Cloud Edge Computing for Internet of Things Elasticity Management

Rayene Moudjari
LIRE Laboratory, Constantine II
University Abdelhamid Mehri
Constantine, Algeria
rayanemath@gmail.com

Zaidi Sahnoun
LIRE Laboratory, Constantine II
University Abdelhamid Mehri
Constantine, Algeria
sahnounz@yahoo.fr

Allel Hadjali
LIAS, ENSMA
Poitiers, France
allelhadjali@ensma.fr

Abstract

A major ingredient that encourages the Cloud computing solution wide adoption is their ability to allocate and des-allocate resources in automatic and optimal manner according to customer's need or what is referred to Cloud Elasticity. Nowadays, with the emerging Cloud edge computing for IoT's systems, the existing centralized elasticity management solution must be revisited to cover the introduced new requirements. In this paper, we provide a first analysis step towards the Cloud Edge computing elasticity management solution.

Keywords: Elasticity; Cloud Computing; Edge Computing; Internet of Things.

1. Introduction

Over the past decades, Cloud Computing has been the model of choice for exposing resources as services, which means shielding users from the complexity of the cyber-physical world's resources, describing resources in a machine-readable format so they can be discovered, and shifting the burden of managing resources internally to Cloud providers' in-return of a fee (pay-as-you-go) [Mel11].

With the progressing development of information technology, Internet of Things (IoT) has come to play a crucial role in our daily lives. It is driving a digital transformation in all current modern lives aspects. IoT refers to the interaction and communication between billions of devices that produce and exchange data related to real-world objects [Atz10].

Cloud Computing has been used as a complementary technology for IoT due to its scalable and distributed data management scheme. In conventional Cloud Computing, all data must be uploaded to centralized servers, and after computation, the results need to be sent back to the sensors and devices. This process creates great pressure on the network, specifically in the data transmission costs of bandwidth and resources. Thus, supporting the billion connected things' huge amount of generated data, which are generally referred to as big data, based on Cloud exclusively is not an efficient solution. Especially, IoT applications that are time-sensitive where very short response times are non-negotiable. To this end, centralized big data processing era have pushed the horizon of a new computing paradigm called "Edge Computing" [Shi16]. This new paradigm proposes substantial computing and storage resources placed at the Internet's edge in close proximity to mobile devices or sensors. The edge computing has the potential to address the computing capability limitation, the concerns of response time requirement, bandwidth cost saving, as well as data safety and privacy.

In recent years, due to elasticity property usefulness, different works [Nas16, Gal12] have investigated the issues related to elasticity improvement. However, most of them consider elasticity control or management in Cloud computing specific environments, characterized by centralized, large datacenters [Par16, Dup15, and Moh15]. Actually, with the cloud edge computing emergence, systems are more complex, with geo-distribution, built upon several abstraction layers (As a result of multi paradigms convergence: IoT, Edge Computing, Cloud Computing). Such new characteristics impose new advanced mechanism rather the traditional centralized one, for a decentralized elasticity management solution able to cover systems new requirements.

The present paper constitutes an initiative or a first step towards elasticity management in cloud edge computing system. We analyze elasticity solution dimensions for complex, geo-distributed systems running on edge and cloud computing, their requirements, and possible or revisited solutions for decentralized Cloud edge computing elasticity management. The paper is organized as follows; Section 2 provides an overview about elasticity; Section 3 presents an analysis of Cloud edge computing elasticity management solution. Finally, we conclude the paper in Section 4 with ongoing works.

In: A. Editor, B. Coeditor (eds.): Proceedings of the XYZ Workshop, Location, Country, DD-MMM-YYYY, published at http://ceur-ws.org

2. Overview of Elasticity

One of the key factors motivating Cloud Computing popularity is *Elasticity* property, which is considered as unique characteristic of Cloud paradigm compared to the other technologies. Elasticity is the degree to which a system is able to adapt to workload changes by allocating and des -allocating resources in autonomic manner, such that at each point in time the available resources match the current demand as close as possible [Her13]. Thus, the elastic service should be able to collect new computing resources when their workload is increased, but also to discard any exceeding resources when workload is decreased. According to Galante and de Bona classification [Gal12], the Cloud elastic resource provisioning is ensured through three elementary methods: Horizontal scale, Vertical scale, and Migration.The emerging systems and services are characterized by the combination and the integration of different paradigms. These paradigms turn out to be particularly effective in moving computation and storage capabilities closer to data production sources and data consumption destinations. Recently, most of introduced solutions consider elasticity at large-scale and centralized Cloud data centers [Par16] or less frequently in distributed edge/fog environments. Managing elasticity at Cloud Edge computing is a new research direction that must be explored.

3. Cloud Edge Computing Elasticity

Cloud and edge computing complement each other to form a mutually beneficial and inter-dependent service continuum. Some functions are naturally more advantageous to carry out in centralized Cloud, while others are better suited to the edge. This complementary relation is described in [Maa18] through possible Cloud Edge Computing scenarios. Thing's life cycle is defined around three functions (Figure 1): Sensing (collecting/ capturing data); Actuating (processing data); and Communicating (distributing data). Similarly, Edge and Cloud Computing life cycle is defined around three functions (Figure2, Figure 3). Depending on application requirement, the edge/ cloud computing supports: Storing (saving data that was submitted by things, other edges, and/or clouds); Processing (acting on data that was submitted by things, other edges, and/or clouds); and relaying (transferring data to other parties). In addition to its internal storage / computing capabilities, edge/ cloud computing may relay data to other edge/ cloud computing so that it uses external storage/computing capabilities (Figure 2, Figure 3).

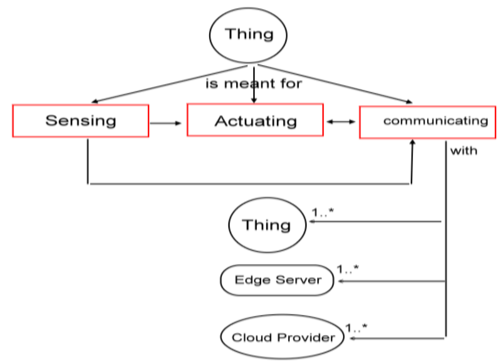

Figure 1: Thing's Life Cycle

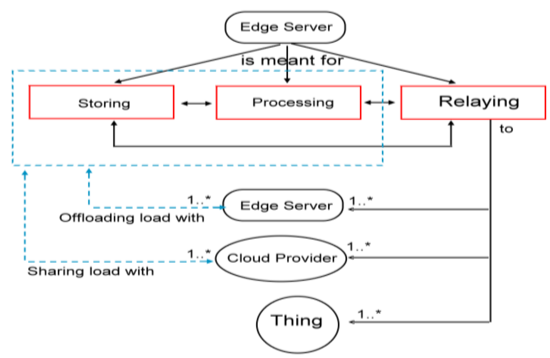

Figure 2: Edge Computing Life Cycle

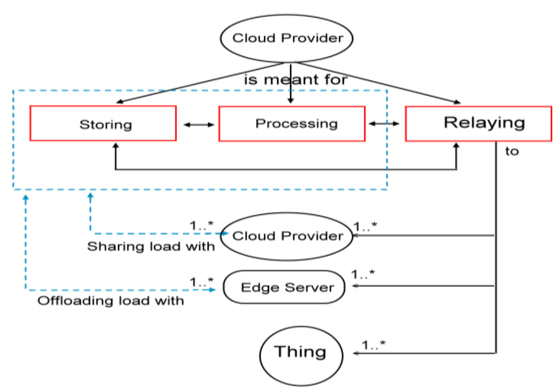

Figure 3: Cloud Computing Life Cycle

The Cloud Edge Computing life cycle analysis conducts us to define new possible elasticity dimensions by coordinating, horizontally and vertically, the system layers (Edge / Cloud Computing). For more visibility, we propose Cloud Edge Computing for IoTs model (Figure 4). The architecture considers three layers:

- *IoT Devices layer (Front-End):* consists of mobile IoT devices that can use sensors to sense information from their surrounding environment and utilize actuators to modify it. They are deployed at the front end of the edge computing structure.
- *Edge Computing Layer (Near-End):* consists of multiple distributed nodes (edge servers with medium computational capacity and limited storage) that provide functionalities for IoT devices under the orchestration of the cloud computing layer.
- *Cloud Computing Layer (Far-End):* defines the cloud servers in the far end environment that provide more computing power and more data storage.

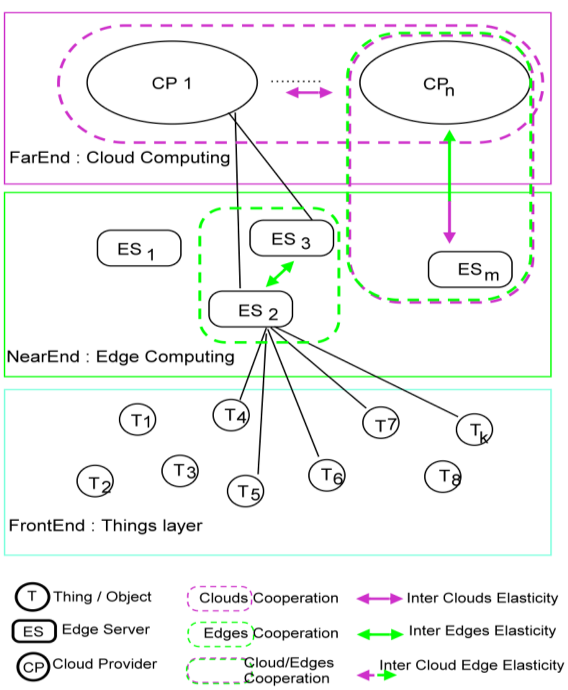

Figure 4: Cloud Edge Computing Elasticity Dimensions

According to pre-described life cycle, the presence of multiple layers with different computing capabilities, leads to new elasticity dimensions definition. The introduced dimensions are a result of system new layers coordination (see Figure 4), principally:

- *Inter Edges elasticity*: defined by Edges cooperation establishment. In this case, edge servers relay some data to external edges for storage/ processing. This solution is performed horizontally in system intra layer level (edge computing layer).
- *Inter Cloud Edge elasticity*: in contrary to the first one, this solution is performed vertically in system inter layers level (Cloud/edge computing layers). It is defined by Cloud-Edge Cooperation establishment. The edge (or Cloud) servers relay some data to external Clouds (or edges) for storage/ processing.

Despite cloud edge computing new elasticity dimensions possibilities, resulting from multiple layers computing capabilities convergence, different challenges must be explored and solved to ensure the desired elastic behavior. As reported by [Car18], managing elasticity for such emerging paradigms is crucial but in the same time challenging task. The first faced problem is determining which elasticity strategy to be executed. In other words, when it is more convenient to use resources from the same layer (edge / or Cloud)? When we should take resources from multiples layers? It is more advantageous, and for which property (cost, time, availability ...) resources should be taken by the same, or multiple layers?

Second, along the past years, the reference model for elasticity management for Cloud systems is the IBM autonomic control MAPE (Monitor, Analyze, Plane, and Execute) loop [Jac04]. To realize elasticity, we need to observe the system state evolution, define the elasticity adaptation to be performed, and finally reconfigure the system. Now, with the emergence of new complex, distributed Cloud edge computing environments, managing elasticity according to a single centralized decisional MAPE loop may not be sufficient, so multiples MAPE loops may be introduced to cooperatively make decision. For this purpose, different suggested design patterns [Wey13] must be explored, analyzed and adapted to Cloud edge computing system context.

In previous works [Mou17, Mou18], we proposed a multi agent system for cloud of clouds elasticity management. The system development is inspired from IBM autonomic control loop, principally three functions are considered: Surveying; Planning Adaptation Strategy, and Executing Adaptation Strategy. The system cooperative agents ensures elasticity management at both intra /and inter cloud levels. Now, with the definition of Cloud edge computing system new dimension, the proposed solutions should be revisited and extended to consider the discussed new system's architecture and requirements. As a plan, we intend to:

- Define an appropriate multiples MAPE loops design to support cooperative decentralized decision for Cloud edge computing elasticity control.
- Enrich the planner decisional logic (Elasticity Controller Agent) to be able to support new elasticity dimensions (*Inter Edges elasticity, Inter Cloud –Edge elasticity).*
- Adopt the proposed Top Clouds Cooperation [Mou18] mechanism in Inter Edges and Inter Clouds-Edges elasticity contexts, to define respectively, the possible Top Edges Cooperation and Top Clouds-Edges Cooperation.

4. Conclusion

In this paper, we presented an analysis of elasticity management in Cloud edge computing systems problem. First, we introduced the elasticity visions in both traditional cloud computing architecture and new Cloud edge computing systems. Next, we discussed the challenges related to designing elasticity mechanisms in those new systems. Finally, and as a result of Cloud edge computing systems analysis process, we identified new elasticity dimensions to be considered in our future elasticity management solutions.

References

[Mel 11] P. Mell, T. Grance, The NIST Definition of Cloud Computing, Special Publication, vol.800, National Institute of Standards & Technology, p.145, 2011.

[Atz 10] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey", Computer networks, vol. 54, no. 15, p. 2787–2805, 2010.

[Shi 16] W.Shi , J.Cao, Q.Zhang , J.Cao, Q.Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," IEEE Internet of Things Journal, vol. 3, no. 5, p.637–646, 2016.

[Nas 16] A. Naskos, Cloud elasticity: a survey, algorithmic aspects of cloud, Computing, p.151–167, 2016.

[Gal 16] G. Galante, L. Bona, A survey on cloud computing elasticity, in: IEEE Fifth International Conference on Utility and Cloud Computing, UCC, p.263–270, 2012.

[Her 13] N.Herbst, S. Kounev, R. Reussner, Elasticity in cloud computing: what is and what is not ?, in: Proceedings of 10th International Conference on Autonomic Computing, p.23–27, 2013.

[Car18] V.Cardellini, T.Galinac Grbac, M.Nardelli, N.Tanković, Truong HL. QoS- Based Elasticity for Service Chains in Distributed Edge Cloud Environments. In: Ganchev I., van der Mei R., van den Berg H. (eds) Autonomous Control for a Reliable Internet of Services. Lecture Notes in Computer Science, vol. 10768, 2018.

[Maa 18] Z. Maamar , T.Baker ,M.Sellami ,M.Asim ,E. Ugljanin , N. Faci, Cloud vs edge: Who serves the Internet-of-Things better?, Internet Technology letters, 2018.

[Mou 17] R. Moudjari, Z. Sahnoun Towards cloud of clouds elasticity management system, in: The 28th International Workshop on Database and Expert Systems Applications, DEXA, p.152–156, 2017.

[Mou 18] R. Moudjari, Z. Sahnoun, F.Belala, Towards a Fuzzy Bigraphical Multi Agent System for Cloud of Clouds Elasticity Management, International Journal of Approximate Reasoning, Vol.102, p.86-107, 2018.

[Par 16] F. Paraiso, P. Merle, L. Seinturier, SoCloud: a service-oriented component-based PaaS for managing portability, provisioning, elasticity, and high availability across multiple clouds , Computing 98, p. 539–565, 2016.

[Dup 15] Dupont, J. Lejeune, F. Alvares, T. Ledoux, Experimental analysis on autonomic strategies for cloud elasticity, in: International Conference on Cloud and Autonomic Computing, p.81–92, 2015.

[Moh 15] M. Mohamed, M. Amziani, D. Belaid, S. Tata, T. Melliti, An autonomic approach to manage elasticity of business processes in the cloud, Future Gener. Comput. Syst. Vol.50, p. 49–61, 2015.

[Jac 04] B. Jacob, A Practical Guide to the IBM Autonomic Computing Toolkit, IBM, International Technical Support Organization, 2004.

[Wey 13] D.Weyns, B.Schmerl, V.Grassi, S.Malek, R. Mirandola, C.Prehofer, J.Wuttke, J.Andersson, H.Giese, K. Göschka, On Patterns for Decentralized Control in Self-Adaptive Systems, Software Engineering for Self-Adaptive Systems II. Lecture Notes in Computer Science, vol 7475, 2013.