Modeling and Calculation of Elasticity in Cloud Computing

A. Outamazirt, D. Aïssani et K. Barkaoui

Doctoriales de Recherche Opérationnelle, le 12 et 13 Décembre 2018

Introduction

cloud computing, elasticity is defined as e degree to which a system is able to apt to workload changes by provisioning d de provisioning resources in an tonomic manner, such that at each point time the available resources match the rrent demand as closely as possible.

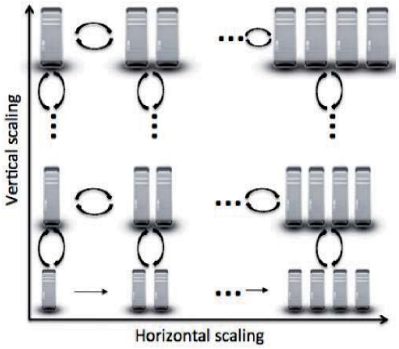

Fig. 1: Vertical vs. horizontal elasticity.

roblems and Motivation:

The long unexpected VM start-up time resource under-provisioning.

The long unexpected VM shut-down me resource over-provisioning.

Minimize the number of active servers.

Minimize the transition from “on” to ff” and vice versa.

Methodology

The proposed mathematical models are based primarily on queuing models and Markov chains (see Figure 2 and Figure 3). These models allow to calculate the elasticity value of a Cloud Computing platform.

$$E = \frac{T_{\text{normal}}}{T} = 1 - \frac{T_{\text{over}} + T_{\text{under}}}{T}. \quad (1)$$

$$E = p_{\text{normal}} = 1 - (p_{\text{over}} + p_{\text{under}}). \quad (2)$$

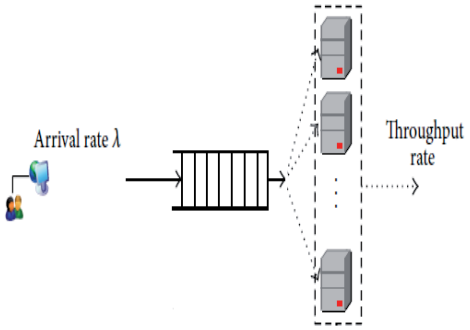

Fig. 2: M/M/c/k queuing system with queue-dependent servers .

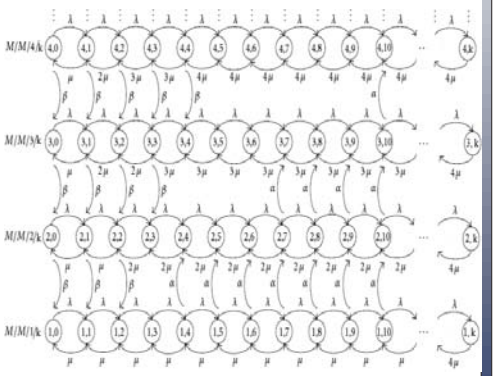

Fig. 3: Markov chain

Results

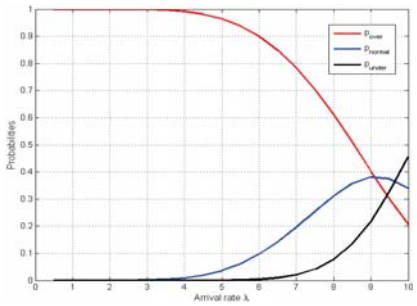

Fig. 4: p_{over}, p_{under}, p_{normal} vs. arrival rate

Discussion: It is observed that as arrival rate increases, $p_{}$ decreases (i.e., more service requests result in less probability of over-provisioning), and p_{under} changes slight (actually, increases and then decreases, i.e., more service requests result in slight change of the probability of under provisioning), and p_{normal} increases (i.e., the elasticity increases).

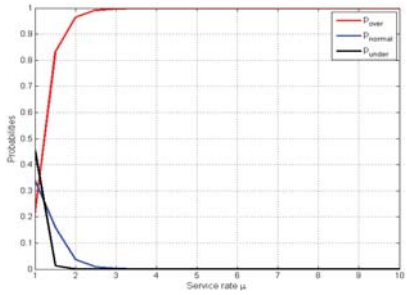

Fig. 5: : p_{over}, p_{under}, p_{normal} vs. service rate

Discussion: It is observed that as service rate increases, $p_{}$ increases significantly (i.e., faster service rate results in greater probability of over-provisioning), and p_{under} chang noticeably (actually, increases and then decreases, i.e., faster service rate results in noticeable change of the probability of under-provisioning), and p_{normal} decreases significantly (i.e., the elasticity decreases significantly).

Conclusion

We developed an analytical model to study elasticity by treating a Cloud platform as a queuing system, and we used a continuous time Markov chain model to precisely calculate the elasticity value of a Cloud platform.

References

1. K. Li, *Quantitative Modeling and Analytical Calculation of Elasticity in Cloud Computing*, IEEE Transactions on Cloud Computing, pp. 1-1, 2017
2. W. Ai, K. Li, S. Lan, F. Zhang, J. Mei, K. Li, and R. Buyya, *On Elasticity Measuremen in Cloud Computing*, pp.13, 2016.
3. Y. Al-Dhuraibi, F. Paraiso, N. Djarallah and *Ph. Merle, Elasticity in Cloud Computin*