms.

The article refers to these aspects and discusses the differences in utilization
of quality of service descriptions for web services in overprovisioned networks and
in disadvantaged networks. It also describes approaches to QoS descriptions for web
services and their applicability for web services frameworks invoked through disad-
vantaged networks. The main part of the article is a case study showing an example
of semantic description of QoS attributes for web services based on OASIS WSQF
and OWL-Q, and discussion of problems that arise when using this description.

The work presented in this paper will be included in the results of RTO/IST
090 Work Group titled “SOA challenges over real time and disadvantaged grids”.

2. Quality of Service

There are many aspects and layers of the quality of service (QoS). The term
QoS is extensively used today, not just in the telecommunications world from
which it derives, but more and more often regarding IP-based services. Systems and
networks are being designed taking into consideration end-to-end performance
required by user applications. Looking for the definition of QoS, one can find out
that the term QoS is usually not well defined, is used in different meaning, or worst
of all, misused. The most accurate definitions are presented by the ITU-T Rec-
mendations. For instance, ITU-T Rec. E.800 [4] defines QoS as:

“the collective effect of service performance which determines the degree of sat-
isfaction of a user of the service”.

This recommendation considers QoS in scope of service performance per-
ceived by the user that consists of support, operability, serviceability and security.
All of them are dependent on network characteristics. The ITU-T Rec. E.800 defines
the quality of service concept, relates quality of service and network performance
and provides many usable performance measures that can be also used to define
web services quality.

On the basis of the E.800 concept, ITU-T Rec. G.1000 proposed service quality
criteria (speed, accuracy, availability, reliability, security, simplicity, flexibility) and
defined four viewpoints of the quality of service:

- Customer’s QoS requirements,
- Service provider’s offerings of QoS,
- QoS achieved or delivered,
- Customer survey ratings of QoS.

![Figure 1. Four QoS viewpoints on the basis of ITU-T Rec. G.1000](image)

The definitions proposed in ITU-T recommendations can be used to reflect
web services quality. However, it is necessary to consider it according to the four
aforementioned viewpoints.

QoS requirements of the user state the level of quality required by the applica-
tions of service customers. They may be expressed non-technically. QoS offered
by the provider is the level of quality expected to be offered to the customer by
the service provider. These two QoS viewpoints are stated by the two parties and
may be different from the real achieved and perceived values.

QoS perceived by the users expresses the level of quality that customers believe
they have experienced. QoS achieved by the provider is a statement of the level
of the actual quality achieved and delivered to the customer.
There are different metrics to measure QoS achieved by the provider and QoS perceived by the consumer. Their values can differ from each other due to the fact that QoS experienced by the user is influenced by the network performance (which is the case in terms of disadvantaged networks).

QoS of the service in being influenced by the actually experienced network performance. In order to minimize this influence, network QoS mechanisms are introduced. (e.g. according to the QoS architecture described in ITU-T Rec. Y1291).

Quality of service is often referred to in terms of network performance stated as, for instance, high bit rate, low latency or low bit error probability. This can be confusing because only particular services (especially multimedia telecommunications services) require reservation of network resources or guarantees defined using network parameters (like bandwidth, error rate, delay, jitter). These services generate traffic that is usually referred to as non-elastic. QoS offered and achieved by the service can be in this case defined using network performance metrics. This is very comfortable when defining QoS mechanisms for the network (like those defined in the ITU-T Rec. Y.1291) that are able to support provision of required QoS level. The achieved QoS level can be then evaluated by the metrics reflecting network performance. For instance, ITU-T Rec. G.1010 defines performance targets for audio and video applications with required one-way delay, delay variation (jitter) and information loss. Throughput is usually dependent on the audio/video codec that the application utilizes. Network QoS mechanisms are designed and employed to meet these targets.

However, apart from non-elastic traffic, we have also an elastic one that describes bulk transfer applications. It is difficult for these services to define network performance metrics that would enable providing appropriate QoS level perceived by the client.

Web services (WSs) are a good example of applications that generate elastic traffic. They are considered by the ITU-T Rec. G.1010 data type applications. For these applications, G.1010 defines performance targets, as a matter of fact, just in terms of delay (considered as response time). Web services are, however, not sensitive to delay variation. In order for the application to process the data, information loss at the application layer must be equal to zero. Therefore, appropriate mechanisms that enforce reliability must be employed. It is difficult to define required throughput of a web service since this will only influence the time needed to deliver particular piece of data (depending on the amount of data).

Taking this into account, it is visible that using link performance metrics for stating QoS requirements or offerings is problematic and very often subjective. The only QoS parameter the application of which is clear is response time, influenced by the actually experienced network performance.

3. Scenario

Our work assumes the scenario where the armed forces operate in a multinational mission, using heterogeneous system based on the Federation of Systems
(FoS) approach. Network domains can be realized in any technique, assumed that they are able to transfer IP traffic. The system is based on the Service Oriented Architecture and uses web services as the main technology to distribute information to the users. Web service traffic is sent to users located in different levels of command, according to the actual needs and security constraints.

Service providers are located in the fixed or deployable network (strategic and operational levels), and the clients are located in the deployable and mobile networks (tactical edge and tactical level). They use terminals with limited resources (e.g. battery, computation power, etc.). What is more they move while the operational situation changes, changing also their location, communications networks and their tasks. While in operation, they need to be fed with information that are crucial for the successful mission accomplishment. The scenario is to investigate the problem of information delivery with required quality of service, stated by the client and provided by the service with regard to the network QoS mechanisms that support web service traffic transmission in the network.

4. Problem overview

The article focuses on one of the problems of QoS guarantee in SOA and web services environment defined in [1], which is “How to state and process QoS requirements of the web service client”. It presents results of the case study of user and service QoS specification that can support web services delivery with end-to-end quality of service.

It assumes that QoS attributes are incorporated in the user profile and service descriptions. They can be used mainly in two ways. Firstly – to provide the services discovery and selection process that takes into account QoS parameters. Secondly, to support web services realization parameterizing the network QoS mechanisms and, if necessary, employing additional mechanisms that adapt web service realization to the current network performance, user requirements, possibilities of its terminal, etc. The list of the parameters that can be used for the purpose of the adaptation is long, and selection of the appropriate ones depends on the idea of the system architect.

Web services selection is being investigated by many research institutes all over the world. It follows the loose coupling concept and provides the basis for services orchestration, providing more complex services based on their descriptions, inputs and outputs. There are also solutions that incorporate QoS attributes for the purpose of service discovery [2], [3][10][11]. An interesting study in QoS attributes has been presented in [3]. Matching and discovery process for web services in mobile environments are shown in [3] and [2]. However, as stated in [2], “adding QoS awareness to Web services is complex” and among many other mechanisms, needs “the ability for a client to specify QoS needs during service invocation”. This aspect is usually omitted in literature. Moreover, these QoS parameters need
to be communicated to the lower layers and mapped into networking parameters to appropriately parameterize network QoS mechanisms.

The QoS attributes can be used for web service provision, however, not only for the cooperation with QoS mechanisms, but also for the purpose of the adaptation process. It should take place when the network is not able to realize the service according to the user preferences and can also take into account parameters of the user’s device, e.g. its software and hardware. The more parameters are defined, the more complex the environment becomes. The adaptation based on such an approach follows the concept of context – aware service provision.

5. QoS attributes presentation

The aspect of stating user QoS requirements and service QoS offers is considered from the perspective of the context – aware service provision. It assumes adaptation of services invoked by users located in disadvantaged networks on the basis of the context of the service call. Among many other parameters, it includes QoS attributes required by the user and QoS attributes offered by the service provider. Proposed solution is part of a larger Adaptation FRamework for web service provision in disadvantaged environment (AFRO) [16]. The QoS attributes description and definition is part of a more general semantic description of the context of the service call (AFRO ontology), which allows performing appropriate adaptation actions in a proxy that is located at the border of the wireless network. According to the concept, the context is modeled in OWL (Web Ontology Language) and consists of:

- user context:
  - user QoS profile,
  - adaptation preferences,
  - environment (device, network, actions),
- service context:
  - service QoS profile,
- network context:
  - Link performance.

The adaptation proxy operates on the basis of the context ontology and the adaptation ontology describing all the actions that need to be taken by the proxy, reflecting the user preferences and the adaptation rules defined using the Semantic Web Rule Language (SWRL).

Most of the proposals of the context representation omit the problem of the relationship between context data and communication networks. However, in the considered scenario, web services are to be provided in a disadvantaged environment so that the impact of network performance, even though not direct, should be reflected in the QoS framework. The approach presented in this paper is willing to relate web service delivery to the current network performance taking into
account QoS attributes defined as required by the user, QoS attributes defined as provided by the service provider and QoS attributes related to the available network performance.

6. Related work

There are many approaches to describing aspects of web services Quality of Service as part of a larger services management framework, but any of them is mature enough to set its position as a standard. They focus, for instance, on defining QoS information (attributes) together with appropriate values or even describing QoS specifications, which are descriptions of what/where/when/how to monitor and control. The latter ones are mostly industry propositions (e.g. by IBM). The QoS information definitions are usually in the form of:

- descriptions (e.g. description of QoS attributes – by W3C, OASIS),
- contracts (Service Level Agreements – SLAs, e.g. Web Service Level Agreement: WSLA; Classes of Service, e.g. Web Service Offerings Language: WSOL; general contracts – e.g. OWL-S, WS-Agreement),
- extensions to UDDI (e.g. UDDIe),
- policies (WS-Policy).

These approaches are unfortunately most of all competitive or overlapping. There is no well-defined and complete framework for web services management that would solve the problem of of QoS definition, provision, monitoring and control, and would be beneficial for web service – based system's architect. Therefore, there is a need for common framework for measuring and processing the QoS information that, inter alia, guarantees common interpretation. As an introduction to the case study on semantic QoS description, the following subchapters will focus on well-defined QoS attributes description provided by W3C and OASIS, seen from the perspective of their possible application for web services in a disadvantaged environment.

6.1. QoS attributes description for web services

Web service invocation is successful when the data is sent from the sender and received by the receiver without errors (no fault messages are related to this message flow). However, web service invocations can have different levels of performance depending on a number of quality factors. As indicated in Reference [8], since web service is a combination of a software application and a service, the characteristic of its quality differs both from the typical software quality [13] and typical service QoS [4-7].

Loose coupling between architectural components is what distinguishes web services from general set of services. It is also a typical SOA identifier. Service consumers can change service providers at runtime whenever their quality is un-
acceptable. However, the quality of web services, especially in the commercial applications, is seen on the business level. When web services are part of a business process, it is important, e.g. what reputation the service provider has or what is the price of running particular service. Such an approach to web services quality is not of much benefit to the evaluation of web services in disadvantaged environment.

Apart from business level characteristics of web services quality, there are many other factors that affect web service behaviour and successful web service invocation. Examples are, e.g. security measures, hardware resources of the service provider, and the most important in case of disadvantaged networks – network performance. The quality of the transport media may degrade the web services quality (perceived by the user) significantly, since they are not invoked directly (between directly connected devices) but through the network. For instance, even if a web service is very fast and has a very small response time on the server side, it is not equivalent to have a rapid response time measured at the client side if the bandwidth of a transport network is narrow. Similarly, although a web service has been implemented efficiently, it is not expected to respond very fast if a service provider application was installed on a poor performance hardware.

There have been several different and not fully coherent nor standardized initiatives on describing the web service QoS attributes and metrics [8, 9, 14]. None of them is anyhow promoted over the rest and they are not frequently referenced in the other recommendations and reports on web service QoS framework (e.g. WS-Policy, WSOL, WS-Agreements). However, they give some good background for evaluating performance of web services, also for those invoked over unreliable and dynamically changing network.

The most interesting and well defined proposal for describing quality attributes of web services has been proposed by the OASIS organization in the document entitled “Web Services Quality Factors” (WSQF) [8]. This document defines so called “web service quality factors” which are groups “of items which represent web service’s functional and non-functional properties (or values) to share the concept of web services quality among web service stakeholders” [8].

Based on two different perspectives, WSQF categorizes the quality factors into two Quality Groups: Business and System (see Fig. 2). The first one includes only the business value quality factor. The second one is composed of two subgroups: variant quality subgroup, which includes quality factors whose values can dynamically change in runtime while a service is being used (Service Level Measurement Quality), and invariant quality subgroup that refers to quality factors whose values are determined after the service has been developed. The invariant quality part comprises interoperability quality, business processing quality, manageability quality and security quality.

The System Quality Group factors are dependent on a few characteristics: implementation of the service, service deployment and underlying network performance. The first two are inevitably invariant and determined by the service
producer after the service has been coded (implementation of the service) and made available for users (service deployment). Then, they should be monitored by the management system for keeping to the declared values.

![Web service quality factors defined by OASIS (on the basis of [8])](image)

The variant quality factors change in time. They should also be monitored, however, some of them can be measured at the provider’s side, the other ones – at the consumer’s side (see Fig. 2). The quality factors that are measured at the consumer’s side are those affected by the network performance – Response Time and Successability. They are the most important ones in terms of user perception (QoS perceived by the consumer).

Successability and response time has been defined in [8] as:

- **Successability** – “probability of returning responses after web services are successfully processed. In other words, it refers to a ratio of the number of response messages” (defined in the WSDL) “to the number of” error free “request messages after successfully processing services in a given time”.

\[
\text{Successability} = \frac{\text{Number of Response Message}}{\text{Number of Requested Message}}
\]
Response time – “duration from the time of sending a request to the time of receiving a response. The response time can be varied by the point of measurement and affected by three types of latency: client latency, network latency and server latency”, i.e.:

- Client latency is the delay time caused by a client system to process a service request. It is a sum of the time taken between two events during processing a request: when a client application requests a service and when the request is sent by a client (t1 \(\rightarrow\) t2 in Fig. 3), and then, when processing response, between the time taken between the response arrives to the client and the time the application system receives the response (t7 \(\rightarrow\) t8 in Fig. 3).

- Network latency refers to the time taken on a network for transmitting request message and response message. For request message, it is the time taken between a client sends a request and the time the web services server receives the request (t2 \(\rightarrow\) t3 in Fig. 3). For response message, it is the time taken between the server sends a response and the time the client receives the response (t6 \(\rightarrow\) t7 in Fig. 3).

- Server latency is a delay time caused by a server system to process the request and develop response (may also require contact with other services – in case of service composition). It the time taken between the time the server receives the request (t3 in Fig. 3), then delivers it to the web service (t4 in Fig. 3), web service processes the request and generates the response (t5 in Fig. 3), the response is delivered to the server (t6 in Fig. 3).

Response time can be, therefore, defined by the following equations (CL, NL, SL as declared in Fig. 3,):

\[
\text{Response Time} = \text{ClientLatency} + \text{NetworkLatency} + \text{ServerLatency} \quad (2)
\]

\[
\text{Response Time} = (CL1 + CL2) + (NL1 + NL2) + (SL1 + SL2 + SL3) \quad (3)
\]

Successability and response time are quality factors that are measured at the client side and can be used for determining the efficiency of web services invocation.
Other quality factors defined for web services are measured at the server side so that the network performance does not have direct influence on them.

Another interesting document, W3C QoS for Web Services: Requirements and Possible Approaches [9], describes quality-of-service (QoS) requirements for web services and defines them in the form of the following QoS attributes: performance (throughput, response time – i.e. the server response time, latency – i.e. Round Trip Delay at the client’s side, execution time, transaction time), reliability, scalability, capacity, robustness, exception handling, accuracy, integrity, accessibility, availability, interoperability, security, and network-related QoS requirements (network delay, delay variation, packet loss) (see Fig. 4). W3C in [9] provides a high level description without detailed definitions and equations. Some explanations are not explicit enough to become definitions (e.g. it is difficult to define differences between response time and execution time). What should be noted is that W3C raises the question of mapping of QoS attributes defined in the application layer to the QoS parameters in the IP layer, used to provide appropriate traffic handling (e.g.: using RSVP, DiffServ, MPLS, etc.). The parameters’ set that has been proposed for this purpose is called Network related QoS requirements and consists of Network Delay, Delay Variation and Packet Loss. However, introduction to this mapping is out of scope of this recommendation. It is also emphasized that in order for the web services QoS framework to work properly, network management mechanisms should also be involved in controlling and managing QoS for web services.

![Diagram of W3C QoS attributes](image)

**Figure 4.** W3C QoS attributes (on the basis of [9])

### 6.2. OWL-Q – semantic model for QoS description

QoS attributes descriptions presented in subchapter 6.1 define QoS attributes with different level of detail but have no formal semantic meaning. In order to lift such descriptions to provide semantics it is necessary to model them in a semantically rich language.
An interesting model comes from the semantic description of ontology extending the OWL-S (Semantic Markup for Web Services [12] called OWL-Q [15]. This ontology describes QoS from a few points of view which are defined in so-called facets. The proposed ontology is connected with the OWL-S model (by the connecting facet), classifies QoS attributes and defines them as static and dynamic ones, which is very relevant for the web services provision (the basic facet). It also defines QoS Metric facet which, together with Function, Measurement Directive and Schedule facets, Unit facet and QoSValueType facet, precisely describe possible QoS metrics and their utilization for measuring particular QoS attributes.

The OWL-Q is very relevant for the semantic descriptions of web service QoS and can be used to describe any QoS attributes. In the case study provided for the purpose of this article, OWL-Q has been used together with a set of selected QoS attributes. Such QoS description can be incorporated in the OWL-S service description (as the service QoS offer). Additionally, it is proposed to use the QoS description in the user profile. This would be the QoS required by the user. The case study is to show the advantages and disadvantages of this solution as well as the perspectives of its utilization for service discovery and invocation in disadvantaged environment.

7. Semantic description of QoS attributes in the AFRO ontology

The Adaptation FRamework for web service provision strongly utilizes the semantic description of the context of the service call. It is mostly used for automatic decision on adaptation actions to be performed on messages flowing to the client (as responses from the target service). One of the elements of the context is user profile defining user preferences in terms of adaptation, his device properties and his QoS requirements. It is assumed that services will be described using OWL-S. On the basis of OWL-Q, it is possible to incorporate QoS profile of the service and define QoS offered by the service provider. Network profile defines the link performance QoS attributes that are related to the influence that the network has on the other QoS attributes of web services. Fig. 5. shows upper ontology linking together all the major elements of the AFRO ontology.

![Figure 5. AFRO Upper ontology including user and service profiles](image-url)
The QoS Profile of the user incorporating the user requirements in terms of QoS has been linked with the user class by the hasQoSProfile object property (see Fig. 6).

AFRO ontology can import the OWL-Q ontology and describe QoS offers and requirements on its basis. The OWL-Q (see Fig. 7) covers the QoS domain extensively enough to define static and dynamic QoS attributes measured by simple and complex metrics. Definition of the metrics enables provision of common understanding of their meaning (even if they are complicated equations) across system boundaries.
8. Semantic QoS model in context-aware service provision

For the purpose of the case study, a set of well described QoS attributes which have appropriate metrics defined has been selected, i.e.:

- Availability – measured by the Availability Index (AI) defined on the basis of WSQF [8]:

\[ AI = 1 - \frac{\text{DownTime}}{\text{MeasuredTime}} \]  

(4)

- Accessibility – measured by the ACCessability Index (ACCI) defined on the basis of WSQF [8]:

\[ ACCI = \frac{\text{NumberOfAckMessages}}{\text{NumberOfRequestedMessages}} \]  

(5)

- Successability – measured by the Successability Index (SI) defined on the basis of WSQF [8]:

\[ SI = \frac{\text{NumberOfResponseMessages}}{\text{NumberOfRequestedMessages}} \]  

(6)

- ResponseTime – measured by the TotalResponseTime (RT) (see subchapter 6.1) defined on the basis of WSQF [8];

- ServicePerformance – measured by the ExecutionTime (ET) (Server latency in Fig. 3) defined on the basis of WSQF [8]:

\[ ET = SL1 + SL2 + SL3 \]  

(7)

(SL1, SL2, SL3 as in Fig. 3).

- LinkPerformance – measured by the Round Trip Time (RTT), Packet Error Rate (PER), delay, and throughput.

![Diagram](image)

Figure 8. Application of the OWL-Q concepts in the AFRO userQoSProfile
The defined attributes have been expressed semantically in the AFRO ontology on the basis of the OWL-Q. The direct application of the OWL-Q was not possible because the owl file was not available as an open source. According to the above definitions, QoS attributes and their metrics were defined. Almost all of them were complex metrics (except for link performance metrics which were simple metrics) with appropriate units for measurement (see Fig. 8).

In the case study, the service description in OWL-S was enhanced with the QoS offer expressed on the basis of the OWL-Q model (see Fig. 9).

Figure 9. QoS attributes offered by the service – OWL-S ontology with OWL-Q enhancements

The QoS user requirements and service offers can be used for the service discovery. An interesting mechanism for semantic matching with QoS has been proposed, e.g. by the FFI [2], however, it does not define the semantic model for the QoS attributes description. An extensible description of a very useful QoS-based Web Service Matchmaking Algorithm based on the OWL-Q has been proposed in [3]. It bases on OWL-Q and is a good starting point for semantic web service discovery mechanisms.

Application of the presented QoS profiles has been also proposed in the AFRO adaptation framework. In this case, it will be also necessary to provide information about the current network state. It is assumed that the network performance parameters on the link to the user will be monitored. On the basis of the predefined rules, captured data will be used to determine the network state. The monitored values define network performance QoS attributes, i.e. Round Trip Time (RTT), Packet Error Rate (PER), delay, and throughput (see Fig. 10).
9. Discussion

The article was focused on aspects of web service QoS attributes description for web services discovery and invocation, especially in terms of matching user requirements with the service offer and showing impact of the network link performance on the values monitored on the client and service side. What is visibly lacking in the available QoS frameworks for web services, is the complete solution for monitoring and controlling the status of web services and web services invocations. This means that the monitoring should be done at the users’ and services’ sides. Moreover, the framework must be based on some form of SLAs that should be agreed between service providers and consumers. Violations of the contract in the commercial world usually are reflected in payment penalties. In the military operational context the consequences may be more serious. Such a framework must consider those issues and provide solutions when the contract is actually violated.

The solution for acting in case of violating the contract is the AFRO proxy. It provides adaptation of the response messages to minimize negative effects of the disadvantaged network.

In terms of service discovery matching of the QoS parameters measured at the user’s and service’s sides, that are not affected by the network performance, can be direct (based on selected matching algorithm, e.g. exact match). However, in terms of Response time, the comparison should take into account the fact that this QoS attribute is affected by the Link performance and service performance.
In terms of services invocation, the actual effect of the degradation needs to be studied in order to provide appropriate mappings of the requested QoS attributes at the application layer on the network QoS attributes. As already emphasized this is difficult due to the fact that web services generate elastic traffic having no direct reflection in the required network resources (e.g. for reservation).

Another problem with WS discovery is that the dynamic QoS attributes should be applied in the service description. In case of frequent QoS attributes’ values changes it is necessary to update the service descriptions. The control over these attributes handled by some management entity will also require quite a big amount of traffic. It is, therefore, a problem to be considered how to efficiently handle the monitoring process in low bandwidth networks.

10. Summary

The case study presented in the paper rises many problems in terms of QoS for web services that is part of a larger management framework. It can be summarized with the following statements.

There are several approaches to QoS attributes description but none of them is ideal. There is no actual standard in terms of monitoring and control over QoS attributes for web services, so interoperability is not achieved. There is also no interoperability of web services management entities with network QoS mechanisms (or brokers) to provide QoS support in the network. To achieve desired QoS for web services, the QoS mechanisms operating at the web service application level must cooperate with the QoS mechanisms of the transport network. Application level QoS parameters should be mapped appropriately to corresponding network level QoS parameters. Additionally, solution for handling support for WS delivery in lower layers (including cross-layer functionality) are needed.

REFERENCES

[1] Sliwa J., Duda D., Adaptive web services supported by QoS IP network, MCC 2009, Prague, Czech Republic.


