Monitoring as-a-service to drive more efficient future system design

Frédéric Lemoine*, Tatiana Aubonnet*§, Ludovic Henrio†, Soumia Kessal§, Eric Madelaine‡, and Noëmie Simoni§
*CEDRIC, Conservatoire National des Arts et Métiers, 292 rue Saint-Martin, 75003 Paris, France
Email: frederic.lemoine@cnam.fr
†CNRS, University of Nice Sophia-Antipolis, Sophia-Antipolis, France
‡INRIA, Sophia-Antipolis, France
§Télécom ParisTech, 46 rue Barrault, 75013 Paris, France

***Abstract*—In the services world, the expected benefits are the fastest time to market, lower costs, greater consistency in the application, and increased agility. The re-use and sharing properties of software components are useful to address these challenges. However, to achieve this, it is necessary to be able to observe each service and to control the service composition. This article proposes to rethink the company's organizational process of application development and use the power of monitoring to help the application design. The proposed Monitoring as-a-service (MaaS), whose properties are detailed, will be used for the computation of the offered Quality of Service (QoS) and the services calibration during the service creation phase and to inform the service QoS Controller during the operational phase. For effective design, the architect will place MaaS at crucial points of its architecture according to its decision-making process. Finally, we present experimental results and a conclusion ends the paper.**

I. Introduction

Cloud computing and Future Internet promise a new ecosystem where everything is "as a service". Architects mutate to the service oriented architecture (SOA). The reusability and loose coupling properties facilitate the implementation of applications. Indeed, applications are built through the composition of services that exist today in the enterprise or can be provided by Cloud providers.

No doubt, we are in the era of the services and the service is at the heart of the architecture. Therefore each component of service has to be defined, controlled and managed. However, to manage, it is necessary to know the values and metrics of the service:

- values allow checking a service status, triggering an alert and sending a notification related to an abnormal behaviour (out contract), which implies immediate action. This is the supervision and control responsibility.
- metrics allow logging and observing each measuring point. This is the metrology responsibility.

It is important that this monitoring, which regroup these two concepts of supervision and metrology, is placed at each service and composition level. Software components provide means to structure service composition and ensure better re-usability, adaptability, and scalability of services. In our preceding works [1], we introduced this vision of the monitor, by proposing a Self Controlled service Component (SCC).

But, the problems of heterogeneous services, their service level agreement (SLA) compliance, and service composition automated management are raised.

For improving the system design and make it more efficient, we need to adapt existing composition models in order to make this design and this automated management converge. For this, we need to answer the following questions:

- What are the properties needed for monitoring services to adapt to heterogeneous environments?
- Where the measuring points have to be placed to have the right information for fast reactions?
- How to know the values and metrics of service in general and of "re-used" service in particular?
- Can we take these problems into account during the design phase?

We show in this paper how the adoption of a component oriented structure helps the service composition to provide a guaranteed quality of service.

Our main contributions are the following:

- We design a generic monitoring component template that can be placed in each hierarchical level.
- We define a calibration technique to compute the nominal/offered QoS and to help their composition.
- We provide a method for the design architect to structure his application (services composition) by respecting SLA compliance.

This paper is organized as follows: The related works of the properties of monitoring systems and their analyses are described in Section II. Section III presents the SCC proposed in the OpenCloudware project, but also extended SOA properties and autonomic capabilities of these SCC components. Section IV is devoted to our propositions for efficient driving, i.e. the advantages of Monitoring as-a-service (MaaS) within SCC architecture, method for design architect, monitoring as-a-service for calibration and design. A prototype implementation of a single SCC, of a SCC components composition, and their calibrations are proposed in Section V. Finally, in Section VI, we highlight the advantages of our approach to drive future system design.

II. Related Works

Monitoring is needed to perform business analytics for improving the operation of systems and applications [2] or for verifying the compliance with an SLA contract.

There are different types of layers to be monitored: Application, middleware, OS, network, hardware [3][4][5]. These layers can be seen as where to put the probes of the monitoring system. In fact, the layer at which the probes are located has direct consequences on the phenomena that can be monitored and observed:

- Application, middleware, and OS: bugs, malfunctions, vulnerabilities, etc.
- Network: bandwidth, throughput, etc.
- Hardware: CPU, memory, temperature, voltage, etc.

The measured value in the upper layers (e.g. the performance of the application) may or may not include the values of the lower layers (e.g. the transfer rates on the network). The processing time for a task (top layer) depends on the hardware (lower layer) on which it runs and the load of the virtualized environment.

We present here our analysis of the properties related to system monitoring. These properties must be the same as those of the monitored system (*Scalability*, *Elasticity*, *Adaptability*, and *Autonomicity*) or system component (*Availability* and *Resilience*). Its integration must be done at lower cost (*Intrusiveness*, *Comprehensiveness*). The *Timeliness* property is needed for agility and quick decision making at runtime. We analyse their issues and discuss how they have been addressed in literature.

Timeliness. A monitoring system is timely if detected events are available on time for their intended use [6].
The difficulties are:

- The time between the occurrence of an event and its treatment can vary depending on the measurement, analysis, and the communication delay.
- To obtain up-to-date information, a trade-off between accuracy and sampling frequency is necessary because the shorter the sampling interval, the smaller the delay between the time a monitored condition happens and is captured.
- Analysis is problematic because it can be complex and require computing time to be relevant.
- The communication delay can be a problem if it is necessary to aggregate multiple data sources in order to process them.

Adaptability. Monitoring requires computing and communication resources that can be costly. Adaptability should be used to find the right compromise between accuracy and invasiveness (environmental disruption).

Autonomicity. A monitoring system is autonomic if it is able to self-manage its distributed resources by automatically reacting to unpredictable changes, i.e., if it is able to react to detected changes, failures, performance degradation without manual intervention [7].
The difficulties are:

- The control loop receives data from a large number of sensors and propagates the action to a large number of actuators which leads to coordination and scaling difficulties.
- The analytical capacity must be adapted to the complexity of the infrastructure (Layers)
- It is difficult to implement steering policies that respond adequately to events detected by the monitoring system.

Elasticity. Elasticity consists in coping with dynamic changes of monitored entities (created or destroyed by expansion and contraction) [8].
Types of changes are:

- New assignment of resources for the user.
- Change in the monitoring needs for the user.
- Change of the number of users.

Intrusiveness and Comprehensiveness. A monitoring system is intrusive if its adoption requires significant modifications of the monitored system [9].

A monitoring system is comprehensive if it supports different types of resources (physical and virtualized) and is multiple tenants [10]. The latter requires:

- To adopt a single monitoring API regardless of the measure that is currently used.
- To deploy and maintain a single monitoring infrastructure.

Having a low Intrusiveness minimizes cost instrumentation.
The difficulties are:

- Comprehensiveness requires supporting different underlying architectures, technology, resources, and multitenancy.
- The heterogeneity of resources and settings of the different layers.

Resilience and Availability. A monitoring system is resilient if it can support a number of faulty components while continuing to operate normally.

It is available if it provides services according to the system design whenever users request them [11].

A system must be resilient and available at least for reasons of payment, SLA compliance, and resource management.
The difficulties are:

- Services can be migrated from a physical computer to another, striking down classical monitoring logics and affecting the reliability of the monitoring system.
- Because of complexity of tracking and managing heterogeneous monitored and monitoring resources, we should take into account possible faults of the monitoring system itself.

Scalability. The aim for a scalable monitoring system is to manage a large number of probes [8]. A system is scalable if it is able to efficiently collect, transfer, and analyse large amounts of data without affecting the functional part.
The difficulties are the large number of parameters to be monitored and the large amount of data from multiple distributed locations to aggregate and filter.

TABLE I
PLATFORMS COMPARATIVE

Platform	Properties	Multi-Layers
AzureWatch [12]	Scalability, Adaptability, Autonomicity	Yes
Boundary [13]	Timeliness, Resilience, Availability	Yes
CloudClimate [14]	Timeliness, Resilience, Availability	No
CloudCruiser [15]	Timeliness, Resilience, Availability	No
Cloudfloor [16]	Timeliness, Resilience, Availability	No
CloudHarmony [17]	Timeliness, Comprehensiveness	No
CloudSleutch [18]	Timeliness	No
CloudStack ZenPack [19]	Timeliness	No
CloudWatch [20]	Elasticity, Timeliness	Yes
Cloudyn [21]	Timeliness, Resilience, Availability	No
Dargos [22]	Adaptability, Intrusiveness	No
New Relic [23]	Timeliness, Resilience, Availability	No
Up.time [24]	Timeliness, Resilience, Availability	Yes
VR. Hyperic [25]	Timeliness	No

The table I show a comparative of different platforms according to their properties.

In the following, we present different works, described in literature, aimed to satisfy or improve preceding properties.

To improve **Timeliness**, [26] propose a behavioural model to predict the best measurement time interval. [6] reduce the time of analysis and communication by assembling and processing information of near nodes and by adapting the analysis and communication topology.

Concerning **Adaptability**, [27], [9], [28], [29], [6] propose to fine-tune the amount of monitored resources and the monitoring frequency. [27] propose to predict the resource consumption for adapting the time interval to push monitoring information Monalytics [29] configures its agents in real time depending on the monitoring topology (collect, process, and transmit) by providing new analysis and monitoring codes or by changing the methods being used.

For **Autonomicity**, focusing on bottlenecks, [30] proposes two methods to detect and resolve them as well as the identification and reduction of resources if too many have been provisioned. These methods require a maximum response time and are useful for the SLA compliance. [31] proposes a monitoring system based on agents having the ability to continuously check the status of virtual machines (VM) and to restore them in case of malfunction. [32] allocates computing resources to services and deploys them on virtualized infrastructures. [32] detects violations of SLAs and offers automatic dynamic reactions combining low-level resource metrics with service level objective (SLO) and a knowledge base for the analysis of monitoring information.

Concerning **Elasticity**, Most of tools were designed for slow changes of the physical infrastructure (Ganglia [33], Nagios [34]) and do not support rapid and dynamic changes. They use a push strategy (the physical host notifies the tool on the status and the presence of the running VMs) [35] or publish-subscribe to decouple communications ends and thus to support dynamism. An hypervisor controller checks the list of virtual execution environment (VEE) and add or remove a monitor according to the detected number [8]. An extension of Nagios [35] allows the use of active verification method (pulling) by remote code execution. An extension of Nagios [9] offers a push-pull model. The monitoring information is sent by agents to a Manager (push) and information consumers can obtain data from it (pull). Monalytics [29] was designed for scalability and efficiency in highly dynamic scenarios: discovery at runtime of resources to monitor and configuration at runtime of monitoring agents. Brokers" at different hierarchical levels, collect process and transmit the monitoring information.

To improve **Intrusiveness** and **Comprehensiveness**, [10] proposes an architecture based on agents that monitor directly the flow of information through the same workflow system. They are connected with adapters which abstract from data of a specific technology. [36] monitors events at the VM level.

In the literature, several works search for the reasons impacting **Resilience**: Resource Volatility [37], [27], virtualization technology [31]. To improve the **Availability**, [38] provides a publish-subscribe paradigm for communication and a set of redundant brokers for events management while providing tolerance to attacks and malfunctions.

To ensure **Scalability**, two methods are commonly used to reduce the amount of data collected by the controller:

- Data aggregation consists in combining multiple metrics into a single one,
- Filtering avoid spreading unnecessary data to the Controller.

Most of the proposed architectures use a subsystem to propagate event announcements [10], [9], [6], [39] or agents for collecting, filtering, and aggregate data [10], [9], [28].

Although each property has been addressed in various studies presented above, no platform includes them all to the best of our knowledge. We think that a monitoring and an analysis placed close to each functional component would have many advantages:

- The volume of data exchanged and thus the communication resources would be extremely low since the analysis would be done on site. Only its result would be sent.
- The code would be simplified and hence require less computing resources (Adaptability).
- The analysis would be faster, more relevant, and reaction times would be minimized (Timeliness).
- At each addition / removal of a functional component, a monitoring and controlling component would be therefore added / removed (Scalability, Elasticity).
- Monitoring and controlling component would be located at any hierarchical level: At the same place that any functional component.

We would not be intrusive if monitoring and analysis were external to the functional component. (Intrusiveness). A generic monitoring and analysis independent of the functional component would be comprehensive (Comprehensiveness) and might

be present at all levels of architecture.

We will show how such a system would also be a valuable aid to application design for the architect. This one could experience during the design and before being put into production if its composition is properly sized i.e. whether resources will be sufficient to operate and meet the requested QoS.

Our motivation is thus double:

- Show that our MaaS, by its design, responds to most of the preceding properties.
- Show that it can also be used to help the architect to choose the best component when designing his application.

III. Background

In the services era, the service is the center of architecture, to enjoy all the benefits expected from this concept, we have proposed in [40] a component called SCC which we recall the description (Section III-A) with SoA extended properties (Section III-B) and autonomic capabilities (Section III-C).

A. Self-controlled service Component

To increase the structural decomposition and the reuse of non-functional QoS components, we have separated its internal functions and proposed an architecture that separates the monitoring and Qos functions of the remaining functions called "control". We have specified this model in the OpenCloudware project to address the behavioural aspects through QoS.

The membrane of our SCC includes (Figure 1):

- An input monitoring component (InMonitor) and an output monitor (OutMonitor). They play an interceptor role. Incoming service requests are intercepted and transmitted (unchanged) to the functional component via the corresponding internal interfaces. The OutMonitor intercepts outgoing service requests. They provide measurement information on the flow they intercept.
- A QoS component (QosControl), associated with the business component.
- A non-functional interface (client) for QoS control (IQoSStatus), by which it will send the information of violation of QoS contracts, i.e. "InContract" notifications when the behaviour is compliant with the contract or "OutContract" otherwise.
- A non-functional interface (server) of configuration (IConfigQoS, IConfigMonitor), whose role is to receive component configuration commands.

The QoSControl component checks the current behaviour of the resource and its conformity with the contract. For this, it triggers a timer and regularly requests to the monitors (InMonitor and OutMonitor) the parameter values (getValues method) of the IControlMonitor interface (Figure 2). It compares each current value to the corresponding threshold value not to exceed. It sends an OutContract notification if the current value is less (or more) than the threshold value; in this case the dynamic management consists in replacing on the fly the failing component by an ubiquitous service fulfilling the requirements. Otherwise, it sends an InContract notification. We define two types of QoS:

1) The requested QoS: client side, SLO.
2) The offered QoS also called nominal QoS is computed under resource conditions of the underlying level: provider side, SCC components based.

The QoS requested by the customer is provided by catalogue components with an offered QoS and/or components with adaptation mechanisms (SCC+). A SCC+ component is indeed necessarily a composition. The provider responds to the client's request (requested QoS) by establishing a user session based entirely on SCC and SCC+ components.

We obtain a SCC component, self monitoring and self controlling Component. The sub-components of the membrane (monitors and QoS) are activated in order to perform monitoring of the quality of service and to notify its degradation.

B. Extended SOA properties

We based on the recommended service SOA with the properties of description, invocation, autonomy, reuse, and loose coupling. In [1], we have added the following properties: stateless, mutualization, ubiquity, and exposability. These properties, named SOA+, allow exposing components in a library (catalogue), sharing components for use in different applications, and assembling them in a personalised session.

In this article we focus on the properties that the architect/developer must particularly take into account:

- Autonomy which will be presented in Section III-C.
- Reusability: A service has an agnostic logic and thanks to this can be positioned as a reusable resource.
- Composability: A service has to be designed so that they can be used in a service composition. This property is used via system information blocks (SIB) in Intelligent Network of Telecommunication services ([41]).

C. Autonomic capabilites

GCM/ProActive [42] is the component platform we used for out experimentation. What motivates this choice is the design of the component model imposing a strong encapsulation between components. In GCM/ProActive each component is seen as an autonomous entity in a much service-oriented manner. GCM/proactive enforces a strong separation of concerns, well separating the component management from the functional behaviour of the components [42]. It also revealed efficient for implementing autonomic services.

In the GCM component model, a structure is defined for the membrane elements: the non-functional part of the component can thus be defined as an assembly of components. These components can then be connected with other components within the same membrane or with non-functional interfaces of other components. This structure has been precisely and formally specified [43], [44].

IV. Towards an efficient driving

In cloud computing, services platforms, and Internet of Things (IoT), the component is the cornerstone. Each component is responsible for its action. It can belong to several

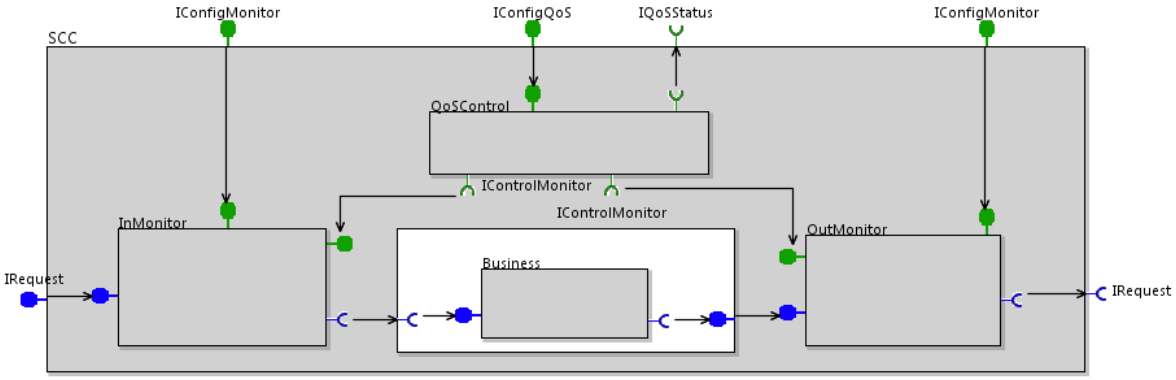

Fig. 1. Self-Controlled service Component (SCC)

providers. It's chosen according to his contract. Each application (service composition) responds to a client request based on the resources and possibilities of its environment. So, the questions are: How to help the service provider to calibrate their service? How to help the application architect designer?

In the next sections, we will present the advantages of our monitoring service and our SCC architecture (Section IV-A) and we will show how a calibration technique based on the SCC (Section IV-C) can help the service provider to create his catalogue (Section IV-D) and the design architect to compose his application (Section IV-B and IV-E).

A. The advantages of MaaS within the SCC architecture

As presented on the section III-A, a SCC component includes a monitoring and an analysis for each functional component. Furthermore, the monitors and the QoScontrol surround it and are close to it. The SCC architecture covers most of the the properties related to monitoring systems defined in section II. Indeed it has many advantages:

1) Our SCC component allows self-control and automatic reacting (*Autonomicity*).
2) The code is simplified and hence require less computing resources (*Adaptability*).
3) The analysis is faster, more relevant, and reaction times is minimized (*Timeliness*).
4) Monitoring and controlling component are:
   - Generic so they are independent of the functional component (*Comprehensiveness*) and may be present at all levels of architecture.
   - Not intrusive because they are external to the functional component (*Intrusiveness*).
5) At each addition/removal of a functional component, a monitoring and controlling component is therefore added/removed (*Scalability, Elasticity*).
6) The volume of data exchanged and thus the communication resources are extremely low since the analysis would be done on site. Only its result would be sent.
7) We measure a QoS (Section III-A) whereas most existing tools monitor network traffic or CPU usage when they should monitor the functional component performance.

In the following section we propose a method compatible with the objectives of self-control, i.e., dynamic reaction as well as the management of service composition.

B. System design: Method for design architect

By principle, component oriented programming requires the programmer to think about the re-use and sharing properties of software components at the time of their creation. Here, we push this methodology further and require the application provider to also consider QoS and monitoring purposes at design time which modifies the company's organizational process. So, we propose a new method based on SCC components to help the design architect when choosing the best component and designing his application.

This method has four steps:

1) We begin with the calibration of SCC components. The technique, using self-tests, described in Section IV-C, consist, for a SCC component, to evaluate his nominal/offered QoS and threshold value under resources conditions of the underlying level.
2) Secondly, the service provider creates his catalogue by putting into the preceding calibrated SCC components (Section IV-D).
3) Thirdly, to design application or service, the architect chooses multi-tenant SCC components in providerss catalogues, based on the specified nominal/offered QoS and thresholds value. He calibrate the composition (SCC+) with the same technique described in Section IV-C to also obtain the nominal QoS and threshold value of the full composition. If the composition is entirely SCC composed then it can be put in a catalogue too.
4) Finally, in Section IV-E, we propose SLA management actions to ensure the adequacy of his nominal QoS to the requested QoS (SLO).

C. Monitoring as-a-service for calibration

As mentioned in Section IV-B, the calibration concerns a single SCC that is intended to be placed in the catalogue's

provider or a composition of SCC components. The calibration consists to compute their offered/nominal QoS and their associated threshold value.

First, we focus on the calibration of a single SCC. Our SCC component includes 4 membrane localised non-functional subcomponents: InMonitor, OutMonitor, QosControl, and Self-test (Figure 2). The two monitors surround the business.

The offered/nominal QoS is obtained by a self-test procedure triggered by the QoSControl. The self-test procedure is performed in a closed loop (in-situ). The InMonitor no longer receive requests from outside during the self-test phase but N intelligently chosen queries are generated to compute the values of QoS. We are no longer dependent on the outside for obtaining the QoS and the measurements obtained are more reliable.

In a normal utilisation, each external request is intercepted by the inMonitor which record it whereas in self-test situation, each request is generated by the Self-test component which is recorded too by the InMonitor. The request is then processed by the business code and the result is intercepted by the OutMonitor which record it.

We measure a QoS (Section III-A) from raw data. The number of incoming or outgoing requests (queries, packets, primitive) and their timestamp are recorded by the two monitors. The QosControl periodically ask the monitor for their records. It can detect if a request has been processed by the business component or not and can compute the number of processed/unprocessed requests and the processing time by subtracting the out and in timestamp. The QosControl can compute other metrics like the availably of the component or the number of processed request by minutes. In a normal situation, it check the compliance with the SLA by comparing the result with a reference threshold and send an in or out contract. In the self-test procedure, it's used to compute the offered/nominal QoS and the threshold values from which the business component stops responding by gradually increasing the numbers of requests. The obtained QoS are given on resources conditions because they depend on their environment. Reference tests can also be processed by modifying the resources to highlight the effect of the environment on the measures.

Second, we focus on the calibration of a composition (SCC+). The same procedure can be used for a composition of SCC components. A composition includes two surroundings monitors and QosControl, the self-test procedure computes the nominal Qos and the threshold value of the composition with the same method as for a single SCC. We determine the threshold value from which the business component stops responding by also gradually increasing the numbers of requests.

D. Cataloque

The catalogue is a showcase for reusable components. But as we mentioned, for re-using, it is better to know the offered QoS and the needed resources to provide this QoS. Indeed, for a same functionality, different algorithms and treatments may be used and therefore different QoS are provided. The consumed resources are not the same.

That is why, the provider's catalogue is filled with SCC calibrated components. If a composition is entirely SCC composed then it can be put in a catalogue too. Each component is given with his offered/nominal Qos and the associated resources conditions. Each component, located at the layer N, depends on the QoS of the layer N-1. A component located at the lower layer depends on hardware resources (CPU, RAM).

E. Monitoring as-a-service for design

Based on SLA, with SCC reusable components selected from the catalogue, the architect and/or developer build the desired application by composing services. In a normal utilisation, each external request (user transaction) is intercepted by the inMonitor/outMonitor of highest level which record it. The QosControl check the compliance with the SLA by comparing the result with a reference threshold and send an in or out contract. But between the behaviour of the composition from end to end (application) and that of each SCC component, there are several subsets which are the responsibility of the architect.

The recommended method, as the decision process progresses, is to progressively build SCC composites with a new membrane containing the InMonitor, OutMonitor, and QoSControl (Figure 5). So the MaS will be the cornerstone of the design of the application structure. The analysis of the composite is complex and is still an open issue, however, some cases are simplified. Namely, if the OutContract come from a primitive component, it will be replaced automatically [45], [46], [47]. If we have only InContracts from primitive components and one OutContract from the final composition, then the composition is faulty. In the literature several solutions have been proposed to perform autonomic adaptation and take the adequate adaptation decision, in particular concerning specific component patterns [48], [49], [50].

We show that we know to locate the problem accurately and timely (Timeliness properties) and to send the notification, generating decision making, to the right place.

V. Implementation

In this section, we bring our SCC component on the Proactive Platform. GCM/ProActive is a Java library that includes a component model and has a strong support for large-scale distributed execution of programs. It relies on an active-object pattern for the interaction between the different entities (i.e. components). According to what has been described in the method (Section II), we present the experimentation of two calibrations: Firstly for an single SCC (Section V-A) and secondly for a SCC composition (Section V-B).

A. From design to configuration (experiments for calibration)

As a reminder, the offered/nominal QoS is obtained by a self-test procedure triggered by the QoSControl (Figure 2). The self-test procedure is performed in a closed loop (in-situ). The InMonitor no longer receive requests from outside

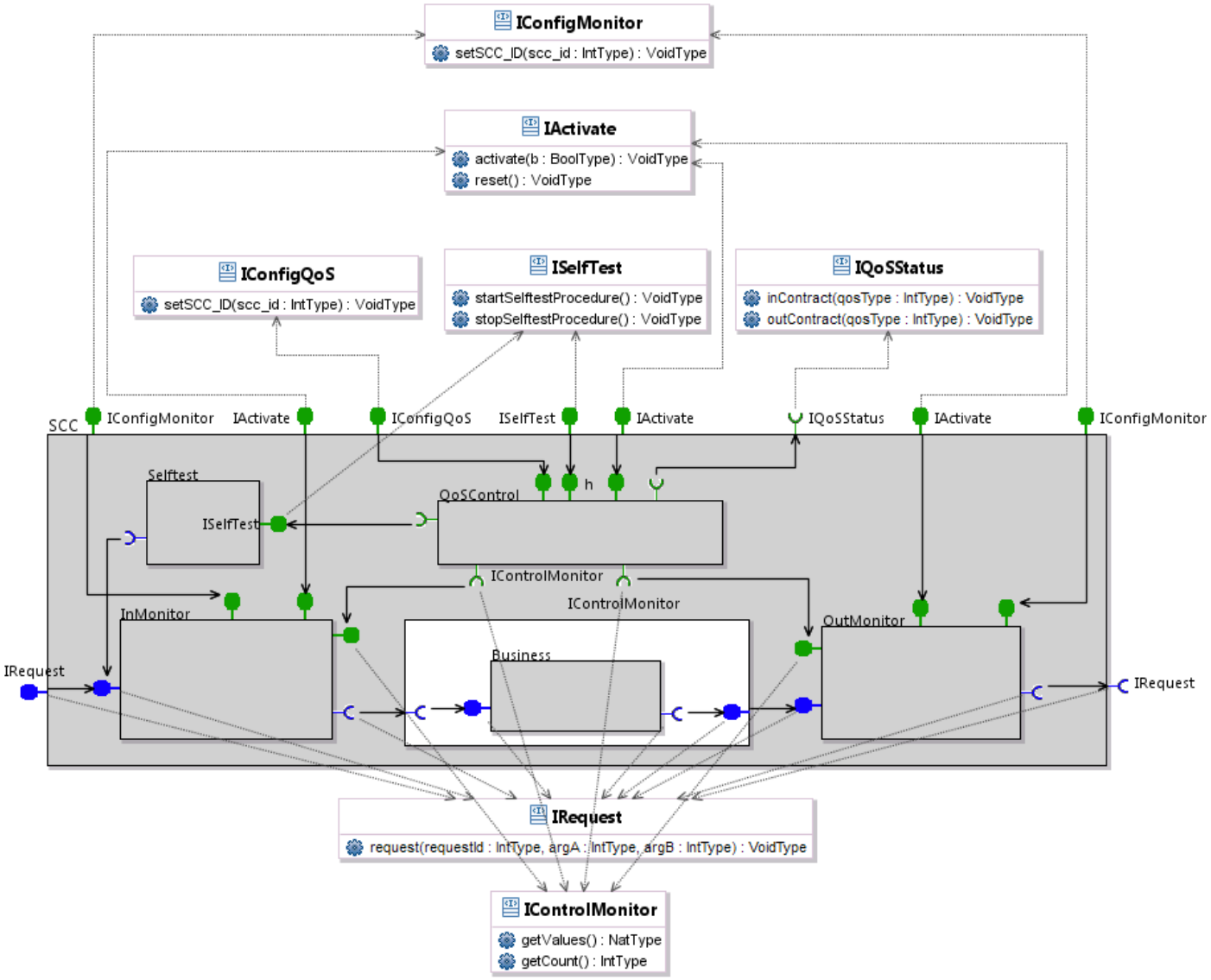

Fig. 2. Self-test SCC component.

during the self-test phase but N intelligently chosen queries are generated to compute the values of QoS.

We bring our SCC component on the Proactive Platform [51].

Each implementation includes five steps:

- Diagram design on VCE with classes and interfaces
- Checking of the validity of the diagram
- Generation of the architecture description language (ADL) file and code template of classes and interfaces.
- Creation of an proactive project with enriched code
- Execution of the application

First, we focus on calibration of only one SCC (Figure 2).

For simplify this implementation, the business role consist only to process a very simple task. In self-test situation, each request is generated by the Self-test component which is recorded by the InMonitor. The record consists of the request number and a timestamp. The request is then processed by the business code and the result is intercepted by the OutMonitor which record it. The request is based on a generic Interface: IRequest which includes his number (requestId) and a list of functional parameters for the business code. The InMonitor, OutMonitor, and QosControl can be (un)activated via the IActivate interface (activate(b: boolType) method). Their records can be erased via the reset() method. The self-test procedure is triggered/stopped by the call of the startSelfProcedure()/stopSelfProcedure() method of the ISelf-test interface. Each non-functional subcomponents receive the number of his community (setSCC_ID() method) via their IConfigQoS interface.

The QosControl is a thread which periodically ask the monitor for their records. It can detect if a request has been processed by the business component and can compute the

```
Controller 38e8d41 compute delay
Controller 38e8d41  31 ms for request 0
Controller 38e8d41   1 ms for request 1
Controller 38e8d41   1 ms for request 2
Controller 38e8d41  18 ms for request 3
Controller 38e8d41 335 ms for request 4
```

Fig. 3. QosControl logs.

processing time by subtracting the out and in timestamp. In the self-test procedure, QoSControl computes the offered/nominal QoS and the threshold values from which the business component stops responding.

Then, we explain the logs of the self-test experiment. As already mentioned, we consider firstly a single SCC component. The business role consist only to process a very simple task. The logs highlights the following events in order:

- The thread of the QosControl starts.
- The Self-test component send 5 requests to the business component via the InMonitor.
- The business receives 5 requests having the identifier 0,1,2,3, and 4 to process.
- The OutMonitor see 5 results coming from the business component.
- The QosControl ask the two monitors their tables of recordings.
- It hears that 5 requests has been recorded. They have successfully been processed by the business.
- The QosControl get the records from the InMonitor and OutMonitor (timestamp and requestId).
- The QosControl compute the processing time (Figure 3).

Note that ProActive supports multi-active objects that allow a single active object to execute several requests in parallel if they do not conflict. This is particularly useful here to ensure that the main flow of requests is handled efficiently while not conflicting with the rest of the behaviour of the monitor.

By repeating the operation and by increasing at each time the number of requests, we can compute the average processing time for a given physical resources level (Figure 4 Single SCC)

Sample:

- Number of request: 220
- Total processing time: 26814 ms
- Average processing time per request: 121.8 ms

Given physical resources:

- Memory: 681616 bytes
- CPU: Intel Core i5 3.6 GHz

By still increasing the number of requests we determine the threshold value from which the business component stops responding. Here for 250 requests. The service provider choose a nominal value which may be defined, for example, at 70 % of the threshold value: 71.2 ms for 175 requests. This experiment show how to use the self-test procedure to compute the offered QoS and the threshold value from which the business component stops responding.

B. Towards the desired architecture (experiments for control)

Second, we focus on the calibration of a composition (SCC+).

The same procedure can be used for a composition of SCC components. An example of composition is given at the figure 5. Two chained SCC components are included in an SCC component called ”Composition”. This composition includes 6 monitors and 3 QoSControl. Thanks to the two surroundings monitors and QosControl, the self-test procedure computes the nominal Qos and the threshold value of the composition (Figure 5 - Composition).

We determine the threshold value from which the business component stops responding. Here for 250 requests. The service provider choose a nominal value which may be defined, for example, at 70 % of the threshold value: 146.6 ms for 175 requests. This experiment show that the self-test procedure is useful too to compute the nominal QoS and the threshold value for a SCC component composition.

VI. Conclusion

In this article, we stood from the point of view of an architect or developer in the new ecosystems that refer to paradigms of Cloud, SOA or IoT and are based on the properties of the ”service”. We started from the reuse property by advocating the SCC component. Thus, during the design of an application or a composite service, the architect and/or the developer will be able to select from a catalogue, for example that of the cloud supplier, the desired(s) service(s) according to the exposed features and associated QoS. The SCC component integrates, during operation, the control of the contract compliance. Furthermore, the most significant proposed help is the use of MaaS to drive more efficiently the design process. Our MaaS as specified allows designers to: (i) assess the offered QoS during the service creation, (ii) test the offered QoS through the catalogue in its deployment environment, (iii) structure, in terms of decisional process, the composite services by placing MaaS at the crucial points of the architecture of the application and allow the control of the SLA contract. All these elements are integrated within a new method for the design architect.

VII. List of abbreviations

- ADL: Architecture description language (Section V-A)
- GCM: Grid component model (Section III-C)
- IoT: Internet of Things (Section IV)
- MaaS: Monitoring as-a-service (Section I)
- MAPE: Monitor-analyse-planning-execute (Section I)
- SCC: Self Controlled service Component (Section I)
- SIB: System information blocks (Section III-B)
- SLA: Service level agreement (Section I)
- SLO: Service level objective (Section II)
- SOA: Service oriented architecture (Section I)
- VM: Virtual machines (Section II)

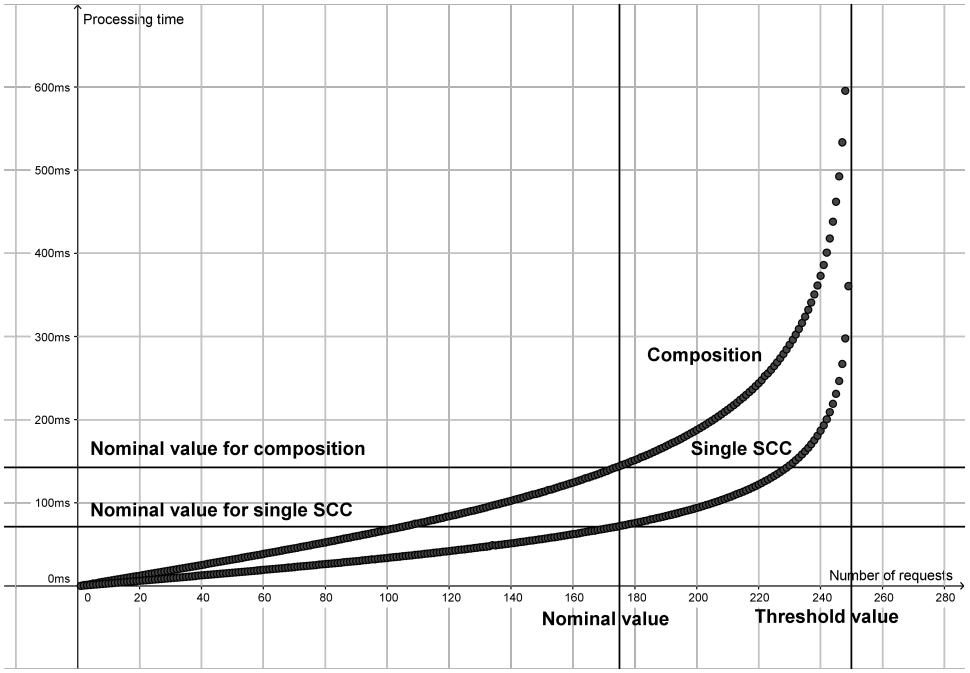

Fig. 4. Number of requests and processing time for the composition.

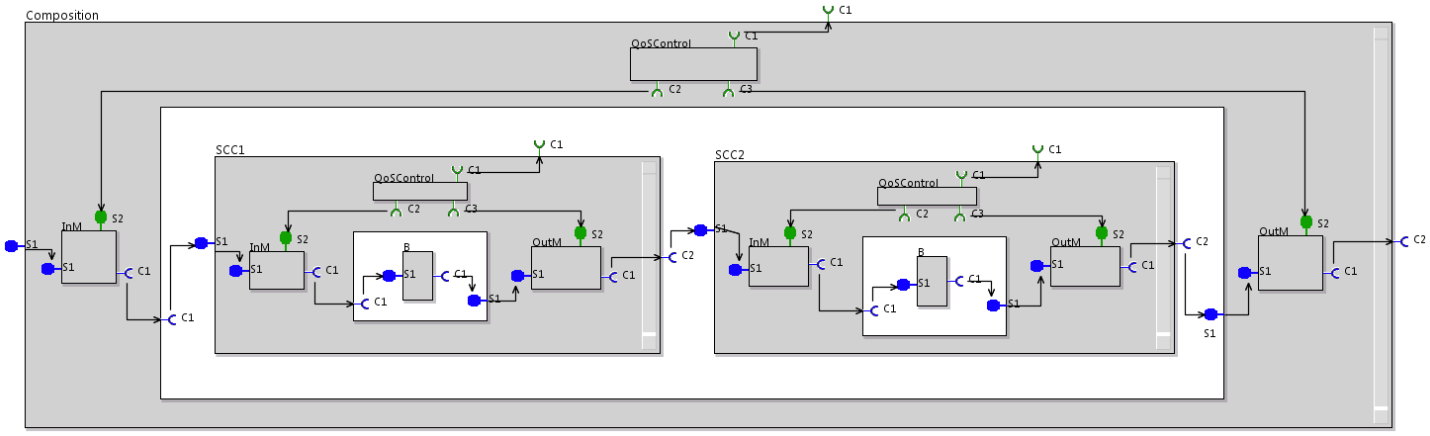

Fig. 5. Example of composition (SCC+).

ACKNOWLEDGMENT

The authors would like to thank for their help and contribution:

- Oleksandra Kulankhina: INRIA, Sophia-Antipolis, France
- Cristian Ruz: Pontificia Universidad Católica de Chile, Santiago de Chili, Chile

This work is supported by the OpenCloudware project. OpenCloudware is funded by the French FSN (Fond national pour la Société Numérique) and is supported by Pôles Minalogic, Systematic, and SCS.

REFERENCES

[1] T. Aubonnet, L. Henrio, S. Kessal, O. Kulankhina, F. Lemoine, E. Madelaine, C. Ruz, and N. Simoni, "Management of service composition based on self-controlled components," *Journal of Internet Services and Applications*, vol. 6, no. 15, p. 17, 2015. [Online]. Available: https://hal.inria.fr/hal-01180627

[2] V. Kumar, Z. Cai, B. F. Cooper, G. Eisenhauer, K. Schwan, M. Mansour, B. Seshasayee, and P. Widener, "Implementing diverse messaging models with self-managing properties using iflow," in *Autonomic Computing, 2006. ICAC'06. IEEE International Conference on*. IEEE, 2006, pp. 243–252.

[3] G. Brunette and R. Mogull, "Security Guidance for critical areas of focus in Cloud Computing V2.1," *CSA (Cloud Security Alliance), USA. Online: http://www.cloudsecurityalliance.org/guidance/csaguide.v2*, vol. 1, 2009.

[4] J. Spring, "Monitoring cloud computing by layer, part 1," *Security Privacy, IEEE*, vol. 9, no. 2, pp. 66–68, March 2011.

[5] ——, "Monitoring cloud computing by layer, part 2," *Security Privacy, IEEE*, vol. 9, no. 3, pp. 52–55, May 2011.

[6] C. Wang, K. Schwan, V. Talwar, G. Eisenhauer, L. Hu, and M. Wolf, "A flexible architecture integrating monitoring and analytics for managing large-scale data centers," in *Proceedings of the 8th ACM*

International Conference on Autonomic Computing, ser. ICAC '11. New York, NY, USA: ACM, 2011, pp. 141–150. [Online]. Available: http://doi.acm.org/10.1145/1998582.1998605

[7] R. Mian, P. Martin, and J. L. Vazquez-Poletti, "Provisioning data analytic workloads in a cloud," *Future Gener. Comput. Syst.*, vol. 29, no. 6, pp. 1452–1458, Aug. 2013. [Online]. Available: http://dx.doi.org/10.1016/j.future.2012.01.008

[8] S. Clayman, A. Galis, and L. Mamatas, "Monitoring virtual networks with lattice," in *Network Operations and Management Symposium Workshops (NOMS Wksps), 2010 IEEE/IFIP*, April 2010, pp. 239–246.

[9] G. Katsaros, R. Kübert, and G. Gallizo, "Building a service-oriented monitoring framework with REST and nagios," in *IEEE International Conference on Services Computing, SCC 2011, Washington, DC, USA, 4-9 July, 2011*, 2011, pp. 426–431. [Online]. Available: http://dx.doi.org/10.1109/SCC.2011.53

[10] P. Hasselmeyer and N. d'Heureuse, "Towards holistic multi-tenant monitoring for virtual data centers," in *Network Operations and Management Symposium Workshops (NOMS Wksps), 2010 IEEE/IFIP*, April 2010, pp. 350–356.

[11] R. Shirey, "Internet Security Glossary, Version 2," RFC 4949 (Informational), Internet Engineering Task Force, August 2007. [Online]. Available: http://www.ietf.org/rfc/rfc4949.txt

[12] "Azurewatch." [Online]. Available: http://www.paraleap.com/azurewatch

[13] "Boundary." [Online]. Available: http://www.bmc.com/truesightpulse/

[14] "Cloudclimate." [Online]. Available: http://www.cloudclimate.com

[15] "Cloudfloor." [Online]. Available: http://cloudfloor.com/

[16] "Cloudcruiser." [Online]. Available: http://cloudcruiser.com/

[17] "Cloudharmony." [Online]. Available: http://cloudharmony.com/

[18] "Cloudsleuth." [Online]. Available: http://www.dynatrace.com

[19] "Cloudstack." [Online]. Available: https://cloudstack.apache.org/

[20] "Cloudwatch." [Online]. Available: https://aws.amazon.com

[21] "Cloudyn." [Online]. Available: http://www.cloudyn.com/

[22] A. Corradi, L. Foschini, J. Povedano-Molina, and J. M. López-Soler, "Dds-enabled cloud management support for fast task offloading," in *2012 IEEE Symposium on Computers and Communications, ISCC 2012, Cappadocia, Turkey, July 1-4, 2012*, 2012, pp. 67–74. [Online]. Available: http://dx.doi.org/10.1109/ISCC.2012.6249270

[23] "Newrelic." [Online]. Available: http://newrelic.com

[24] "Uptimesoftware." [Online]. Available: http://www.uptimesoftware.com

[25] "Vrealize hyperic." [Online]. Available: http://www.vmware.com

[26] D. Zissis and D. Lekkas, "Addressing cloud computing security issues," *Future Gener. Comput. Syst.*, vol. 28, no. 3, pp. 583–592, Mar. 2012. [Online]. Available: http://dx.doi.org/10.1016/j.future.2010.12.006

[27] J. Park, H. Yu, K. Chung, and E. Lee, *Markov chain based monitoring service for fault tolerance in mobile cloud computing*, 5 2011, pp. 520–525.

[28] S. Clayman, R. Clegg, L. Mamatas, G. Pavlou, and A. Galis, "Monitoring, aggregation and filtering for efficient management of virtual networks," in *Proceedings of the 7th International Conference on Network and Services Management*, ser. CNSM '11. Laxenburg, Austria, Austria: International Federation for Information Processing, 2011, pp. 234–240. [Online]. Available: http://dl.acm.org/citation.cfm?id=2147671.2147708

[29] M. Kutare, G. Eisenhauer, C. Wang, K. Schwan, V. Talwar, and M. Wolf, "Monalytics: Online monitoring and analytics for managing large scale data centers," in *Proceedings of the 7th International Conference on Autonomic Computing*, ser. ICAC '10. New York, NY, USA: ACM, 2010, pp. 141–150. [Online]. Available: http://doi.acm.org/10.1145/1809049.1809073

[30] W. Iqbal, M. N. Dailey, D. Carrera, and P. Janecek, "Adaptive resource provisioning for read intensive multi-tier applications in the cloud," *Future Gener. Comput. Syst.*, vol. 27, no. 6, pp. 871–879, Jun. 2011. [Online]. Available: http://dx.doi.org/10.1016/j.future.2010.10.016

[31] A. Ayad and U. Dippel, "Agent-based monitoring of virtual machines," in *Information Technology (ITSim), 2010 International Symposium in*, vol. 1, June 2010, pp. 1–6.

[32] V. C. Emeakaroha, M. A. S. Netto, R. N. Calheiros, I. Brandic, R. Buyya, and C. A. F. De Rose, "Towards autonomic detection of sla violations in cloud infrastructures," *Future Gener. Comput. Syst.*, vol. 28, no. 7, pp. 1017–1029, Jul. 2012. [Online]. Available: http://dx.doi.org/10.1016/j.future.2011.08.018

[33] M. L. Massie, B. N. Chun, and D. E. Culler, "The ganglia distributed monitoring system: design, implementation, and experience," *Parallel Computing*, vol. 30, no. 7, pp. 817 – 840, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167819104000535

[34] "Nagios." [Online]. Available: http://WWW.nagios.org

[35] M. de Carvalho and L. Granville, "Incorporating virtualization awareness in service monitoring systems," in *Integrated Network Management (IM), 2011 IFIP/IEEE International Symposium on*, May 2011, pp. 297–304.

[36] G. Xiang, H. Jin, D. Zou, X. Zhang, S. Wen, and F. Zhao, "Vmdriver: A driver-based monitoring mechanism for virtualization," in *Reliable Distributed Systems, 2010 29th IEEE Symposium on*, Oct 2010, pp. 72–81.

[37] D. T. Hoang, C. Lee, D. Niyato, and P. Wang, "A survey of mobile cloud computing: architecture, applications, and approaches." *Wireless Communications and Mobile Computing*, vol. 13, no. 18, pp. 1587–1611, 2013. [Online]. Available: http://dblp.uni-trier.de/db/journals/wicomm/wicomm13.html

[38] S. Padhy, D. Kreutz, A. Casimiro, and M. Pasin, "Trustworthy and resilient monitoring system for cloud infrastructures," in *Proceedings of the Workshop on Posters and Demos Track*, ser. PDT '11. New York, NY, USA: ACM, 2011, pp. 3:1–3:2. [Online]. Available: http://doi.acm.org/10.1145/2088960.2088963

[39] F. Azmandian, M. Moffie, J. G. Dy, J. A. Aslam, and D. R. Kaeli, "Workload characterization at the virtualization layer." in *MASCOTS*. IEEE Computer Society, 2011, pp. 63–72. [Online]. Available: http://dblp.uni-trier.de/db/conf/mascots/mascots2011.html

[40] T. Aubonnet and N. Simoni, "Self-control cloud services," in *2014 IEEE 13th International Symposium on Network Computing and Applications, NCA 2014, Cambridge, MA, USA, 21-23 August, 2014*, 2014, pp. 282–286.

[41] "Service plane for intelligent network, capability set2," ITU-T Recommendation Q.1222, Tech. Rep., 1997.

[42] F. Baude, L. Henrio, and C. Ruz, "Programming distributed and adaptable autonomous componentsthe gcm/proactive framework," *Software: Practice and Experience*, p. n/a, 2014. [Online]. Available: http://dx.doi.org/10.1002/spe.2270

[43] F. Baude, L. Henrio, and P. Naoumenko, "Structural reconfiguration: An autonomic strategy for gcm components," in *Proceedings of the 2009 Fifth International Conference on Autonomic and Autonomous Systems*. Washington, DC, USA: IEEE Computer Society, 2009, pp. 123–128.

[44] L. Henrio, O. Kulankhina, D. Liu, and E. Madelaine, "Verifying the correct composition of distributed components: Formalisation and Tool," in *FOCLASA*, Rome, Italy, Sep. 2014. [Online]. Available: https://hal.inria.fr/hal-01055370

[45] Tatiana Aubonnet and Noëmie Simoni, "Service creation and self-management mechanisms for mobile cloud computing," in *Wired/Wireless Internet Communication - 11th International Conference, WWIC 2013, St. Petersburg, Russia. Proceedings*, 2013, pp. 43–55.

[46] Noëmie Simoni and Xiaofei Xiong and Chunyang Yin, "Virtual community for the dynamic management of NGN mobility," in *Fifth International Conference on Autonomic and Autonomous Systems, ICAS 2009, Valencia, Spain, 20-25 April 2009*, 2009, pp. 82–87. [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/ICAS.2009.33

[47] Houda Alaoui Soulimani and Philippe Coude and Noëmie Simoni, "User-centric and qos-based service session," in *2011 IEEE Asia-Pacific Services Computing Conference, APSCC 2011, Jeju, Korea (South), December 12-15, 2011*, 2011, pp. 267–274. [Online]. Available: http://dx.doi.org/10.1109/APSCC.2011.64

[48] M. Aldinucci, M. Danelutto, and P. Kilpatrick, "Autonomic management of non-functional concerns in distributed and parallel application programming," in *Proc. of Intl. Parallel and Distributed Processing Symposium (IPDPS)*. IEEE, 2009.

[49] Alexandre Augusto Flores, Rafael de Souza Mendes, Gabriel Beims Brascher, Carlos Becker and Maria Elena Villareal, "Decision-theoretic model to support autonomic cloud computing," in *The Fourteenth International Conference on Networks,ICN 2015, April 19 - 24, Barcelona*, 2015, pp. 219–223.

[50] Rafael Mendes, Rafael Weingartner, Guilherme Geronimo, Gabriel Brascher, Alexandre Flores, Carlos Westphall and Carla Westphall, "Decision-theoretic planning for cloud computing," in *2015 The Fourteenth International Conference on Networks,ICN, February 23 - 27, Nice, France*, 2014, pp. 191–197.

[51] "ProActive Parallel Suite." [Online]. Available: http://proactive.inria.fr/