Research Opportunities in an Intercloud Environment Using MOSt in SLA4CLOUD Project

Emanuel Coutinho, Marcelo F. Santos, Stenio Fernandes, José Neuman de Souza, Thiago Moreira da Costa, Elie Rachkidi, Nazim Agoulmine, Javier Baliosian

Research Opportunities in an Intercloud Environment Using MOSt in SLA4CLOUD Project

Emanuel Coutinho*, Marcelo Santos†, Stenio Fernandes†, José Neuman de Souza*,
Thiago Moreira da Costa¶, Elie Rachkidi‡, Nazim Agoulmine‡ and Javier Baliosian§
*Federal University of Ceará (UFC), Fortaleza - Brazil
†Federal University of Pernambuco (UFPE), Recife - Brazil
‡IBISC/LRSM Lab, University of Evry Val d'Essonne, France
¶IMAG/LIG, Université Grenoble Alpes, France
§Universidad de la Republica, Montevideo - Uruguay
Email: emanuel@virtual.ufc.br, mabs@cin.ufpe.br, sflf@cin.ufpe.br, neuman@ufc.br,
thiago.moreira-da-costa@imag.fr, elie.rachkidy@ibisc.fr, nazim.agoulmine@ibisc.univ-evry.fr, javierba@fing.edu.uy

***Abstract*—Actually, Internet services are becoming essential for different types of users. This evolution impacts how data connections, network routes and resources are configured and used. In this context, the way in which distributed applications and services is becoming more difficult to manage. Cloud computing allows interactions between cloud providers and cloud service providers, and cloud providers can offer deployment services in different datacenters located in different world regions. Much development effort is needed for deploying scalable solutions. One of the these challenges is how to design, develop and deploy cloud solutions that could meet the policies and security requirements of multiple environments needs. The SLA4CLOUD project intends to build an environment where a user can request the deployment of its services anywhere in the underlying infrastructure, using the MOSt platform and its services. This work aims to report some opportunities and research challenges resulting from SLA4CLOUD project in the context of MOSt platform, and the promotion of new projects and partnerships.**

***Index Terms*—Cloud Computing; IaaS; Testbed; Service Level Agreements.**

I. Introduction

Currently, the recent trend of building massive datacenters is becoming part of the current network evolution. This evolution impacts not only the users, but also in how data connections and network routes are configured. In this context, the manner in which distributed applications and services collaborate with each other is also becoming more difficult to manage.

In order to make this cloud-based service ecosystem to work, standardized protocols, suitable data representation formats, well-defined interfaces, and efficient data distribution and dissemination techniques are needed. Indeed, some solutions have already been deployed, such as those used by Google and Amazon in their service platforms.

Cloud computing allows interactions between cloud providers and cloud service providers [1]. Many cloud service providers offer to their customers the possibility to deploy services in different datacenters located in different world regions. Generally, these datacenters are different in hardware capacity and support different variants of cloud services. And either, these datacenters have different strategies for resources management, scheduling and provisioning.

Although several cloud services have emerged, the scenery for research is still promising. Much development effort is still needed for deploying scalable solutions, especially across the Internet, and for mobile devices. One of the most significant challenges is how to design, develop, and deploy cloud-enabled solutions that could meet the policies and security requirements of multiple environments needs.

In this context, the SLA4CLOUD project intends to build an environment where a customer could request the deployment of its services anywhere in the underlying infrastructure, given a set of defined constraints. Therefore, services could be modeled as service graphs that could be deployed in different datacenters located in France, Brazil and Uruguay, depending on the SLA objectives such as localization objectives, QoS objectives or pricing [2].

This article aims to disseminate some opportunities and research challenges resulting from SLA4CLOUD project, which benefited from MOSt platform. The rest of the article is divided into the following sections: Section 2 describes a little of MOSt platform; Section 3 comments about SLA4CLOUD project; Section 4 points out some research opportunities; Section 5 outlines some research challenges that can be performed from the results of the project; and finally, in Section 6, we presented some conclusions.

II. MOSt

Multisite Orchestration System (MOSt) is a component developed under the ITEA EU EASI-CLOUDS project [3]. Its aim is to provide an optimal provisioning plan in a distributed cloud infrastructure of a service request in a service graph way. The service graph represents a customer request in terms of a set of basic services, such as virtual machines and storage, and the links between these services. These links could be required network links between the nodes or available bandwidth [2].

Underlying nodes in the network graph are sites located in different geographical locations. For instance, the sites showed in Figure 3. Each site is an IaaS operated by a site manager which has its own service portfolio. It also may have its service pricing, depending on its deployed features.

MOSt is responsible for the global provisioning only. This means it will decide which service component will be initiated where [2]. Local provisioning depends on the local implementation, and therefore sites may use different scheduling mechanisms to deploy service component on each country. Hence, the MOSt core system is responsible for building the provisioning plan and requesting the underlying sites to instantiate the resources based on the SLA terms.

The main requirements presented in MOSt design are:

- Multi-site deployment: MOSt must be capable of deploying a complex service over different cloud datacenters (sites) geographically distributed;
- Optimal provisioning plan: MOSt must provide to the customer the best provisioning plan and relies on IGM (Iterative Graph Mapping) algorithm [4];
- Site independence: MOSt should be independent from any underlying site, i.e., the system should be stand-alone and not connected by internal details.

The handling of a customer service request deployment is achieved in three phases:

- Multi-site provisioning phase: MOSt calculates the optimal provisioning plan and engage the resources in the multiple underlying sites;
- Post-configuration phase: MOSt launches the post-configuration of the deployed virtual machines based on the customers request;
- Networking provisioning phase: MOSt launches the network configuration connections between datacenters sites to fulfill nodes communication service requirements. This requires the instantiation of specific network gateways in each site to build the necessary VPL (Virtual Private Links) to allow virtual machines of the same project to communicate.

The MOSt system responsibilities are:

- The global provisioning (it will decide which service component will be initiated;
- Allocating resources to zones/sites;
- Orchestrating other sub-components such as network configuration and post-configuration;
- Parsing requests;
- Initiating deployment of requested applications in different zones;
- Consulting current state of deployed applications;
- For applications with reconfiguration, requesting/deleting more resources in case of scale up/down.

III. SLA4CLOUD Project

A. Project Description

The SLA4CLOUD project has as objective to build an environment where a customer could request the deployment of its services anywhere in the underlying infrastructure. This deployment complies a set of defined constraints. Thus, services could be modeled as service graphs of service components that could be deployed in different datacenters located in France, Brazil and Uruguay, depending on the SLA objectives such as localization objectives, QoS objectives or pricing [2].

As cited in [2], this project proposed: (i) the development of different cloud services with a SLA representation that could be used for negotiation in cloud environments; (ii) the implementation of a strategy for dynamic consolidations of virtual machines aiming to reduce energy consumption without compromising performance requirements concerning availability and SLA violation; (iii) the development of an automated security policy composition mechanism for composite services in Cloud, while maintaining consistency with the security policies of the external services; (iv) a rule-based pricing system that implements intuitive ideas to improve the quality of service and to increase the global income of a Cloud Computing provider; and v) to deploy the developed mechanisms in a Mobile Cloud The computing scenario as case study and proof-of-concept demonstrator. To achieve the above goals, the we used the MOSt platform and some development into the all sites infrastructure was necessary. As cloud, we used OpenStack. Only some of the project objectives were deepened. The others remained only as study topics.

The SLA4CLOUD architecture glimpses large scale distributed applications, supported by services and processes based on distributed and integrated facilities, using an infrastructure among different countries covering Europe and South America. In this context, the infrastructure should be flexible and a programmable platform that allows the adding of new functions and capabilities. In order to establish this infrastructure, we need to install full the three sites, one in Brazil, one in France, and one in Uruguay. As illustrated in Figure 3, the SLA4CLOUD testbed is located in different data centers at IBISC Laboratory, Evry University (France), UFC Fortaleza (Brazil), and UDELAR Montevideu (Uruguay).

B. Environment Description

A website is developed as a front end which uses MOSt services. It allows the Cloud Consumer to define his complex service (i.e. interconnected virtual machines), reserve resources, and deploy in pre-configured Cloud sites. We can see the website main features in Figure 1 menu: (i) parser, (ii) keypairs, (iii) manifests, (iv) nodes, (v) virtual links, (vi) maps, and (vii) user information

Using the parser component, it is possible to add nodes and virtual links to a manifest description in JSON format (Figure 2). Once finished, the Cloud Consumer sends this manifest to MOSt component for checking and storage. The keypairs tab manages all created SSH keys when deploying user-defined services. They can be downloaded from the website and used by the Cloud Consumer for accessing his services (virtual machines).

The manifest tab display all defined manifests and allows to apply actions on them. The Cloud Consumer can perform resources reservation, virtual machines deployment and interconnection, and resources release. Nodes and virtual links tabs allows the management of each manifests' elements.

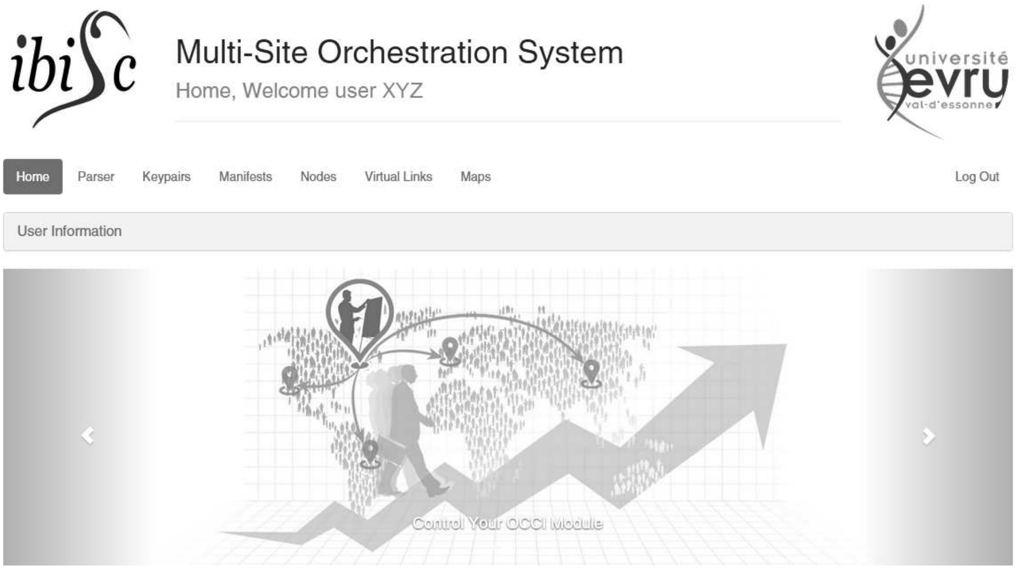

Fig. 1: MOSt initial screen

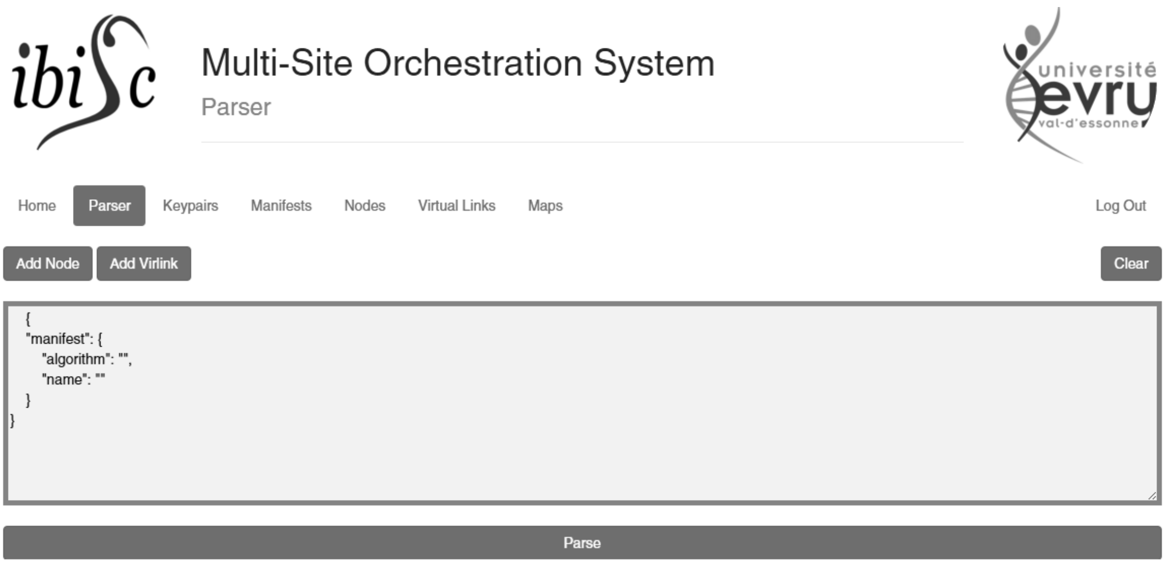

Fig. 2: MOSt parser screen

Finally, the map feature shows a map with all sites and virtual machine (Figure 3 and Figure 4). It allows to visualize Cloud Sites (green circle if the site is active and red circle otherwise), deployed virtual machines (blue circle if the machine is successfully instantiated and orange circle otherwise), and established links between virtual machines (blue lines).

IV. Research Opportunities

We identified some research opportunities during the SLA4CLOUD project. Many of these opportunities are infrastructure aspects. These opportunities may generate new projects and new services to be incorporated into the MOSt platform.

A. *Elasticity*

Elasticity is an aspect of cloud computing widely discussed both in academic and commercial environments. According to the National Institute of Standards and Technology (NIST) definition, resources can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand [5]. To the consumer, the resources available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time. For a detailed study about cloud elasticity, various aspects were described in [6], such as: definitions, metrics, benchmarks, strategies, challenges and trends in the construction of elastic solutions. In a recent work, elasticity was defined as

Fig. 3: MOSt environment showing a functional connection between sites

Fig. 4: MOSt environment showing a problem in the connection between sites

the degree to which a system is able to adapt to changes in workloads by resources provisioning and unprovisioning in an autonomic manner, so that at each point in time the available resources correspond to the workload demand as much as possible [7].

The SLA4CLOUD platform can provide elasticity services for running applications. This strategy can be based on reactive or proactive actions by increasing or decreasing the resources capacity, or replicating resources [6]. All of this depends on the elasticity design to be defined for the platform, providing resources as needed, and therefore avoid resource waste and idleness. Some metrics could be used for elasticity measurement, such as Elasticity of Demand [8] and Elasticity of Resources [9].

B. Load Balance

The platform provides several sites where they can be in communication. If a data center / cloud of a site is overloaded, these requests can be directed or distributed to other sites, and thus maintaining a certain level of service quality and use of resources.

C. Availability

The availability of resources, in most cases virtual machines to be accessed remotely, it is essential for the full functioning of a cloud infrastructure. Since there is a more global management level, it is possible to manage the sites that are active and not active, and thus redirect links and requests, so that running services remain active, and transparently to the user.

D. Monitoring

Monitoring can be performed at several levels: the level of virtual machine, where each site has its instances, their physical features, and their applications; the site level, allowing a view of the total capacity in use and available from the site, and their instances; and the level of connections between sites, allowing the visualization of how the sites and how their instances are communicating. Other monitoring levels may also be defined depending on the need and business rules.

E. Access Control

The access to multiple virtual machines of the whole platform should be controlled and monitored so as not to allow security breaches. Such access should be defined so that each user profile has its proper access credentials. In addition, the applications to be executed on the platform must have a manner to access virtual machines from the other infrastructure sites, in order to use services or simply to benefit from the availability of resources.

V. Research Challenges

Some research challenges can be identified through use of MOSt platform and SLA4CLOUD project. These challenges can be difficulties and problems to be solved to improve access and environmental performance.

A. Cloud SLA Negotiation

An SLA is recognized to be a contract between a Service Provider and a Service Customer. It is designed to create a clear measurable common understanding of the minimal obligations and expectations about what the customer wants, what he is requesting, and what the provider has committed to provide and at which constraints. In this context, the objective of the SLA Negotiation process is to provide a customer an API that allow him to describe the required SLA for a Cloud Service as well as for negotiating the terms of this SLA with the provider in term of QoS level and associated cost. The Cloud Service could itself be composed of other Cloud Service and in this case there should be a way to describe the semantic dependencies between the services.

Currently, there is no real standard for Cloud Computing SLA definition, and the Service Level Objective (SLO) should be identified in a precise manner. Besides, we have to handle with important parameters, such as: performance, availability reliability, privacy, cost, and compensations in case of SLA Violation.

Negotiation process is other very important issue. Traditional negotiation process, involving human representatives from both parties, can not be considered since the number of served consumers could be extremely large and worldwide. Therefore, there is a need to automate completely the process via an automatic SLA Negotiation Process. Negotiation is the process whereby the service provider and consumer discuss and agree upon expectations of the service provision. Service consumers will require specific QoS, for example, consumers with critical operations will desire faster response time, and their expectations may change over time due to continuing changes in the environment.

B. Energy-efficient Resource Management in Cloud

Given the environmental and economic impact caused by high-energy consumption in large data centers, many works in literature have proposed techniques capable of providing energy efficiency in such infrastructure [10]–[13]. As a challenges for energy management, we have: migration of virtual machines to shut down idle hosts; tradeoff analysis between virtual machines migration and reducing the physical servers number to host virtual machines; optimization and heuristic approaches for selecting the best frequency/voltage combination to each host, and to reduce the total energy consumption of the system; balance between overload and idle states of the computing environment; and establishing virtual machines selection policies for minimizing the migration impact on system performance.

In this context, an allocation controller of applications that minimizes energy consumption and cost of migration, with good performance in virtualized clusters was proposed in [10]. This controller uses resizing and migration of virtual machines (VM), at runtime, to consolidate the load in an optimal set of hosts so that the idle hosts are switched to low power states. The dynamic allocation of VMs on hosts can be seen as a problem of bin packing with items of different sizes, which involves a tradeoff between VM migration and reducing the number of physical servers to host VMs [11]. Therefore, the strategies to save energy by changing the CPU frequency and/or voltage must ensure that a task does not use much of the CPU processing power. Moreover, DVFS (Dynamic Voltage and Frequency Scaling) does not apply to other computing resources of a system besides the CPU. An optimization solution for energy management and performance in virtualized clusters was proposed in [12]. The optimization decision consists of selecting the best combination frequency/voltage to each host. Although the proposal ensures optimal solutions, it becomes critical especially in virtualized computing environments in large scale, where there is constant change in resources demand. Therefore, heuristic approaches must be considered because they are able to deliver a solution in short time, even if this is not the optimal solution. A balance between overload and idle states of the computing environment was proposed in [13]. In this work, the redundant hosts are shut down when the system is in non-intensive computing state and the load of the overloaded hosts is transferred to the free hosts when the system is in its intensive computing state.

C. SLA-aware Service Composition in Cloud Environments

One of research challenge in service composition area is to develop automated or semi automated methods for identifying general SLA inconsistencies in the existing policies. Several solutions aim to build applications by composing services by means of a composite service that invokes external services. As an example of a solution, Service-Oriented Architecture (SOA) has the flexibility to adapt to changing business processes and can meet functional requirements, but faces difficulties to satisfy the non-functional requirements, such as security. Consequently, due to the growth of the number of users and emergence of new network scenarios the traditional approach for service composition can not accommodate the variety of user requirements.

As services share a common resource infrastructure, guaranteeing SLA compliance is a challenging task. Hence, SLA elements, such as security issues, must be fully integrated into the composition techniques as one of the objectives. In this context, several challenges still remain open, such as derivation of rules for compositing policies that take into account SLA service and management elements for most types of composite processes [14] [15] [16]. The problem becomes more complexity take into account elasticity in Cloud environments through virtualization mechanisms and components [17] facing inherent security challenges.

A specific challenge is to develop a semi or fully automated security policy composition mechanism for composite services while maintaining consistency with the security policies of the external service [18]. The Cloud Security Alliance (CSA) defines security domains over specific functional aspects and presents a number of security concerns, best practices, and recommendations regarding at all levels, from data privacy to infrastructural configuration, which should be all included in SLA negotiations [19]. As security is a major area of concern for Cloud environments, data confidentiality and audibility have been specifically considered as the major obstacles to its wide-spread adoption. In addition, deploying the whole workflow or business process on a secure private cloud may result in poor performance. The use of hybrid and heterogeneous Clouds has the potential of solving such scalability and performance issues, but such workload partitioning requires strict security of data and services.

In other words, research challenges must provide mechanisms to evaluate the set of services (individually or composed) in a workflow, and find the optimal set of deployments over hybrid clouds, while ensuring SLA requirements are fulfilled.

D. SDN and Cloud Environments

Software Defined Networking (SDN) is a paradigm that promises to provide an extensive control over network traffic flows [20] [21]. In a nutshell, SDN resides in a clear separation of the data (or forwarding) plane from the control plane enabling a centralized programmable control plane. Consequently, the emergence of the SDN paradigm provides a new opportunity to integrate virtualized components in the cloud environment through programmable interfaces that may provide control flexibility, adaptive network policies and a greater degree of automation.

Challenges in this field of research come from the fact that SDN architectural design breaks a traditional paradigm created many years ago on the Internet. New protocols and abstractions are needed to support this transition from traditional networks to an SDN paradigm, requiring an effort from academia and industry. As a result, to adopt an SDN paradigm it needs to make a large number of adaptations in the physical infrastructure in order to create an integrated functional environment and it aggregates a high hardware solution cost (e.g.: acquiring SDN controller and switches). On the other hand, a Cloud Computing environment needs to have a good orchestration engine in order to take advantage of available resources through the use of an SDN network, but even in a centralized architecture is difficult to find an optimum solution in a short period of time for network mapping problems (usually NP-hard problems) [22]. In the very recent past, researchers on SDN have been also putting efforts on very narrow focuses, such as on data and control plane performance, software updates issues in production SDN-based networks, parallel processing of SDN controllers (i.e., at the control plane), and the like. In this context, as a centralized point, an SDN controller can become a bottleneck. As example, the POX SDN controller can handles 30K flows/s [23]. A big datacenter with 2 million virtual Machines (VMs) may generate 20 million flows/s [24]. Consequently, when the design of the SDN control plane comes into play, several research issues become evident. Research challenges for scalable SDN design can be found in several aspects, from data plane optimizations to distributing control plane functionalities. However, as a new paradigm for the future cloud computing and datacenter environments, scalable mechanisms for SDN are expected to leverage potentially transformative business opportunities.

E. Pricing Strategies for Cloud Computing

Cloud computing services have yet to explore several pricing paradigms, but is common to find the ideas of Cloud computing and pay-as-you-go together, where consumers have a loose contractual relationship with the provider. It is natural to extend the idea to a world in which clients choose between two or more Cloud providers and establish short or micro term SLAs with particular QoS specifications and price. However, the price must be chosen so that the revenue of the sellers is maximized while the highest satisfaction is achieved by the buyers.

Typically, there are two main factors influencing the price setting: user demand and competition among service providers. If the services are substitutable, users buy the service that provides the highest satisfaction at the lowest price. Obviously, the Cloud-service providers will want to recover their investment charging the right price for their services. It may not be so simple though. The price of simple goods it is often determined by just one parameter such as the number of copies, their weight or the length of a lease. Cloud services, on the contrary, are specified by several parameters such as number of transactions, transferred data or storage space. Since cloud services can be specified in terms of that many variables, the number of different possible contracts is enormous and complicates the design of a reasonable and coherent pricing strategy. On the other hand, contracts are more than simple price agreements. For example, a contract may be an incentive for the user to use the service conforming to the agreed parameters. This, at time, will impact positively on the service quality and the price paid by the clients in general.

All this motivates an effort to develop a pricing technique simple enough to be implemented by the operators, but, at the same time, sophisticated enough to compete successfully with other strategies and work as a scalable feedback mechanism to control how the cloud is used.

F. Privacy Enforcement in the Cloud

Concerns about privacy have globally raised several discussions involving different sectors of societies and nations. This is due to the complex nature on which privacy is based on, such as temporal, cultural, social, political, and geographical dimensions. For this reason, the concept of privacy is continuously evolving.

The advance in networking, pervasive computing, and data analytics have changed the way individual privacy is threatened today. In addition to that, cloud computing increases its complexity by exposing these privacy violations to different contextual and cultural differences. In response to that, countries and political blocks have revised and updated their regulations and laws to protect political, economical and individual interests, as observed in Europe with the "Right to be forgotten", "Marco Civil" in Brazil, in Russian with the federal laws "On Personal Data" and "On Data Localisation", and "Protección de Datos Personales" in Argentine.

This rapid volatility of the concept of privacy and its influences over regulations justifies the need to develop controls to support efficient and automatic privacy enforcement. In this context, the following challenges can be remarked regarding privacy:

- Jurisdiction: Cloud computing contemplates the storage and processing of data in any and everywhere simultaneously. Multiple jurisdictions can be involved during the service lifetime or only during a specific data analysis process. For this reason, future cloud service providers need to address jurisdiction compliance by considering regulation jurisdiction in the cloud elasticity process. Nowadays, legal offices and knowledge databases manually updated assess today's regulations compliance, supporting the legal verification process. Cloud computing may support this verification automatically by providing formal methods to verify compliance in different jurisdiction regulations and considering different time frame.
- Accountability: Cloud providers must offer data processing and data storage models that safeguards data confidentiality. Security is the most important criteria mentioned by companies' CIOs concerning the portability of their business into the cloud, and it is the base for providing privacy. In order to guarantee business confidentiality, organizational privacy needs to be addressed by solid cloud-based infrastructures driven by the capacity to control security and privacy and trace possible confidentiality violations. This is achieved by tracing evidences of legal requirements, such as those provided by independent assessment, and report of compromised data (in case of malicious attack or information leak). Cloud provider may need to deliver data security and privacy accountability by providing indicators and controllers to be integrated as part of their business governance.

G. Secure Interoperability for Policy-Based Cloud Management

Policies are often used as a means of implementing flexible and adaptive systems for the management of Internet services, distributed systems, and security systems. Among the different domains where policy-based control has been applied, one of the most successful is security and access control.

Security policies are usually implemented as sets of rules expressing permissions, prohibitions, and obligations. When two cloud providers or institutions with cloud deployments cooperate and set their systems to exchange data and their users to access each other's infrastructure, the interoperability between the systems of both entities must be studied and, in particular, the security of such interoperability is a particularly important issue.

Therefore, the security policies of both organizations must be analysed and, potentially, modified by higher level policies to fit the cooperation objectives. These new policies can be viewed as a set of contracts negotiated between the different entities to control their interoperation. Therefore, secure interoperability of cloud services rise the following questions: (i) what is permitted if we apply the security rules of two organizations at the same time?; (ii) are the obligated actions consistent between them and with the prohibitions of both organizations?; and (iii) is the emerging policy in line with their cooperation objectives? Answering these questions requires mathematical modelling of policies, operate on that model to map and translate policies from one domain to another, detect contradictions, create high-level policies, and produce emerging new policies automatically in order to study the security aspects of the interaction between several cloud providers.

VI. Conclusion

This work showed some results from SLA4CLOUD project. The research opportunities listed in this work are related to infrastructure features. The research challenges cited here are desirable characteristics for a more complete and complex environment, where various services would be available for the users.

With this work, we hope to contribute with new resource opportunities that can be performed in MOSt environment, such as: cloud SLA negotiation, energy-efficient resource management in cloud, SLA-aware service composition in cloud, SDN in cloud, pricing Strategies for cloud, privacy enforcement in cloud, secure interoperability for policy-based cloud management.

We know the importance of developing the recognized research opportunities in the context of MOSt platform. As future work, we intend to put in practice the identified research challenges as solutions provided by MOSt platform, and to promote new projects and partnerships.

References

[1] L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "A break in the clouds: Towards a cloud definition," *SIGCOMM Comput.*

Commun. Rev., vol. 39, no. 1, pp. 50–55, Dec. 2008. [Online]. Available: http://doi.acm.org/10.1145/1496091.1496100

[2] E. H. Cherkaoui, E. Rachkidy, M. Santos, P. A. L. Rego, J. Baliosian, J. N. de Souza, and N. Agoulmine, "3rd international workshop on advances in ict infrastructures and services," in *SLA4CLOUD: Measurement and SLA Management of Heterogeneous Cloud Infrastructures Testbeds*, Miami, USA, 2014.

[3] EASI-CLOUDS, "Easi-clouds — extensible architecture and service infrastructure for cloud-aware software," http://easi-clouds.eu/, 2015, online; last access october-2015.

[4] K.-T. Tran, N. Agoulmine, and Y. Iraqi, "Cost-effective complex service mapping in cloud infrastructures," in *Network Operations and Management Symposium (NOMS), 2012 IEEE*, April 2012, pp. 1–8.

[5] P. Mell and T. Grance, "The nist definition of cloud computing," *National Institute of Standards and Technology*, vol. 53, no. 6, p. 50, 2009. [Online]. Available: http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc

[6] E. F. Coutinho, F. R. de Carvalho Sousa, P. A. L. Rego, D. G. Gomes, and J. N. de Souza, "Elasticity in cloud computing: a survey," *annals of telecommunications - annales des télécommunications*, vol. 70, no. 7-8, pp. 289–309, 2015. [Online]. Available: http://dx.doi.org/10.1007/s12243-014-0450-7

[7] N. R. Herbst, S. Kounev, and R. Reussner, "Elasticity in cloud computing: What it is, and what it is not," in *Proceedings of the 10th International Conference on Autonomic Computing(ICAC 2013), San Jose, CA*. USENIX, 2013, pp. 23–27.

[8] E. Coutinho, D. G. Gomes, and J. D. Souza, "On applying microeconomics concepts to cloud elasticity evaluation," in *XIII Workshop em Clouds e Aplicações (WCGA2015)*, may 2015. [Online]. Available: http://sbrc2015.ufes.br/wp-content/uploads/proceedingsWCGA20151.pdf

[9] E. Coutinho, P. Rego, D. G. Gomes, and J. D. Souza, "Uma métrica para avaliação da elasticidade em computação em nuvem baseada em conceitos da física," in *XII Workshop em Clouds e Aplicações (WCGA2014)*, Florianópolis - Brasil, may 2014. [Online]. Available: http://www.sbrc2014.ufsc.br/anais/files/wcga/ST2-2.pdf

[10] A. Verma, P. Ahuja, and A. Neogi, "pmapper: Power and migration cost aware application placement in virtualized systems," in *Proceedings of the 9th ACM/IFIP/USENIX International Conference on Middleware*, ser. Middleware '08. New York, NY, USA: Springer-Verlag New York, Inc., 2008, pp. 243–264. [Online]. Available: http://dl.acm.org/citation.cfm?id=1496950.1496966

[11] T. C. Ferreto, M. A. Netto, R. N. Calheiros, and C. A. D. Rose, "Server consolidation with migration control for virtualized data centers," *Future Generation Computer Systems*, vol. 27, no. 8, pp. 1027 – 1034, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167739X11000677

[12] V. Petrucci, O. Loques, and D. Mossé, "A dynamic optimization model for power and performance management of virtualized clusters," in *Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking*, ser. e-Energy '10. New York, NY, USA: ACM, 2010, pp. 225–233. [Online]. Available: http://doi.acm.org/10.1145/1791314.1791350

[13] X. Liao, L. Hu, and H. Jin, "Energy optimization schemes in cluster withÂ virtual machines," *Cluster Computing*, vol. 13, no. 2, pp. 113–126, 2010. [Online]. Available: http://dx.doi.org/10.1007/s10586-009-0110-2

[14] X. Zhao, L. Shen, X. Peng, and W. Zhao, "Toward sla-constrained service composition: An approach based on a fuzzy linguistic preference model and an evolutionary algorithm," *Information Sciences*, vol. 316, pp. 370–396, 2015.

[15] H. Yin, C. Zhang, B. Zhang, Y. Guo, and T. Liu, "A hybrid multiobjective discrete particle swarm optimization algorithm for a sla-aware service composition problem," *Mathematical Problems in Engineering*, vol. 2014, 2014.

[16] A. Jula, E. Sundararajan, and Z. Othman, "Cloud computing service composition: A systematic literature review," *Expert Systems with Applications*, vol. 41, no. 8, pp. 3809–3824, 2014.

[17] H. He, "Virtual resource provision based on elastic reservation in cloud computing," *International Journal of Networking and Virtual Organisations*, vol. 15, no. 1, pp. 30–47, 2015.

[18] R. Dilmaghani, S. Geyik, K. Grueneberg, J. Lobo, S. Y. Shah, B. K. Szymanski, and P. Zerfos, "Policy-aware service composition in sensor networks," in *Services Computing (SCC), 2012 IEEE Ninth International Conference on*. IEEE, 2012, pp. 186–193.

[19] C. Coles and J. Yeoh, "Cloud adoption practices & priorities survey report," *Cloud Security Alliance*, 2015.

[20] D. Kreutz, F. M. V. Ramos, P. E. Verissimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, "Software-Defined Networking: A Comprehensive Survey," *Proceedings of the IEEE*, vol. 103, no. 1, pp. 14–76, jan 2015. [Online]. Available: http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=6994333

[21] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "Openflow: enabling innovation in campus networks," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 2, pp. 69–74, 2008.

[22] M. Abu Sharkh, M. Jammal, A. Shami, and A. Ouda, "Resource allocation in a network-based cloud computing environment: design challenges," *Communications Magazine, IEEE*, vol. 51, no. 11, pp. 46–52, 2013.

[23] F. Benamrane, R. Benaini *et al.*, "Performances of openflow-based software-defined networks: An overview," *Journal of Networks*, vol. 10, no. 6, pp. 329–337, 2015.

[24] M. Jammal, T. Singh, A. Shami, R. Asal, and Y. Li, "Software defined networking: State of the art and research challenges," *Computer Networks*, vol. 72, pp. 74–98, 2014.