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

There's no doubt that the Internet of Things (IoT) is a great resource. It comprises *things* that have unique identities and are connected to the Internet (e.g., vehicles, home appliances, wearable devices). The number of mobile devices are predicted to reach 11.6 billion by 2021, exceeding the world's projected population at that time (7.8 billion), where the subset of IoT ones are expected to be 929 million [1].

Although this kind of device has evolved radically in the last years, battery life, computation and storage capacity remain limited. This means that they are not suitable for running heavy applications, being necessary, in this case, to resort to third parties.

Cloud computing has been imperative in expanding the reach and capabilities of IoT devices. It enables that clients outsource the allocation and management of resources (hardware or software) that they rely upon to the cloud. In addition, to avoid over- or under-provisioning, cloud service providers also afford dynamic resources for a scalable workload, applying a pay-as-you-go cost model. This way, besides overcoming the aforementioned limitations, it also brings other advantages such as availability, flexibility, scalability, reliability, to mention a few.

Despite the benefits of using cloud computing, two main problems, linked to IoT applications, remain unresolved and they can not be underestimated. The first, and the most obvious, is the fact that cloud servers reside in remote data centers and, consequently, the end-to-end communication have long delays (characteristic of multi-hops transmissions over the Internet). Some applications, with ultra-low latency requirements, can't support such delays. Augmented reality applications that use head-tracked systems, for example, require end-to-end latencies to be less than 16 ms [2]. Cloud-based virtual desktop applications require end-to-end latency below 60 ms if they are to match QoS of local execution [3]. Remotely rendered video conference, on the other hand, demand end-to-end latency below 150 ms [4]. The other problem is related with the constantly growing number of IoT devices. In a sense-process-actuate model, where the processing is done in the cloud, core network traffic (i.e. bandwidth usage) grows depending on the number of IoT devices.

To overcome this limitations, the solution that has already been proposed is to bring the cloud closer to the end devices, where entities such as base-stations would host smaller sized clouds. This idea has been variously termed as Cloudlets [5], Fog Computing [6], Edge Computing [7], and Follow Me Cloud [8], to name a few.

Fog computing is a new computing architecture that 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. Fog nodes, also known as fog servers or cloudlets (smaller sized clouds with lower computational capacity), are Virtual-Machine (VM) based, which means that they promote flexibility and mobility. They can be placed close to IoT source nodes, due to low hardware footprint and low power consumption. This allows latency to be much smaller, through geographical distribution, compared to traditional cloud computing and still cut off a significant amount of core network traffic. Nevertheless, cloud is still more suitable than fog for massive data processing, when the latency constraints are not so tight. 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 to IoT applications. 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].

1.1 Motivation

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 efforts to overcome them.

There is lack of support for mobile fog computing. Most of the existing literature assumes that the fog nodes are fixed, or only considers the mobility of IoT devices [9]. Less attention has been paid to mobile fog computing and how it can improve the QoS, cost, and energy consumption. For instance, it does not foresee that a bus could have computational power; as a cloudlet, it cloud provide offloading support to elements (i.e. IoT devices and other cloudlets) inside and outside it. The same could be applied to cars that are nowadays getting increasingly better in terms of computational power. Both would be extremely useful in order to enhance the resources and capabilities of fog computing, for example, in large urban areas where traffic congestion is common or when they are parked. Apart from QoS enhancement, this would reduce the implementation costs since it would no longer require such computational power in the roadside cloudlets. Also, the costs in terms of latency to the client would be minimized. On top of that, it would minimize the energy consumption of mobile devices once the access points where they are connected may be even closer.

Another limitation of fog computing is to take into account few parameters in the decision making of migration. Most of the existing schemes that are proposed for fog systems, such as offloading, load balancing, or service provisioning, only consider few objectives (e.g., QoS, cost) and assume other objectives do not affect the problem [9]. Fog servers are less powerful than clouds due to the high deployment cost. If many requests are made 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: 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 seems to be the solution, however, migrate the VM that was initially one-hop away from the IoT device to a multi-hop away server, will increase the network distance. Consequently, raises the end-to-end latency and the bandwidth usage by the intermediate links. Besides, this decision still has to take into account the cost for both the client (e.g., migration time, computational delay) and the provider (e.g., computing and migration energy). Ignoring some of these parameters can lead to wrong decisions, what will both violate latency constraints of user's application and damage or defeat the credibility of fog computing.

1.2 Objectives

Summing up, this work 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 mobility support in fog computing environments, not exclusively to the end devices but also to the fog nodes and the other is to achieve multi-objective fog system design. These objectives shall be implemented in a toolkit allowing the simulation of resource management techniques in IoT, and mobile fog computing environments. In order to achieve the aforementioned goals, this work must follow the sequence of tasks presented below.

- Start with analysis of current mobility approaches publicly available with respect to IoT nodes or cloudlets;
- Propose mobile fog computing, where fog nodes can move, provisioning a method to keep the service always-available for IoT nodes;

- Analyze the most suitable optimization algorithms for fog computing and select the most appropriate ones;
- Design of mobility-aware task offloading when fog nodes are mobile, taking into account the related costs to the client (i.e. migration, communication and processing delay) and discovery;
- Propose a variation to achieve multi-objective (i.e. QoS, QoE, cost, handover, load balancing, energy, bandwidth and VMs or virtual objects migration);
- Study the current open source simulators for fog computing environments;
- Implement the proposed algorithms in the simulator and compare them.

The remainder of the document is structured as follows. Section 2 xxx. Section 3 describes xxxx. Section 4 defines the xxx. Finally, Section 5 presents xxxx and Section 6 xxxx.

2 Related Work

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

3 Architecture

xxxxx

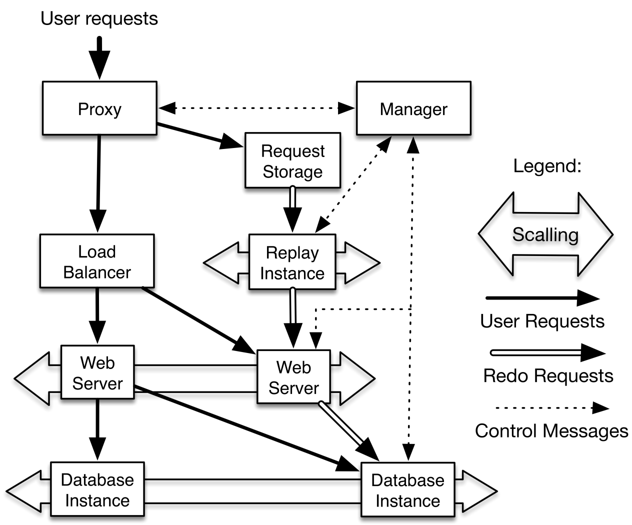

Fig. 1: Overview of the proposed service

xxxx

4 Evaluation

The evaluation of the proposed architecture will be done xxxx

5 Schedule of Future Work

Future work is scheduled as follows:

- xxxx
- xxxx

6 Conclusion

xxxxx

References

1. "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
2. 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.
3. 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.
4. T. Szigeti and C. Hattingh, *End-to-end qos network design.* Cisco press, 2005.
5. 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.
6. 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.
7. 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.
8. 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.
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.