Agent based Framework for QoS Measurement applied in SOA

A uniform Approach based on a QoS Meta Model

Andreas Hausotter, Arne Koschel
University of Applied Sciences & Arts Hannover
Faculty IV, Department of Computer Science,
Hannover, Germany
email: Andreas.Hausotter@hs-hannover.de
email: Arne.Koschel@hs-hannover.de

Johannes Busch, Markus Petzsch, Malte Zuch
University of Applied Sciences & Arts Hannover
Faculty IV, Department of Computer Science,
Hannover, Germany
email: Johannes.Busch@stud.hs-hannover.de
email: Markus.Petzsch@stud.hs-hannover.de
email: Malte.Zuch@hs-hannover.de

Abstract—Nowadays, enterprises are faced with a variety of major challenges, such as the cut-throat competition in a global market, a decreasing customer loyalty, and the strategic adjustment moving from a product centric perspective to a customer centric perspective. Therefore, businesses need to change their operational processes in a flexible and agile manner to keep their competitive edge. A Service-oriented Architecture (SOA) may help to meet these needs. As the application landscape of enterprises is inherently heterogeneous and highly distributed it is a great challenge to provide services with a certain quality. This is particularly the case when services are requested externally via the web. Therefore, quality of service (QoS) measurement and analysis is a crucial issue in Service-oriented Architectures. As the key contribution of this paper we present a generic SOA Quality Model (SOA QM) based on the measurement standard ISO/IEC 15939, a SOA Information Model (SOA IM), and an architectural concept of a QoS System. The SOA IM is an XML-based specification for the measurement to be performed. The QoS System provides an execution platform for the SOA IM, based on a Complex Event Processing (CEP) approach and guarantees minimal impact on the SOA environment. The concepts are explained in detail using a standard process of the German insurance domain.

Keywords—Service-oriented Architecture (SOA); Quality of Service (QoS); Measurement Process; Complex Event Processing (CEP).

I. INTRODUCTION

Distributed IT-systems are commonly used in today’s companies to fulfill the needs of agility and scalability of their business processes to manage the highly variable demand of the market. Typical scenarios are real time logistics and delivery, just in time supply chain management and in general, handling services in real time to fit market demands.

The latter is commonly used within the finance and insurance industry during their internal computation of risk and money management and for their external customer services, like proposal calculations (including the current market conditions). Especially the external services must have a high quality in terms of time behavior. Google has shown that a latency of 100 ms up to 400 ms causes an impact of -0.2 % up to -0.6 % concerning the daily usage of web services by the customers [1]. The integration of those services to run business processes in a stable way, fulfilling the varying demand of the market, is commonly realized with Service-oriented Architectures (SOA).

Those architectures integrate (micro)-services within distributed systems to run business processes with a high capability in terms of agility. Especially the distribution of the services over several systems allows to scale with the market demands.

Distributing and handling several services is a common concern of the insurance industry. But an increasing distribution and more complex business processes will only gain more agility with SOA, if the distribution of the services over several systems is realized in a reasonable way. For getting the required control of the distribution of those services, a measurement system is required. Measuring the general QoS in distributed systems is part of the motivation of this work and is explained in detail in the next subsection. The subsection after the discussion of the general motivation will show the contribution to the general problem in measuring the QoS in SOA within the application scenario of the insurance industry, explained in section III.

A. Motivation

In many cases, it is not foreseeable to forecast, how much computing power and bandwidth the infrastructure needs to host the allocated services within the distributed computing system. Beside these design decisions of the infrastructure, there is a further problem in allocating the services to the right locations within the distributed system. This allocation will influence how much bandwidth and calculation power is available for the services and how many services will share identical resources during the same time. So if several services will use the same part of the infrastructure, this could lead to increasing latencies over the whole system, resulting in an unfavorable time behavior for the users. Especially if some services are requested with intense demands of the market, latencies could rise in an unpredictable manner.

Such a scenario is typical for the German insurance industry. At the end of the year, millions of users are able to switch their insurance contracts and will request therefore designated online services. The general demand is not foreseeable and the intensely interaction between the insurance industry and
the finance industry requires a high quality of those services. Especially the historically low interest rates in today’s market provokes fast changing business models and the need of a fast adoption to new business processes and the ability to offer services in a high quality to fulfill the external user demands and the internal interaction within the finance industry. To fulfill those demands, distributed systems with SOA will benefit from an across boarder measurement of the quality of those services, especially in terms of latency. Such a measurement system is the contribution of this work and is explained in the following subsection.

B. Contribution

The need for a new development of a flexible measurement system is influenced by the limitations of common solutions. The scenario of this work is based on a German insurance company, which already uses Dynatrace as a measurement solution [2].

The partner from the insurance industry is currently restructuring and modernizing his business processes and therefore, he needs a more flexible and generic approach to integrate an external measurement system for monitoring and analyzing the time behavior of his services. Additionally, a more detailed analyzer component was required to process the measured data.

So on the one hand, the approach has to be integrated in a generic way with minimal interaction points within the SOA of the partner from the insurance industry to guarantee a simple integration during the continuous development process. But on the other hand, the solution should offer a flexible and detailed analyzer component.

This article will present our currently ongoing applied research work. Since it is still "work in progress", we will mostly focus on measurement concepts and an adequate measurement model here. We combine this with an initial description of the main insurance application scenario used by us. More technical details on our actual prototypic implementation as well as QoS measurement results, will be presented in future work.

The required solution was defined by the following:

• generic approach to generate the measurement system,
• automatic integration of the measurement system in the existing SOA,
• lose couplings within the existing SOA,
• flexible agent based approach,
• technology independent approach using standards (XML),
• individual and customizable analyzer component.

As stated above, beside technical concepts we will also present some details from our mainly utilized application scenario, which is based upon the ideas from the "Check 24" process. Within this process different offerings for the same kind of insurance are compared. The offerings typically origin from several insurance companies. They are, for example, different offerings for car insurances. Based on certain input parameters, the end user gets eventually different insurance offers by this process. The proposal service used by "Check 24" is a common service throughout the German insurance sector is implemented by various insurance companies.

This service can be called externally by applications such as "Check 24" through a common interface given by a so called "BiPro specification". BiPro is widely used throughout the German insurance sector and the availability of these services has a significant impact on competitiveness. Internally the proposal service is, for example, used in the process "Angebot erstellen" ("create proposal") of the general German "Versicherungsanwendungsarchitektur (VAA)” (cf. [32]), which describes a set of standardized insurance processes working within a generalized “insurance application architecture”. Our project partner has implemented a similar process for it’s own agent respectively customer portal.

The remainder of this paper is structured as follows: In the next Section II we discuss some related work. Section III describes our application scenario in some detail. In Section IV and Section V our general Quality of Service (QoS) measurement model and concept are described. Eventually Section VI concludes this paper and gives some outlook to future work.

II. PRIOR AND RELATED WORK

In prior work, we already discussed several aspects of the combination of SOA, Business Process Management (BPM), Workflow Management Systems (WfMS), Business Rules Management (BRM), and Business Activity Monitoring (BAM) [16][17][15] as well as Distributed Event Monitoring and Distributed Event-Condition-Action (ECA) rule processing [20][21]. Building on this experience, we now address the area of QoS measurement for combined BRM, BPM, and SOA environments within the (German) insurance domain context.

Work related to our research falls into several categories. We will discuss those categories in turn.

General work on (event) monitoring has a long history (cf. [12][13] or the ACM DEBS conference series for overviews). Monitoring techniques in such (distributed) event based systems are well understood, thus such work can well contribute general monitoring principles to the work presented here. This includes also commercial solutions, such as the Dynatrace [2] system or open source monitoring software like, for example, the NAGIOS [14] solution. In those systems there is however, generally not a focus on QoS measurement within SOAs. Also, they usually do not take application domain specific requirements into account (as we do with the insurance domain).

Active DBMS (ADBMS) offer some elements for use in our work (see [18][19] for overviews). Event monitoring techniques in ADBMSs are partially useful, but concentrate mostly on monitoring ADBMS internal events, and tend to neglect external and heterogeneous event sources. A major contribution of ADBMSs is their very well defined and proven semantics for definition and execution of Event-Condition-Action (ECA) rules. This leads to general classifications for parameters and options in ADBMS core functionality [19]. We may capture options that are relevant to event monitoring within parts of our general event model. QoS aspects are handled within ADBMS, for example, within the context of
database transactions. However, since ADBMSs mostly do not concentrate on heterogeneity (and distribution), let alone SOAs, our work extends research into such directions.

The closest relationship to our research has work, which directly combines the aspects QoS and SOA. Since about 2002 several articles fall into this category. However, in almost all known articles the SOA part focuses on WS-* technologies. This is in contrast to our work, which takes the operational environment of our insurance industry partners into account.

Examples of WS-* related QoS work include QoS-based dynamic service bind [26][27], related WS-* standards such as WS-Policy [22], and general research questions for QoS in SOA environments [23].

Design aspects and models for QoS and SOA are, for example, addressed in [28][24][33][25][26]. SOA performance including QoS in [34], and monitoring for SOA is discussed in articles such as [30][31][29][35].

III. APPLICATION SCENARIO

Customers are using online platforms to compare the conditions and proposals offered by different companies. The online platform check24.com allows customers to compare different insurance proposals. Therefore, the insurance companies need to respond to those requests to be aware for potential customers on such platforms. The underlying scenario for this work is a service for calculating individual proposals for such online platforms. This scenario is automatically requested by the online customer information platform and needs to respond in a timely manner. The business process for calculating the proposal follows four steps:

- check input parameters for plausibility,
- call all additional relevant services to get required data,
- calculate the proposal based on internal business rules,
- deliver the proposal to the requesting online platform.

The partner from the insurance industry has already developed a distributed system to create and run such business processes. This system uses the approach of SOA and integrates various micro-services located across several locations.

Measuring the time behavior is a feasible approach to maintain the overall system and scale it to changing market demands to fulfill the required quality of such services (QoS). The distributed system is designed with the concept illustrated in Fig. 1. The system part alpha is the enterprise service bus (ESB) of the system, which is responsible to integrate the business processes with further applications and services. Those business processes are parameterized by specific business rules, stored in a business rule database.

The communication with this business rule database is realized via web service calls. So in general, alpha is the central communication component of the system.

The system part beta is the current process engine to run the business processes and is connected via JMS with alpha. Those business processes are influenced by the stored business rules and the business process data, which are stored in a separated database.

This distributed system defines the scenario where several services are parameterized, called and integrated (via alpha) over several locations. The generic and XML-based measurement concept of this work will use this scenario to measure QoS-Parameters, especially the time behavior of services. The specific measurement model is described in the next section.

IV. MEASUREMENT MODEL

The assessment of the QoS in Service-oriented Architectures is based on a SOA Quality Model (SOA QM), which combines characteristics and sub-characteristics in a multilevel hierarchy. For this purpose we adjusted the ISO/IEC-Standard 9126 to meet the SOA-specific requirements. Fig. 2 illustrates the characteristics, sub-characteristics and relationships between these concepts. In our research work we will focus on Time Behavior, which contributes to Efficiency.

Although ISO/IEC 9126 was revised by the ISO/IEC-Standard 25010 (Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models, cf. [9][6]) we use ISO/IEC 9126 as a starting point because of it’s high degree of awareness in German-speaking countries (cf. [7]). Moreover, the German version of the ISO/IEC 25000 series has been prepared by the German Institute for Standardization (DIN) but is not yet available (cf. [11]).

Instead of applying the quality metrics division of SQuaRE (i.e. ISO/IEC 2503x), our approach is based on the comprehensive ISO/IEC-Standard 15939 (cf. [9]). The basic model, as to be found in similar form in the contribution of Garcia et al. ([10]) has been aligned and extended by quality requirements, quality models, and some system components. In the following subsections we describe the main concepts of our SOA Measurement Information Model (SOA MM) as shown in Fig. 4.

A. Information Need and Information Product

The determination of the QoS in a SOA is always demand-driven, since both the specification (’What and how should be measured?’) as well as the execution of the measurement itself and the subsequent interpretation of the results can cause a significant organizational and technical effort.
Therefore, first of all the Information Need with objectives, potential risks and expected problems is to be defined and documented properly. In terms of the application scenario presented in section III the objective is to assess the performance of the business process for calculating the offer in order to identify and resolve problems in time.

B. Core measurement process

The Information Product is the result of the execution of the Core Measurement Process as depicted in Fig. 3 (cf. [8]). The Information Need provides the input for the subprocess Plan the Measurement Process (planning stage), the subprocess Perform the Measurement Process (execution stage) generates the output, i.e., the Information Product. The process goal is to satisfy the Information Need. All concepts presented below directly or indirectly contribute to the Information Product.

C. Concepts of the planning stage

In the focus of our research work are SOA Services whose QoS is to be investigated. For this purpose, Quality Attributes are measured. In this context, the Measurable Concept outlines in an abstract way, how the attributes values are determined to satisfy the required Information Needs. In doing so, it references one or more sub-characteristics of the SOA QM.

For the application scenario described in section III, the process performance is to be determined first and then evaluated. The corresponding Measurable Concept is the calculation of the processing time. To do this, instantiation of a process and termination of the process instance are to be determined. The process identification represents the Quality Attribute to be measured, and the sub-characteristic, referenced by the Measurable Concept, is the Time Behavior.

In order to implement the Measurable Concept and to perform measurements of attributes, first of all Measures are to be specified. A Measure assigns each Quality Attribute a value on a Scale of a particular Type. The ISO/IEC-Standard 15939 provides 3 different types of Measures, namely Base Measures, Derived Measures, and Indicators respectively.

A Base Measure specifies by its Measurement Method how the value of a Quality Attribute is to be determined. It’s always atomic and therefore independent of other Measures.

A Derived Measure uses one or more Basic Measures or other Derived Measures, whilst the Measurement Function specifies the calculation method and thus the combination of the Measures used.

For the application scenario illustrated in section III, the Basic Measures process instantiation $t_{inst}$ and process
stance termination \( t_{\text{term}} \) are specified. The identification of the processes instance \( \pi ID \) represents the Quality Attribute measured by \( t_{\text{inst}} \) and \( t_{\text{term}} \). As the Measurement Method, we select the time of the start and end event respectively. The Derived Measure processing time of the instance \( T_{\text{Proc}} \) will be calculated by the Measurement Function

\[
T_{\text{Proc}}(\pi ID) = \Delta t = t_{\text{term}}(\pi ID) - t_{\text{inst}}(\pi ID).
\]

Finally, an Indicator is a qualitative evaluation of Quality Attributes, which directly addresses the issue raised in the Information Needs. Indicators always use a nominal scale with qualifying values and thus show if necessary action or the need for further root cause analysis. An Indicator is derived from other Quality Measures, i.e., Base and Derived Measures, and Indicators. The combination of the Quality Measures used and the method of calculation is based on an Analysis Model in conjunction with Decision Criteria using thresholds and target values.

For the application scenario illustrated in section III, the indicator adequacy of the processing time of a process instance \( SLoT_{\text{Proc}}(T_{\text{Proc}}) \) according to table I:

<table>
<thead>
<tr>
<th>( T_{\text{Proc}} )</th>
<th>( SLoT_{\text{Proc}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \in (0, 3000,\text{ms}] )</td>
<td>high</td>
</tr>
<tr>
<td>( \in (3000, 7000,\text{ms}] )</td>
<td>medium</td>
</tr>
<tr>
<td>( \in (7000, \infty) )</td>
<td>low</td>
</tr>
</tbody>
</table>

**D. Concepts of the execution stage**

After the concepts of the planning stage have been presented, now those of the execution phase will be explained briefly (subprocess Perform the Measurement Process, depicted in Fig. 3). Section V will discuss their conceptual implementation more detailed.

The actual measuring procedure, i.e., the execution of the instructions for determining the value of a Quality Attribute, is called Measurement. Hereby, Measurement Results are created, collected in a container, namely Data, which is inserted into a Data Store.

The measurement system comprises different supporting software components, which are conceptually presented in section V. The QoS Measurement performs the instructions specified in the Measurement Method or Measurement Function respectively, to generate the Measurement Results for further processing. The QoS Analyser performs the statistical analysis and evaluation of the collected data and creates the Information Product. The QoS Reporting makes the Information Product available to the Measurement User (cf. Fig. 3)

**E. QoS Measurement Information Model**

We designed a domain-specific language to specify the values of the concepts introduced above according to the Information Need. This specification document is referred to as QoS Information Model (QoS IM). The aim of this approach is to automate the measurement process by the generation of artifacts required by the QoS system to execute a measurement.

The QoS IM consists of an abstract and a concrete section. In the abstract section, the concepts of the Planning Stage and partly the Execution Stage are specified. In the concrete section, the implementation specific definitions are done. Since our QoS-System is based on a complex event processing (CEP) approach, the specification of events, agents and rules is subject of this section.

A sophisticated XML Schema was developed to realize the domain-specific language. We opted for XML as a universally accepted standard that is highly flexible, platform and vendor independent and supported by a wide variety of tools. In a follow-up project an XText-based tool will be developed that generates the (XML) QoS IM from a (XText) source code.

Its semantic model is shown in Fig. 4. The following rules for modeling apply:

- Concepts are mapped to XML elements (graphically represented by UML classes).
- Details of a concept are mapped to XML attributes of the owning element (graphically represented by UML instance variables).
- If possible, relationships between concepts are mapped to element hierarchies (graphically represented by UML associations).
- Otherwise they are mapped to constraints (i.e. keyrefs) (graphically represented by UML dependencies).

**V. MEASUREMENT CONCEPT**

In section IV, a QoS IM based upon a SOA QM is described. To execute a specific QoS IM (and thus subprocess "Perform the Measurement Process") an execution platform is needed. This platform and the underlying QoS architecture is given in this section. First reasons for choosing this specific architecture are discussed shortly. Furthermore an overview is shown, detailing in the central agent concept and CEP.

**A. Design Decisions**

As described above the goal of the measurement concept is to provide the execution platform for a specific QoS IM. Therefore basic design criteria for the measurement concept are derived from the QoS IM. Furthermore quality requirements are given, which also have to be considered in the architecture design. These criteria are:

- measurement of Quality Attributes as described by QoS IM,
- flexibility of measurement and computation,
- low impact (modification, performance, etc.) onto SOA components.

The proposed Measurement Concept is based upon a general architecture given in [3]. The basic idea is to separate the measurement (e.g. sensors, agents, etc.) and "analysis and statistics" functionality into different modules. This separation opens the opportunity to cater each module to their specific functional and quality requirements.
Overall the given general architecture already fulfills the requirement to measure Quality Attributes and provide the needed evaluations to produce Measurement Results and Information Products.

The measurement module has to provide the QoS System with information about the observed service. To provide the needed flexibility a sensor has to be placed into it. To keep the impact onto the SOA at a low level an agent based approach was chosen. Agents capsule the needed parsing and computation and thus can be easily integrated into arbitrary SOA modules. Furthermore minimizing the performance impact (through threading, non-blocking, etc.) can be integrated into the agents.

The "analysis and statistics" module does not have these strict requirements on performance impact. Flexibility of computation and measurement execution is the main quality requirement. Thus a platform approach was chosen. Basically artifacts generated through the QoS IM are placed into the QoS platform and executed.

**B. Overall system architecture**

On a high level the QoS system splits the measurement agents and further processing (QoS platform) into different components. This approach allows to easily split these components into different processes to comply to the quality requirements. While the measurement agents (encapsulating the agent concept) represents the client component, the server component is represented by the QoS platform and contains the CEP engine and further analysis processing. Fig. 5 shows a high level overview of important components and their relationships.

The general purpose of the measurement agents is to emit specific events based on the defined Base Measures. As described in section IV events are emitted, e.g., for process instance instantiation/termination. In general, concepts for agent implementation can be categorized by agent location and time of execution (cf. [5] and [4]). To measure a specific process instance agents can be placed into corresponding SOA service calls thus measurement agents are only logically placed into the QoS System component. One and currently used approach is to use the concept of interceptors, which offers a low modification impact and can deliver precise Measurement Results.

The QoS Platform consist of several components, most notable the QoS Measurement and QoS Analyzer. In general the purpose of these modules are to collect, clean and compute...
Before any event is given to the Measurement Method, it will be handled by the control module. Purpose of this module is event routing, general cleaning steps and an optional filter step. Cleaning (or formatting) events in the analyzer is needed because measurement agents are placed in the monitored system, thus shall minimize their performance impact. The Measurement Method is implemented as a CEP rule executed by the engine and emits complex events for further near real-time processing and long term analysis.

The QoS Analyzer module provides a basis for statistical analysis and evaluations. Every complex event is stored into a Data Store implemented as a relational database. The different analysis and evaluations defined by Derived Measures and Indicators are implemented through SQL and plain Java. Furthermore the module provides an interface to the computed Information Product.

C. Applying the described measurement concept

In Fig. 6 the given measurement concept (QoS platform) is applied onto the application scenario thus providing the missing link between the QoS IM and the insurance based application scenario. In this example a simplified scenario is used consisting only of an external "Check24" mock-up service (representing a simple consumer), the central ESB and the proposal service (which represents the producer). The task of the measurement model and thus the concept is to measure the processing time of this service and to compute the Information Product for further evaluations.

To measure this service the measurement agents, defined through base measures, are placed directly into the ESB. This offers a measurement independent of service location and different load balancing scenarios. To minimize the integration effort functionality given by Spring Integration is extensively used (especially the interceptors for message queues). In this simplified example base measures (and thus the agents) only determine service call start / end times and announces these to the QoS platform. Furthermore the agents try to minimize their performance impact by using non-blocking techniques and performing only necessary parsing steps (e.g. service call id’s, etc.). These will be shown in detail in further publications.

As described above the QoS platform performs further cleaning and processing steps to compute the QoS IM indicators \( SLoT_{Proc}(T_{Proc}) \) and provides these to downstream systems (e.g. reporting, presentation, load balancing, etc.).

VI. CONCLUSION AND FUTURE WORK

The presented approach for monitoring a distributed SOA environment is a promising path to take: The SOAQM is aiming to follow the ISO/IEC-Standard 15939 (cf. [8]), which enables a wide range of use cases. The Measurement Concept outlines an execution platform for the specific QoS IM, which should cause minimal impact on the SOA environment. The separation of Measurement Agents and QoS-Analyzer allows lightweight agents on the one hand and a very capable analyzer component on the other hand.

The still ongoing work of applying the QoS System to an application scenario relevant to our partner in the insurance industry (the "Check 24 process"), will provide evidence of the practical usability of the created framework. In this paper the framework and the corresponding platform are applied onto a basic, business relevant scenario (the proposal service). Furthermore it is planned to apply these technique to the more complex process "Angebot erstellen" ("create individual proposal") of the VAA thus implementing a more complex scenario. It is expected that the monitoring system will help
to discover potential bottlenecks in the current system design of our partners distributed services and therefore creating high value in the process of solving these issues.

In future work, the actual measurement and analysis of the results are to be done. It is also planned to apply these results onto cloud based environments.

REFERENCES


