Journal of Network and Computer Applications 63 (2016) 159–172

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

# Physics and microeconomics-based metrics for evaluating cloud computing elasticity

Emanuel F. Coutinho [a,d,*], Paulo A. L. Rego [d], Danielo G. Gomes [b,d], José N. de Souza [c,d]

[a] Virtual University Institute (UFC VIRTUAL), Brazil
[b] Department of Teleinformatics Engineering, Brazil
[c] Computer Science Department, Brazil
[d] Federal University of Ceara (UFC), Fortaleza, Brazil

ARTICLE INFO

*Article history:*
Received 4 July 2015
Received in revised form
16 November 2015
Accepted 3 January 2016
Available online 6 February 2016

*Keywords:*
Cloud computing
Elasticity
Metrics
Performance analysis

ABSTRACT

Currently, many customers and broadband providers are using cloud resources, such as processing and storage, for their applications and services. With the increase of computational resources usage, elasticity has become quite attractive and a key feature in cloud computing paradigm. However, due to many technologies and strategies for providing cloud elasticity, its measurement is not obvious. Here we propose a set of metrics for measuring cloud computing elasticity based on Physics' concepts, such as strain and stress, and Microeconomics. Our experimental results show that the proposed elasticity metrics behave well towards the CPU usage. Besides, the metrics computation is simpler because it is straightforward, which speeds up the cloud elasticity evaluation.



## 1. Introduction

Elasticity can be defined as how a cloud adapts to variations in its workload through resources provisioning and deprovisioning. 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 (Mell and Grance, 2009). To the consumer, the resources available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time. Coutinho et al. (2015) performed a study in cloud elasticity, highlighting various aspects such as definitions, metrics, benchmarks, elastic strategies, challenges and trends in the construction of elastic solutions.

Currently, resources are pooled on virtualization platforms, and an elasticity degree is present (Manvi and Shyam, 2014). Services are more on demand than ever before, and there is some automation level present. A most recent work (Herbst et al., 2013) defines elasticity as 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 combine with the demand of workload as close as possible.

In cloud computing, there is a diversity of definitions and metrics for scalability, elasticity, and efficiency (Lehrig et al., 2015). Moreover, stakeholders (e.g., software architects, cloud consumers, and cloud providers) have little guidance for choosing suitable definitions and metrics for these quality properties, thus leading to potential misunderstandings.

Due to the dynamic characteristics imposed by different workloads, elasticity is becoming an increasing need in cloud computing environments. In order to maintain the Service Level Agreement (SLA), providers are developing resource monitoring services to support elasticity.

In this context, various cloud providers offer elasticity mechanisms for their users, such as Amazon EC2,[1] Microsoft Azure,[2] and HP Public Cloud.[3] However, the way to evaluate the elasticity solution performance of an environment is not an easy task due to many aspects, such as (i) lack of standardization; (ii) each service provider performs this task in a particular way and different from other providers; (iii) there are many technologies and strategies for providing elasticity; (iv) finally, it is not common to use specific metrics for measuring the elasticity of cloud environments, but indirectly associated metrics.

Different strategies for evaluating elasticity have been proposed in the literature (Islam et al., 2012; Shawky and Ali, 2012; Herbst

* Principal corresponding author at: Virtual University Institute (UFC VIRTUAL), Brazil. Tel.: +55 85 98875 1977.
*E-mail address:* emanuel@virtual.ufc.br (E.F. Coutinho).

[1] Amazon EC2 – http://aws.amazon.com/.
[2] Microsoft Azure – https://azure.microsoft.com/.
[3] HP Public Cloud – http://hpcloud.com/.

http://dx.doi.org/10.1016/j.jnca.2016.01.015
1084-8045/

et al., 2013; Ferreira Coutinho et al., 2013). The manner to measure the elasticity is shown to be quite varied in several work, and there is still no standardized way for measuring elasticity. A common aspect to evaluate elasticity is the use of resources such as CPU, memory, and throughput to indirectly assess the elasticity, even without a specific metric for this task.

Apart from Computer Science, the concept of elasticity is present in different areas such as Physics, Biology and Microeconomics. In Physics, all structural material has a certain extent of the elasticity property (Timoshenko and Goodier, 1970). If external forces applied to the material produce a deformation in the structure, this deformation disappears with the removal of forces, since they do not exceed a certain limit. In Biology, specifically in Metabolic Control Analysis, elasticity is used to quantify the effect of an effector or substrate concentration on an enzyme rate under locally specified conditions (Woods and Sauro, 1997). The elasticity coefficient describes the response of the isolated rate of an enzyme or transporter to a movement in concentration of a substrate, product or other effector. In Microeconomics, elasticity can be defined as a measure of the responsiveness of quantity demanded or quantity supplied to a change in one of its determinants (Availability of Close Substitutes, Necessities versus Luxuries, Definition of the Market and Time Horizon) (Mankiw, 2012). In cloud computing context, some authors used these ideas to discuss about elasticity, such as Gambi et al. (2013) and Shawky and Ali (2012). They described some elasticity analogies from other areas.

In this scenario, this work proposes a set of metrics to measure elasticity in cloud environments. For validating these metrics, two experiments were designed in a private cloud, using workloads generated by microbenchmarks and scientific applications to evaluate the elasticity. Our main contributions are as follows:

- Metrics for evaluating elasticity in cloud computing based on Physics' concepts of strain and stress (Section 2.1);
- Metrics for evaluating elasticity in cloud computing based on Microeconomics' concepts of Price Elasticity of Demand (Section 2.2);
- Case study with experiments performed in a computational cloud for elasticity evaluation, using workloads generated through microbenchmarks (Section 4).

## 2. Metrics for measuring elasticity in cloud computing

The concept of elasticity is common in fields of knowledge different of Computing, such as Physics (Physics of Materials, Pneumatics, Hydraulics), Biology, Chemistry, Control Theory and Microeconomics. In this section, metrics based on Physics' and Microeconomics' concepts are proposed. Both areas have specific metrics and secondary metrics to calculate the value of elasticity, and thus, analogies to these metrics are presented with their respective calculation.

### 2.1. Physics-based metrics

In Physics, according to Timoshenko and Goodier (1970), all structural material has a certain extent of the elasticity property. If external forces applied to the material produce a deformation in the structure, this deformation disappears with the removal of forces, since they do not exceed a certain limit. Consequently, we studied the stress and strain concepts of Physics, and we proposed metrics for elasticity's evaluation in computational clouds.

Figure 1 shows how stress and tension act on a material. Stress can be calculated by Eq. (1), where $F$ is an applied force and $A$ is the cross sectional area of the material. The strain can be calculated by Eq. (2), where $\Delta l$ corresponds to the variation in the material size and $l_0$ is the original size of the material. Finally, Hooke's law states that deforming forces are proportional to the produced elastic deformations, and thus the elasticity can be measured dividing the stress by the strain, according to Eq. (3).

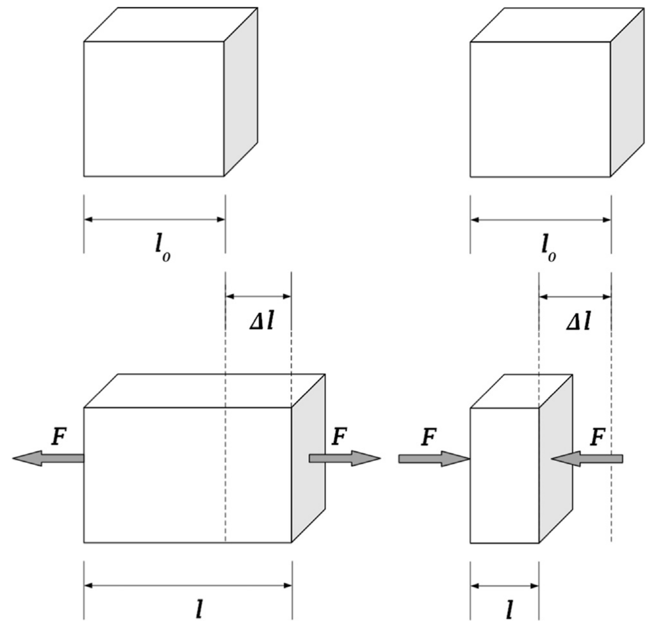


**Fig. 1.** Representation of stress and strain effects on a material (Young and Freedman, 2011).

$$\text{Stress} = \frac{F}{A} \tag{1}$$

$$\text{Strain} = \frac{\Delta l}{l_0} \tag{2}$$

$$\text{Elasticity} = \frac{\text{Stress}}{\text{Strain}} \tag{3}$$

In this context, we propose an analogy between the Physics elasticity concept and the cloud computing elasticity concept, because of its conceptual similarity. Table 1 describes the variables related to proposed metrics, as the nomenclature to be used. The metrics associated with resources have their values collected from the environment, as CPU utilization or number of allocated virtual machines. The variable $t$ is the number of iterations in which the different parameters are collected, defined by the collection interval. The $i$ in these variables is to reinforce that their values are obtained in each time interval defined by the collection interval.

In Physics, stress indicates how strong a material is. Timoshenko and Goodier (1970) and Shawky and Ali (2012) define stress as the amount of pressure that the material can withstand without suffering from some kind of physical change. Whereas stress in Physics is a force acting on the area of cross section of the material, making the analogy with cloud computing, where $D_{R_i}$ are the demanded resources imposed by workloads, and $A_{R_i}$ are the allocated resources in current infrastructure configuration. Thereby, we have that Cloud Quality Stress ($CQS_i$) can be measured by the following Eq. (4)

$$CQS_i = \begin{cases} 0 & \text{if } A_{R_i} = 0 \\ \dfrac{D_{R_i}}{A_{R_i}} & \text{otherwise} \end{cases} \tag{4}$$

Some specific situations must be observed. In the case that the demanded and allocated resources are zero, the stress is considered as zero. Given that resources are different from zero and allocated resources are equal to zero, the stress tends to infinity.