Fog and Cloud Computing Optimization in Mobile IoT Environments

José Carlos Ribeiro Vieira
josecarlosvieira@tecnico.ulisboa.pt

Instituto Superior Técnico, Universidade de Lisboa
Advisors: Prof. António Manuel Raminhos Cordeiro Grilo
Prof. João Coelho Garcia

Abstract. We introduce a xxxxx

Keywords: Cloud computing, fog computing, mobility, optimization, multi-objective

Table of Contents

1 Introduction . . . 3
2 Objectives . . . 5
3 Related Work . . . 5
  3.1 xxxx . . . 5
  3.2 xxxx . . . 5
  3.3 xxxx . . . 5
  3.4 xxxx . . . 5
4 Architecture . . . 5
5 Evaluation . . . 6
6 Schedule of Future Work . . . 6
7 Conclusion . . . 6

1 Introduction

Cloud computing is a computing technology that became popular at the beginning of the twenty-first century, which provides users online access to services, employing large groups of computers, servers, disks, and routers interlinked together in a distributed and complex manner. Cloud computing has been imperative in expanding the reach and capabilities of computing, storage, data management, and networking infrastructure to the applications. The key idea in this model is that clients outsource the allocation and management of resources (hardware or software) that they rely upon to the cloud. Clouds can provide different service models according to the users' requirements, like infrastructure as a service (IaaS), platform as a service (PaaS) and software as a service (SaaS). Since the demand for cloud resources will change over time, setting a fixed amount of resources results in either over- or under-provisioning, so cloud service providers (CSPs) afford dynamic resources for a scalable workload, applying a pay-as-you-go cost model where clients only pay for the amount of resources they actually use. Cloud computing brings many advantages to the end user applications like high availability, flexibility, scalability, reliability, to mention a few.

Although cloud computing has brought forth many advantages, it has certain limitations. Since cloud servers reside in remote data centers, end-to-end communication may have long delays (characteristic of multi-hops transmissions over the Internet), so the time required to access cloud-based services may not be suitable for some applications with ultra-low latency requirements (real-time). Augmented reality applications that use head-tracked systems, for example, require end-to-end latencies to be less than 16 ms [1]. Cloud-based virtual desktop applications require end-to-end latency below 60 ms if they are to match QoS of local execution [2]. Remotely rendered video conference, on the other hand, demand end-to-end latency below 150 ms [3].

Despite the fact that mobile devices have evolved radically in the last years, battery life, computation and storage capacity remain limited. This means that, heavy application executions must be offloaded to cloud servers, which then return processed results. The solution that has already been proposed is to bring the cloud closer to the end users, where entities such as base-stations would host smaller sized clouds. This idea has been variously termed as Cloudlets [4], Fog Computing [5], Edge Computing [6], and Follow Me Cloud [7], to name a few.

Fog computing is a new computing architecture introduced in 2012 by Bonomi et al. [5], that later, in 2015, big companies such as Cisco Systems, ARM Holdings, Dell, Intel, Microsoft, and Princeton University, founded the OpenFog Consortium, to promote interests and development in this field [8]. It aims to enable computing, storage, networking, and data management not only in the cloud, but also along the cloud-to-thing path as data traverses to the cloud (preferably close to the IoT devices). Fog nodes, also known as fog servers or cloudlets (smaller sized clouds with lower computational capacity), can be placed close to IoT source nodes, due to low hardware footprint and low power consumption (e.g., small servers, routers, switches, gateways, set-top boxes, access points). This allows latency to be much smaller, through geographical distribution, compared to traditional cloud computing. Nevertheless, cloud is still more suitable than fog for massive data processing. So even though, fog computing has been proposed to grant support for IoT applications, it does not replace the needs of cloud-based services. In fact, fog and cloud complement each other, and one cannot replace the need of the other. Together, they offer services even further optimized. Allowing enhanced capabilities for data aggregation, processing, and storage, where cloudlets are fundamental to both improving latencies and reducing network traffic to the cloud. Moreover, Internet connectivity is not essential for the fog-based services to work, what means that services can work independently and send necessary updates to the cloud whenever the connection is available [9].

Fog computing will be crucial in a diversity of scenarios. For instance, heterogeneous sensory nodes (e.g., sensors, controllers, actuators) on a self-driving vehicle, are estimated to generate about 1GB data per second [10]. As the number of features grow, the data deluge grows out of control. Moreover, these types of systems, where people's lives depends on it, are hard real-time what means that it is absolutely imperative that all deadlines are met. Offloading tasks to fog nodes will be the best solution, once a big effort in mobility support has been done through the migration of VMs using cloudlets [11]. Also, in this context, Puliafito et al. address three types of applications where fog is required, namely, citizen's healthcare, drones for smart urban surveillance and tourists as time travellers [12], addressing the needs of low latency and mobility support. On top of that, the number of mobile devices are predicted to reach 11.6 billion by 2021, where the subset of IoT ones are expected to be 929 million [13]. With more and more devices sending data to the cloud, leads to and increasing bandwidth usage, which is a bottleneck in terms of response time. To this end, fog computing is a good solution to prevent it.

Despite the benefits that fog promises to offer such as low latency, heterogeneity, scalability and mobility, the current model suffer from some limitations that still require more efforts to overcome them.

Fog servers are supposed to be closer to end devices so its location has to be ubiquitous. Consequently, less powerful than clouds due to the high deployment cost. If many IoT devices make requests to the same fog node at the same time, it will not have enough computational and storage power to give a prompt response. So it raises the question of should a service currently running in one fog node be migrated to another one, and if yes, where? While conceptually simple, it is challenging to make these decisions in an optimal manner. Offloading tasks to the next server appears to be the solution, however, migrate the related data and processing that initially was one-hop away to the end device to a multi-hop away server, will increase network distance. Consequently, increase the response delays and bandwidth usage by the intermediate links. Besides, this decision still has to take into account the cost for both the client, with respect to the time of migration, and the provider, in terms of energy. Ignoring some of these parameters can lead to make incorrect decisions, what will both violate acceptable latency QoS constraints user's application and damage or defeat the credibility of fog computing.

Another limitation, is that fog assumes that cloulets are always fixed. For instance, it does not foresee that a bus could have computational power. A bus could serve as a fog and provide support for the final devices from inside or outside it. The same can be applied to cars that nowadays are increasingly better with respect to computational power. Moreover, in a near future, unmanned aerial vehicles (UAVs) could be used for the same purpose for some social phenomena where a large number of people join in the same place as, for example, in sports games, manifestations, etc.

Summing up, we intend to tackle two of the current limitations which are, to the best of our knowledge, untreated problems in the literature. One is to provide support mobility in fog computing, not exclusively to the end devices but also to the fog nodes and the other is to achieve multi-objective fog system design.

This research will, therefore, focus on (...)

The remainder of the document is structured as follows. Section 2 xxx. Section 3 xxx.

Section 4 describes xxxx. Section 5 defines the xxx. Finally, Section 6 presents xxxx and Section 7 xxxx.

2 Objectives

This work has the main objective to present

3 Related Work

In this section we will give some contextual information about concepts and techniques that are relevant to our work.

3.1 xxxx

3.2 xxxx

3.3 xxxx

3.4 xxxx

4 Architecture

xxxxx

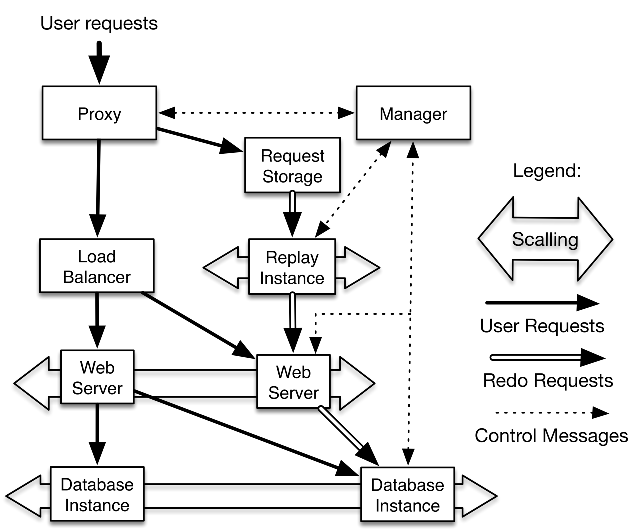

Fig. 1: Overview of the proposed service

xxxx

5 Evaluation

The evaluation of the proposed architecture will be done xxxx

6 Schedule of Future Work

Future work is scheduled as follows:

- xxxx
- xxxx

7 Conclusion

xxxxx

References

1. S. R. Ellis, K. Mania, B. D. Adelstein, and M. I. Hill, "Generalizeability of latency detection in a variety of virtual environments," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 48, no. 23. SAGE Publications Sage CA: Los Angeles, CA, 2004, pp. 2632–2636.
2. B. Taylor, Y. Abe, A. Dey, M. Satyanarayanan, D. Siewiorek, and A. Smailagic, "Virtual machines for remote computing: Measuring the user experience," *Carnegie Mellon University*, 2015.
3. T. Szigeti and C. Hattingh, *End-to-end qos network design.* Cisco press, 2005.
4. M. Satyanarayanan, "Cloudlets: at the leading edge of cloud-mobile convergence," in *Proceedings of the 9th international ACM Sigsoft conference on Quality of software architectures.* ACM, 2013, pp. 1–2.
5. F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the internet of things," in *Proceedings of the first edition of the MCC workshop on Mobile cloud computing.* ACM, 2012, pp. 13–16.
6. S. Davy, J. Famaey, J. Serrat-Fernandez, J. L. Gorricho, A. Miron, M. Dramitinos, P. M. Neves, S. Latré, and E. Goshen, "Challenges to support edge-as-a-service," *IEEE Communications Magazine*, vol. 52, no. 1, pp. 132–139, 2014.
7. T. Taleb and A. Ksentini, "Follow me cloud: interworking federated clouds and distributed mobile networks," *IEEE Network*, vol. 27, no. 5, pp. 12–19, 2013.
8. J. MSV, "Is fog computing the next big thing in internet of things?" Apr 2016. [Online]. Available: https://www.forbes.com/sites/janakirammsv/2016/04/18/is-fog-computing-the-next-big-thing-in-internet-of-things/#6bc7b2c4608d
9. A. Yousefpour, C. Fung, T. Nguyen, K. Kadiyala, F. Jalali, A. Niakanlahiji, J. Kong, and J. P. Jue, "All one needs to know about fog computing and related edge computing paradigms: A complete survey," *arXiv preprint arXiv:1808.05283*, 2018.
10. A. Angelica, "Google's self-driving car gathers nearly 1gb/sec," 2013. [Online]. Available: http://www.kurzweilai.net/googles-self-driving-car-gathers-nearly-1-gbsec
11. M. M. Lopes, W. A. Higashino, M. A. Capretz, and L. F. Bittencourt, "Myifogsim: A simulator for virtual machine migration in fog computing," in *Companion Proceedings of the10th International Conference on Utility and Cloud Computing.* ACM, 2017, pp. 47–52.
12. C. Puliafito, E. Mingozzi, and G. Anastasi, "Fog computing for the internet of mobile things: issues and challenges," in *Smart Computing (SMARTCOMP), 2017 IEEE International Conference on.* IEEE, 2017, pp. 1–6.

13. “Cisco visual networking index: Global mobile data traffic forecast update, 2016–2021 white paper,” Mar 2017. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html