SLA Management Framework to Avoid Violation in Cloud

Walayat Hussain[1(✉)], Farookh Khadeer Hussain[1], and Omar Khadeer Hussain[2]

[1] Decision Support and e-Service Intelligence Lab, School of Software, Centre for Quantum Computation and Intelligent Systems, University of Technology Sydney, Sydney, NSW 2007, Australia
{walayat.hussain, farookh.hussain}@uts.edu.au
[2] School of Business, University of New South Wales, Canberra, Australia
o.hussain@adfa.edu.au

Abstract. Cloud computing is an emerging technology that have a broad scope to offers a wide range of services to revolutionize the existing IT infrastructure. This internet based technology offers a services like – on demand service, shared resources, multitenant architecture, scalability, portability, elasticity and giving an illusion of having an infinite resource by a consumer through virtualization. Because of the elastic nature of a cloud it is very critical of a service provider specially for a small/medium cloud provider to form a viable SLA with a consumer to avoid any service violation. SLA is a key agreement that need to be intelligently form and monitor, and if there is a chance of service violation then a provider should be informed to take necessary remedial action to avoid violation. In this paper we propose our viable SLA management framework that comprise of two time phases – pre-interaction time phase and post-interaction time phase. Our viable SLA framework help a service provider in making a decision of a consumer request, offer the amount of resources to consumer, predict QoS parameters, monitor run time QoS parameters and take an appropriate action to mitigate risks when there is a variation between a predicted and an agreed QoS parameters.

Keywords: Cloud computing · SLA management framework · SLA monitoring · Risk management in cloud

1 Introduction

Cloud computing is a recent technology trend that is attracting the attention of a wide range of businesses and enterprise. A cloud computing is combination of different technologies – grid computing, parallel computing, distributed computing, virtualization and multitenant architecture [1]. Due to its wide range of services small and large enterprises are shifting their businesses on cloud. According to a recent press release by Gartner, Inc [2] that state that the expected growth for a public cloud service market is 16.5 % with a total amount of $204 billion increased from $175 billion in 2015. Just in IaaS the market growth is expected from 31.9 % in 38.4 % in 2016 and a global cloud service market is predicted to grow by 13.6 % in 2016 that will reach to $90.3 billion.

A. Hirose et al. (Eds.): ICONIP 2016, Part III, LNCS 9949, pp. 309–316, 2016.
DOI: 10.1007/978-3-319-46675-0_34

This increase in cloud market raises new challenges for cloud providers. One of the main issue is the formation of a viable SLA and predicting likely violation to alarm a service provider for an early remedial actions.

Service level agreement (SLA) is a key business agreement that bond a consumer and a provider for their commitment and promises for a specified period of a time. To avoid from violation penalties a service provider always need a system which intelligently predict a risk of a likely service violation and generate recommendations to manage those risks. There are number of approaches to monitor SLA violation [3] but the problem with most of the existing approaches is that it start monitoring when a provider and a consumer execute their SLA, however for an optimal SLA management a system need to assure the SLA violation avoidance from a pre-interaction phase.

In our previous work [4] we proposed a viable SLA management framework that comprise of two time phases – pre-interaction time phase and a post-interaction time phase. A pre-interaction phase consists of all steps before SLA execution and when both parties agreed and signed the agreement then a post-interaction phase start. A proposed SLA management framework is presented in Fig. 1. The pre-interaction phase comprises of two modules Identity manager module (IMM) and a viable SLA module (VSLAM). We described the pre-interaction section in our previous work [5, 6]. The post-interaction section is comprised of four modules – threshold formation module, runtime QoS monitoring module, QoS prediction module and risk management module. Modules in pre-interaction phase are responsible to authenticate requesting consumers and by considering their previous profile take a decision on consumer request for marginal resources and the amount of resources offer to them. In post-interaction phase, a threshold formation module (TFM) form a threshold and by observing a runtime behaviour of a consumer the Qos prediction module (PQoSM) QoS parameters for future intervals. If a system finds a difference between a values of PQoSM and the agreed QoS parameters then the risk management module (RMM) is activated, which consider reliability of a consumer, risk attitude of a provider and the predicted trajectory to decide an appropriate action.

The rest of the paper is organized as follows. Section 2 discuss related literature. Section 3 describe different components of our framework. Section 4 discuss implemention of a framework and Sect. 5 conclude a paper.

2 Related Work

Authors in [7] proposed risk based model to ensure the fulfilment of a SLA by a service provider and to maximize financial competence by considering a risk in a decision making process. A system identifies a risk and categorizes it into one of three levels – average risk, less risk and very less risk. Authors proposed three policies to minimize cost and risk both at node level and at graph level. To calculate the probability of a failure authors used statistical information from history data, however there is no comparisons for the optimality of a method with other methods like machine learning or non-linear regression methods. Authors did not use systematic estimation methods to estimate business values. Authors in [8] proposed a lightweight cloud platform for quickly access of changing resource information like CPU, memory etc. and to identify

a specific need of resources. The platform helps for efficient monitoring of SLA violation. Their framework is comprised of five modules and the SLA management module is responsible to monitor SLA violations in the application layer based on existing approach CASViD. Morin et al. in [9] identified the problems and challenges linked with the SLA violation in cloud. The exhaustive use of the Internet of Services raises serious issues about data security and privacy. They proposed that due to frequent changes of the status of services the existing information security risk management methods are insufficient which need to be improved for better performance. Consumer's profile history plays a key role to identify service violation. In our previous work [10], we categorized requesting consumer's based on their previous transactions. A consumer who has previous already communicated with a provider has a transaction record. We consider its track record to calculate its transaction trend and for all new consumers we find its nearest neighbors and calculate the transaction trend of a new consumer based on its nearest neighbor's transaction history. From evaluation results we found that a profile history has a significant impact in prediction of SLA violation. To avoid SLA violation a provider, need an optimal prediction method which can intelligently predict likely violation and alarm service provider to mitigate it. There are a number of prediction methods that have different prediction accuracy depending on a type of a dataset. In our previous work [11, 12] we considered a cloud dataset from Amazon and applied neural network, stochastic and other time series prediction methods to evaluate their prediction accuracy. From the evaluation results we found that ARIMA method has the most optimal results.

3 SLA Management Framework

In this section we discuss about our SLA management framework. As presented in Fig. 1, the framework is comprised of two time phases- pre-interaction and post-interaction time phase. Modules in each time phase is explained below:

3.1 Pre-interaction Time Phase

There are two modules in pre-interaction time phase, identity manager module and a viable SLA module. These are explained below:

- Identity manager module (IMM): This is the first module in our framework which is responsible for authentication and validation of a consumer. The transaction record of each consumer is stored in a profile repository. When a module receives a request of a consumer after validation it passes the request to viable SLA module along with its previous history.
- Viable SLA module (VSLAM): The module receives a consumer request from IMM along with the transaction history if it is an existing consumer. For a new consumer the module selects a transaction history from its top-K nearest neighbors. The module use FIS at two levels. First it finds the suitability of a consumer by considering a reliability of a consumer and contract duration and then it combines

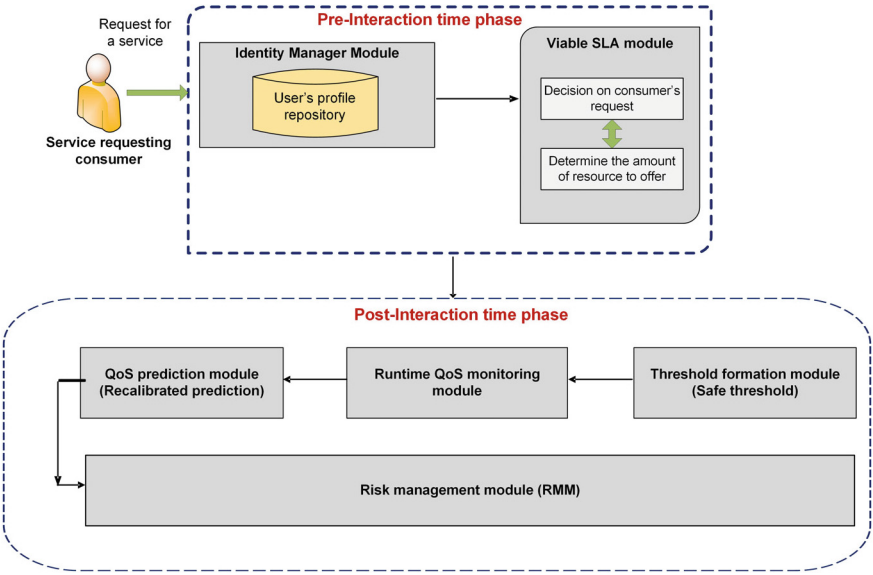

Fig. 1. Viable SLA management framework

the output with the risk attitude of a provider to decide the amount of resources offer to them. The detail of pre-interaction is explained in our previous work [5, 6].

3.2 Post-interaction Time Phase

The post-interaction phase is comprised of four modules – threshold formation module (TFM), runtime QoS monitoring module (RQoSM), QoS prediction module (QoSPM) and risk management module (RMM). The modules are explained below:

- Threshold formation module (TFM): Once both parties signed and execute a SLA, and then based on all agreed QoS parameters a provider defined its violation threshold. We propose two thresholds one is agreed threshold (Ta) which is defined in SLA and the second is safe threshold (Ts) which is a provider's threat threshold. Ts is more strict than a Ta, and when the runtime behavior reaches or exceed this threshold then it alert a service provider for managing a risk of service violation.
- Runtime QoS monitoring module (RQoSM): The module is responsible for monitoring runtime QoS parameters and send it to QoS prediction module (QoSPM) for recalibrated results.
- QoS prediction module (QoSPM): The module takes the input from a RQoSM and predict expected QoS parameters for future intervals. The value of QoSPM is compared with the agreed QoS parameters. If it finds that the value of QoSPM is reached or exceed the Ts value then the risk management module (RMM) is activated to manage risk.

- Risk management module (RMM): The module is started when a system find that predicted QoS parameters has reached or exceed the threat threshold. The module use FIS and take inputs – risk attitude of a provider, reliability of a consumer and predicted trajectory to determine an appropriate action to mitigate a risk. The action is either immediate action, delayed action or no action.

4 Implementation and Evaluation

In this section we evaluate our framework. We use two datasets from different sources for two phases of our framework. One from an existing dataset [13] which comprised of 142 users using 4,532 web services. We consider a throughput and a response time for 10 web services. For a second dataset we consider Amazon EC2 IaaS cloud services – EC2 US East, collected from cloudclimate [14] through the PRTG monitoring service [15]. We consider a CPU performance with 5 min of intervals for a duration of 4 days starting from 21 April 2015 to 25th April 2015 with 1007 observations. For evaluation we used Microsoft Visual Studio 2010 with Microsoft SQL Server Management Studio 2008 for the databases and MATLAB to design the FIS application. Figure 2 presents CPU, memory and I/O for the mentioned period.

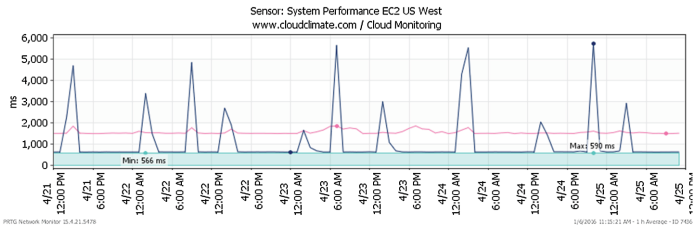

Fig. 2. Data from EC2 US West [15]

We consider a scenario in which a provider receives two request from a consumer A with a consumer ID 806 and from a consumer B with consumer ID 809. We divide the working of our framework into different steps which are explained below:

Step1: When a viable SLA management framework receives a request it is forwarded to the IMM module where it is checked from its stored repository for the identity and its previous history. In this case the IMM module found that both consumers have a profile history. The IMM forward the request to the VSLAM to decide about the request and the amount of resources offer to accepted requests.
Step 2: The VSLAM is a key module in our framework that decide about the request of consumer and the amount of resources offer to them. VSLAM receives previous history of consumers and based on their previous profile history it calculates the Ttrend of both consumers. Which is then compared with the threshold value. A provider has fixed a threshold value of 50 %. The T_{trend} calculated from the previous record of consumer 806 and 809 are 59.22 % and 44.88 % respectively.

In a case of first consumer the T_{trend} value is greater than the threshold however for second consumer the T_{trend} value is less than the threshold value hence a request is rejected. After a decision the VSLAM use the FIS by taking contract duration, reliability of consumer to calculate suitability of a consumer and then by considering the risk attitude of a provider then calculate the amount of resources offer to consumer A. The output of the VSLAM for consumer A is 49.01 % that means a system offers 49.01 % of the requested resource for the marginal resources [6]. The output by VSLAM is presented in Table 1.

Table 1. Request determination and resource allocation of requesting consumers [5]

ID	Transaction trend	Reliability	Suitability of consumer	Current suitability value		Risk attitude of provider	Risk propensity		Required suitability value	Resource allocation	Decision of consumer request
806	59.22 %	42	45.17	Med = 0.90	Low = 0.10	0.6	RN = 0.80	RT = 0.20	Medium = 0.7	49.01 %	Accept
809	44.88 %	13.6	–	–	–	–	–	–	–	–	Reject

Step 3: In a third step a consumer and a provider both are agreed on each service level objectives and each QoS parameters, mentioned in SLA. At this stage the process of post-interaction start. The first module in post-interaction phase is a threshold formation module. We considered EC2 cloud dataset and consider QoS parameter - CPU. The agreed threshold T_a value between consumer and a provider is 290 ms and a provider set its T_s value as 250 ms. For this phase we considered 10 time intervals that starts from 5:00 AM and end at 7:30 AM. The T_s and T_a is presented in Fig. 3.

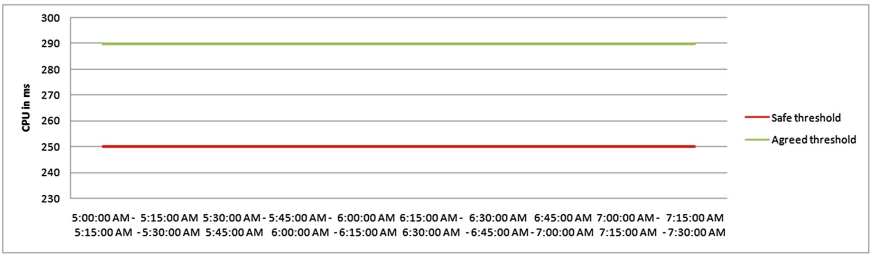

Fig. 3. Agreed and safe threshold

Step 4: When the transection start the runtime behaviour of consumer is recored and the value of RQoSM is forwarded to QoSPM, that use an intelligent prediction method to predict for future intervals. For this experiment we considered ARIMA method because it has an optimal result [11]. The predicted result for 10 intervals are presented in Fig. 4.

Step 5: From a Fig. 4 we see that till 6:00:00 AM the predicted result is below the T_s value so framework let a system to execute, but we see at time interval 6:15:00 AM a predicted result has touch T_s value and but at the next interval 6:20:00 AM come back below the T_s value. At time interval 7:15:00 AM we see that predicted result exceed T_s value and moving towards T_a. At this instance the risk management module is activated to manage a risk.

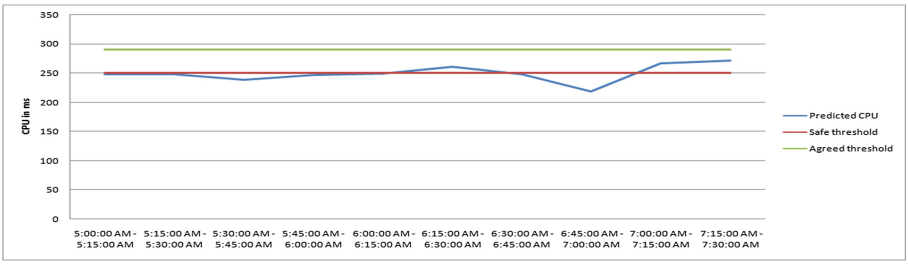

Fig. 4. Prediction result for 10 intervals

Step 6: The RMM take reliability of consumer, risk attitude of a provider and predicted trajectory and use FIS to decide either to take immediate action, delayed action or no action.

5 Conclusion

SLA is a crucial contract between a consumer and a provider that let them for executing their business. To enhance its trust value and to avoid from penalties a service provider need a framework that help in decision making for SLA formation, its monitoring and a mechanism that should inform a service provider to take immediate action when there is a risk of a service violation. Our viable SLA management framework assist a service provider to achieve all mentioned objectives. From the evaluation result we observed that our framework not only enable a service provider to monitor SLA in both phases of SLA life cycle but it also helps a service provider to manage a risk of a service violation.

References

1. Bahtovski, A., Gusev, M.: Analysis of cloud portability. In: The 10th Conference for Informatics and Information Technology, pp. 280–284 (2013)
2. Gartner, I.: Forecast: public cloud services. Worldwide, 2013-2019 (2016). (Gartner)
3. Hussain, W., Hussain, F.K., Hussain, O.K.: Maintaining trust in cloud computing through SLA monitoring. In: Loo, C.K., Yap, K.S., Wong, K.W., Beng Jin, A.T., Huang, K. (eds.) ICONIP 2014, Part III. LNCS, vol. 8836, pp. 690–697. Springer, Heidelberg (2014)

4. Hussain, W., et al.: Profile-based viable service level agreement (SLA) violation prediction model in the cloud. In: 2015 10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), Krakow, Poland, pp. 268–272. IEEE (2015)
5. Hussain, W., Hussain, F.K., Hussain, O.: Allocating optimized resources in the cloud by a viable SLA model. In: 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). IEEE, Vancouver, Canada (2016)
6. Hussain, W., et al.: Provider-based optimized personalized viable SLA (OPV-SLA) framework to prevent SLA violation. Comput. J., 24 (2016)
7. Macias, M., Guitart, J.: A risk-based model for service level agreement differentiation in cloud market providers. In: Magoutis, K., Pietzuch, P. (eds.) DAIS 2014. LNCS, vol. 8460, pp. 1–15. Springer, Heidelberg (2014)
8. Liu, D., Kanabar, U., Lung, C.-H.: A light weight SLA management infrastructure for cloud computing. In: 2013 26th Annual IEEE Canadian Conference on Electrical and Computer Engineering (CCECE). IEEE (2013)
9. Morin, J.-H., Aubert, J., Gateau, B.: Towards cloud computing SLA risk management: issues and challenges. In: 2012 45th Hawaii International Conference on System Science (HICSS). IEEE (2012)
10. Hussain, W., Hussain, F.K., Hussain, O.: Comparative analysis of consumer profile-based methods to predict SLA violation. In: FUZZ-IEEE. IEEE, Istanbul Turkey (2015)
11. Hussain, W., Hussain, F.K., Hussain, O.: QoS prediction methods to avoid SLA violation in post-interaction time phase. In: 11th IEEE Conference on Industrial Electronics and Applications (ICIEA 2016). IEEE, Hefei (2016)
12. Hussain, W., Hussain, F.K., Hussain, O.K.: Towards soft computing approaches for formulating viable service level agreements in cloud. In: Arik, S., Huang, T., Lai, W.K., Liu, Q. (eds.) ICONIP 2015. LNCS, vol. 9492, pp. 639–646. Springer, Heidelberg (2015). doi:10.1007/978-3-319-26561-2_75
13. Zhang, Y., Zheng, Z., Lyu, M.R.: WSPred: a time-aware personalized QoS prediction framework for Web services. In: 2011 IEEE 22nd International Symposium on Software Reliability Engineering (ISSRE). IEEE (2011)
14. CloudClimate: Watching the Cloud. http://www.cloudclimate.com/
15. Monitor, P.N.: PRTG Network Monitor. https://prtg.paessler.com/