Trust-aware Operation of Providers in Cloud Markets

Mario Macías and Jordi Guitart

Barcelona Supercomputing Center (BSC) and
Universitat Politecnica de Catalunya - Barcelona Tech (UPC)
Jordi Girona 29, 08034 Barcelona, Spain
{mario.macias, jordi.guitart}@bsc.es

Abstract. Online Reputation Systems allow markets to exclude providers providers that are untrustworthy or unreliable. System failures and outages may decrease the reputation of honest providers, which would lose potential clients. For that reason, providers require trust-aware management policies aimed at retaining their reputation when unexpected failures occur. This paper proposes policies to operate cloud resources to minimise the impact of system failures in the reputation. On the one side, we discriminate clients under conflicting situations to favour those that would impact more positively the reputation of the provider. On the other side, we analyse the impact of management actions in the reputation and the revenue of the provider to select those with less impact when an actuation is required. The validity of these policies is demonstrated through experiments for various use cases.

1 Introduction

Cloud Computing allows clients to acquire resources (usually Virtual Machines, VMs) and size them dynamically according to their spot requirements and pay only for what they use. Our research is framed in Open Cloud Markets [4] where both clients and providers are autonomous agents that negotiate the terms of the Quality of Service (QoS) and the pricing. After the negotiation, the terms of the contract are stored in a Service Level Agreement (SLA).

Cloud providers may not always fulfil the SLAs they agree with the clients. Online reputation systems [3] allow their users to submit and retrieve information about the fulfilment rate of the SLAs for each provider. Reputation systems enforce the confidence between parties and boost the number of commercial transactions, but they are vulnerable to reputation attacks: a group of dishonest clients may report false values about the QoS that a cloud provider is actually offering [1]. In consequence, reputation systems must also establish trust relationships between peers to avoid dishonest reports.

Reputation allows markets to exclude dishonest providers. However, spot failures or system outages may decrease the reputation of honest providers, having a double economic impact: the provider must pay penalties as agreed in the SLAs and it will lose potential clients due to the loss of reputation. For this reason,

providers operating in a Cloud market require trust-aware management policies aimed at retaining their reputation when unexpected failures occur.

We propose policies to discriminate users in function of their reputation under some situations that force provider to violate SLAs, such as errors in resources allocation, or an outage that makes unavailable part of the resources in a data center, dealing with the issue raised by Xiong et al. [6] and Kerr et al. [1], which considered the necessity of a community-context factor to incentivise peers for reporting true feedbacks.

The main contributions of this paper are: (1) introduction of policies to minimize the impact of system failures in the reputation. We discriminate clients according to their reputation to favour those with high reputation under some conflicting situations, because the reports of those clients will impact more positively the reputation of the provider; (2) evaluation of the impact of the management actions in the reputation and the revenue of the provider to select those with less impact when an actuation is required.

The rest of this paper is organised as follows: Section 2 summarizes the reputation model and describes the policies, which are evaluated in Section 3. Then, we present the conclusions and some future research.

2 Trust-aware SLA management

We consider a group of providers that are competing in a market to sell their resources to the potential clients. The clients are also communicated between them by means of a Peer-to-Peer network. When a client wants to buy a resource, it sends an offer to the providers to start a negotiation and agree a SLA, which is described as $\{\overrightarrow{S}, \Delta t, Rev(vt)\}$. $\overrightarrow{S}$ are the Service Level Objectives (SLOs) that describe the QoS to be purchased by the client. Δt is the time period when the task will be allocated. $Rev(vt)$ is a revenue function that describes how much money the provider earns or loses after finishing a task. The Violation Time (vt) is the amount of time in which the provider has not provided the agreed QoS to the client. Let MR the Maximum Revenue, MP be the Maximum Penalty (a negative revenue), MPT the Maximum Penalty Threshold, and MRT the Maximum Revenue Threshold, we describe our revenue function as follows:

$$Rev(vt) = \frac{MP - MR}{MPT - MRT}(vt - MRT) + MR$$

If $vt < MRT$ the SLA is not violated (0 violations); if $vt > MPT$, the SLA is completely violated (1 violations). $MPT > vt > MRT$ is a partial violation. Please refer to Section 3 for more details about $Rev(vt)$ and its concrete values.

Both clients and providers are entities that have a degree of trust between them as individuals. Trust relations are provided by the reputation model described in our previous work [3], which demonstrated its validity to identify trustworthy providers and expel dishonest peers to protect the system against reputation attacks. In this model, a trust relation is expressed as $\overrightarrow{T}(A, B) = \omega_1 \overrightarrow{D}(A, B) + \omega_2 \overrightarrow{R}(B)$; $\overrightarrow{D}(A, B)$ is the direct trust from A to B, which is built

based on previous experiences between A and B; $\overrightarrow{R}(B)$ is the reputation trust, which is calculated by asking to other clients about their past experiences with B; ω_1 and ω_2 are used to weight each term, and may vary in each particular client. Trust values vary from 0 (no trust) to 1 (maximum trust). Trust relations are continuously updated assuming that most peers are honest and, when asked, they report their true trust toward the provider. Related work considers many incentives to peers to report honestly [7]. Our contribution is complimentary to them, since we deal with the minimization of the impact of the dishonest reports.

To select a provider, a client sends SLA templates to all the providers that match its requirements. If the providers have enough resources to handle the request, they return a price. The client then scores all the providers and chooses the provider with the highest score, which may vary depending on the client preferences. In this work, we score providers as $\frac{-\overrightarrow{T}(c_x, cp_y)}{Price}$: the client would accept sending tasks to providers to which the trust is lower if the price they establish is low enough. That would motivate providers to keep its maximum trust level and, if not possible, to lower prices. The higher QoS is provided to a client, the higher trust values he will report to the reputation system; unless the client is behaving dishonestly and reporting false values.

We propose to maximise the reputation as a key objective that will help providers to increase their revenue due to the enforcement of the trust relationship with their clients. This paper considers selective SLA violation and/or cancellation for minimizing the impact of resource failures and overloading: to prioritise trustworthy users under certain situations in which a set of SLAs that are already allocated must be violated temporarily or directly cancelled.

When the monitoring system of a provider detects that there are not enough resources to fulfil the workload for all the VMs in a given node, the next process is triggered: the VMs are ordered according to a *given criterion* and the provider pauses the VMs on the top positions during t time. When the node can provide the QoS for the VMs that are still running, the provider stops pausing VMs. In reputation maximisation policies, the criterion to order VMs is the trustworthiness to the client that owns it. To calculate the trustworthiness to a client, the provider can join the reputation system as a normal peer, and poll several clients about several providers. If a given client is usually reporting values that are far away from the average, it will be considered unreliable.

The SLA violation algorithm is generic enough to achieve other BLOs, such as revenue maximisation [2]. Next section will evaluate the effectiveness of the trust maximisation policy by comparing it with the same policy for other BLOs.

We must emphasize that our policy cancels SLAs only when the provider is not able to fulfil them all. This should be infrequent, only when the violation is unavoidable, because the economic penalty is paid whatever the client trustworthiness is. The idea is at least to minimize the impact in the reputation of the provider. A bad usage of this policy could make the clients lose the confidence in the provider, thus losing profit.

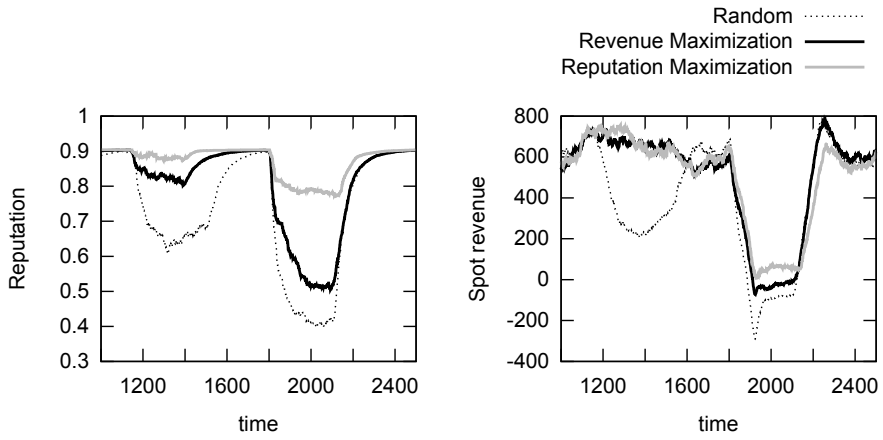
3 Evaluation

To evaluate the effectiveness of our models, we have created a simulated environment [5] that is available online to facilitate replicating the experiments. The simulator reproduces the model and market negotiation steps that are described in Section 2 and the related work [3]. The trustworthiness of the clients follows a folded normal distribution with mean=0.5 and standard deviation=0.2. That means that most clients have trust values near 1 and a few clients are reporting dishonestly. The values for $Rev(vt)$ are the next: $MRT = 0.05$ (that means that if the agreed QoS is not provided during the 5% of the time or less, the SLA is not considered as violated); $MPT = 0.3$ (when the agreed QoS is not provided during the 30% of time or more, the SLA is completely violated); MR is dynamically established according to $\overrightarrow{S}$ and the market status [2]; $MP = -1.5MR$ (if the provider completely violates a SLA, it must pay back the 150% of the price that the client paid initially). The providers normally provide the 100% of the agreed QoS during off-peak hours and around 97% during peak hours. The workload follows a web pattern that varies in function of the hour of the day and the day of the week. The simulations rely on constant values that do not intend to reflect real market data, but to evaluate the model in terms of relative values and tendencies.

To evaluate the reputation-aware resources operation, we have simulated 5 days of a market operation with three types of providers, according to their policy for discriminating SLAs during resources overload: (1) a provider that randomly discards SLAs, used as a baseline to evaluate the system behaviour without any policy; (2) a provider that discards the SLAs that report less revenue [2]; and (3) a provider that discards the SLAs from the clients to which there is low trust. We introduced a global outage of the data centre at day 3 of the simulation. During the outage, the providers only have the 20% of their usual resources. The outage has been programmed to happen during a peak of workload. That means that about 80% of the allocated SLAs are to be violated during the outage.

Figure 1a shows reputation for day 2 (normal day) and day 3 (outage). Reputation is near the maximum value for all the providers at off-peak hours. During the peaks, the graph shows the effect of the increase of SLA violations in the reputation. The reputation maximisation policy keeps reputation near 1 during both peaks and off-peaks. The random policy shows that reputation decreases when no policies are applied. Although the revenue maximisation policy does not consider reputation, it indirectly keeps it between random and reputation maximisation policies: the provider tries to first pause the VMs whose SLA violation time is over MRT. In consequence, it violates less SLAs and, indirectly, the reputation of the provider is higher than the reputation of the provider that does not apply any policy.

Figure 1b shows the evolution of the spot revenue for the three providers during the outage, demonstrating that maintaining high reputation during the outage has a real impact in the revenue of the provider. However, we measured that during normal operation the revenue of the provider that maximises the reputation is slightly lower.



(a) Evolution of reputation

(b) Spot revenue

Fig. 1: Measurements after and before an outage

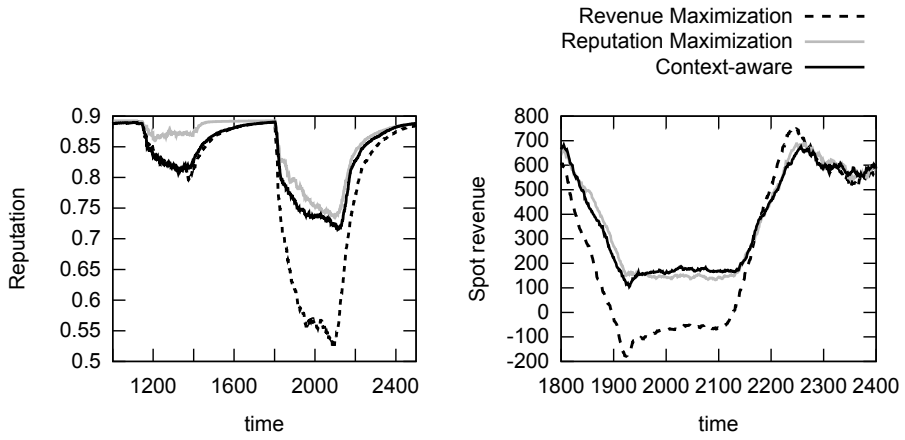
Considering the previous observations, we have introduced a new provider that is aware of the environment (normal operation or outage) and dynamically switches the revenue/reputation maximisation policy in function of what is expected to report the highest economic profit. When the provider is operating normally it uses the revenue maximization policy; if the monitoring information shows that there is an outage in part of the resources, the provider switches to the reputation maximization policy and returns back to revenue maximization policy when the systems are again in normal operation.

Figure 2a shows that the context-aware provider maintains a reputation rate similar to the revenue-maximisation provider during normal operation and a rate similar to the reputation-maximisation provider during the outage.

The revenue of the context-aware policy is similar to the revenue maximisation policy during normal operation. Figure 2b shows the time window that comprehends an outage of the system and the subsequent recovery, demonstrating that the revenue of the context-aware policy is similar to the reputation maximisation policy during an outage.

4 Conclusions and future work

Under certain situations, such as errors in the estimation of resources or an outage in a cloud provider, part of the VMs that are being hosted must be paused to allow enough free resources for fulfilling the other SLAs. This paper introduces a policy to prioritize users according to their trustworthiness. This policy has a double goal: (1) to minimize the impact of the SLA violations in the reputation of the provider and, in consequence, in the revenue; and (2) incentivise users to report true validations of the providers.



(a) Evolution of reputation

(b) Spot revenue

Fig. 2: Measurements for context-aware policy

After analysing the policy in comparison with others, we show that providers that behave honestly and apply revenue maximisation policies, in most cases indirectly keep a good enough reputation rate and achieve higher revenue than the providers that apply reputation maximisation. The benefits of reputation maximisation in terms of revenue are noticeable under conditions that imply a high rate of SLA violations. Considering the aforementioned, we introduce a new type of provider that switches between reputation or revenue maximisation policies depending on the context. This provider achieves the best revenue in all the cases, and always keeps good-enough reputation rates.

A key issue of reputation systems is to incentivise their users to report true validations of the providers [3]. The policies of this paper help solving it because clients that report true validations have high reputation to their peers. Providers that have interest on keeping high reputation will prioritize the QoS for trustworthy clients under certain situations such as peaks of demand or an outage. As a consequence, clients that want to benefit from this positive discrimination will report true validations of the providers. Since the reputation model is P2P, any provider could join a network for polling the trustworthiness of a client.

This paper validates the model by means of a simulated environment because there are no real market traces of this type of market model. In addition, we would need to fully use a large data centre during many days to generate data that is statistically representative enough. Using a simulated environment allows to solve such issues. Although the results do not reflect real information in *quantitative* terms, the objective of this paper is to evaluate the policies in *qualitative* terms: we measured that the application of a given policy can improve the trust or the revenue of a provider in a significant proportion.

This paper does not consider the ethical issues of using such policies by dishonest providers for always cheating the clients with low reputation: not only dishonest clients, but also clients that recently joined the reputation network.

In future work we will improve the context-aware provider by adding statistical analysis to dynamically learn how the actions of the provider during negotiation and operation can influence the future reputation. We also plan to improve the policy for selecting the SLAs that are going to be violated. The objective is to achieve a policy that is able to ponder both reputation and revenue maximisation objectives. In addition to the improvement in the model, future work will include new policies to complement the selective violation/cancellation of SLAs. For example, to apply price discounts to the clients to which there is high trust or to use migration capabilities to redistribute VMs for decreasing the violation rate of those SLAs from trusted clients.

Acknowledgment

This work is supported by the Ministry of Science and Technology of Spain and the European Union (FEDER funds) under contract TIN2012-34557 and by the Generalitat de Catalunya under contract 2009-SGR-980.

References

1. Kerr, R., Cohen, R.: Smart cheaters do prosper: defeating trust and reputation systems. In: 8th International Conference on Autonomous Agents and Multiagent Systems. AAMAS '09, vol. 2, pp. 993–1000. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2009)
2. Macias, M., Fito, O., Guitart, J.: Rule-based SLA management for revenue maximisation in cloud computing markets. In: 2010 Intl. Conf. of Network and Service Management (CNSM'10). pp. 354–357. Niagara Falls, Canada (October 2010)
3. Macias, M., Guitart, J.: Cheat-proof trust model for cloud computing markets. In: Proceedings of the 9th International Conference on Economics of Grids, Clouds, Systems and Services (GECON 2012). Lecture Notes in Computer Science (LNCS), vol. 7714, pp. 154–168. Berlin, Germany (2012)
4. Neumann, D., Stoesser, J., Anandasivam, A., Borissov, N.: SORMA - building an open grid market for grid resource allocation. In: 4th international conference on Grid economics and business models. pp. 194–200. Springer-Verlag, Berlin (2007)
5. Reputation-aware cloud market simulator. Online, https://github.com/mariomac/reputation
6. Xiong, L., Liu, L.: Peertrust: Supporting reputation-based trust for peer-to-peer electronic communities. Knowledge and Data Engineering, IEEE Transactions on 16(7), 843–857 (2004)
7. Zhang, J.: Promoting Honesty in Electronic Marketplaces: Combining Trust Modeling and Incentive Mechanism Design. Ph.D. thesis, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada (May 2009)