Edge Computing via Dynamic In-network Processing

Julien Gedeon
Telecooperation Lab, Technische Universität Darmstadt
Email: gedeon@tk.tu-darmstadt.de

***Abstract*—With the widespread adoption of smart devices and the rapid evolvement of the Internet of Things, the traditional Cloud Computing paradigm is more and more being shifted towards performing data processing close to the network edge, thus reducing latency and traffic in the core network while operating in an energy-efficient way, as to prolong the lifetime of those smart devices. In this paper, we consider enabling edge computing through the concept of in-network processing, where we envision a world with decentralized, opportunistic in-network cloud infrastructures, located at the user's vicinity to provide agile, low-latency computing and storage capabilities. Based on the key challenges of this new approach, we outline five research questions we plan to addres.**

I. Introduction

Today's cloud computing infrastructures offer vast computational resources via an affordable pay-as-you-go model. However, this approach falls short for emerging latency-critical applications (e.g., real-time event detection or augmented reality). Pushing all computations into the cloud makes the core network susceptible to congestion and thus creates bottlenecks in the entire network infrastructure. In addition, cloud-based applications are data-driven instead of user-driven and provide no context awareness. Furthermore, there is limited or no support for mobility. The increasing prevalence of resource- and battery-constrained mobile devices such as smartphones and the emerging IoT [1], which is predicted to grow to billions of devices, will further feed the need to have more powerful devices nearby to offload data and process it. Each of these devices will be delivering massive amounts of data to be used in real-time analytics, event detection or complex event processing. An especially useful and palpable usage context can be found in the vision of *Smart Cities* [2], where urban areas are augmented to provide services to its citizen.

The new paradigm of *In-network Processing* [3] moves the Cloud capabilities closer to the networks' edge and tries to overcome these drawbacks. Other terms are often used to describe similar ideas, such as *Fog Computing* [4], [5] or *Edge Computing* [6], [7]. One possible concept to use is the notion of *Cloudlets* [8], which are micro clouds located at the edge of the network. Cloudlets therefore can run on a variety of devices, including the ones with constrained resources such as network routers. In the next chapters, we will highlight how this new approach differs from existing cloud computing infrastructures and identify the major research challenges we plan to address.

Table I
Comparison between Cloud Computing and In-network Clouds

	Cloud Computing	In-network Clouds
Proximity	low	high
Latency	high	low
Geo-distribution	locally clustered	widespread
Infrastructure	centralized datacenters	decentralized cloudlets
Heterogeneity	low	high
Deployment	fixed, static	dynamic, opportunistic
Availability	high	varying
Connections	long-thin	short-fat
Mobility support	limited	yes
Context Awareness	no	yes
Applications	data-driven	user-driven
User control	low	high

II. The next era of Cloud Computing

We envision a future where we are surrounded by these new kinds of in-network micro-clouds (i.e., cloudlets), that we can leverage to perform real-time processing, analytics and storage at the edge - something currently not possible in today's cloud computing environments [9]. Tomorrow's in-network cloud infrastructure will be located in the network or at the edge between the distant cloud and the mobile end user. Opportunistically available devices will be present in every public and private space to provide services to users. The cloud therefore becomes more user-centric and as users move, so does their personal cloud. In many ways, this new paradigm is orthogonal to the cloud computing infrastructure of today. Table I highlights some of these differences.

In combination with lightweight virtualization techniques such as containers, cloudlets will be able to dynamically instantiate applications and services for nearby users. The deployment of cloudlets also enables new business opportunities for network operators, service providers, as well as end users, who can rent out part of their devices' capabilities for processing, storage, and communication.

III. Research Challenges

The newly introduced concept brings along new challenges and opportunities. In the following, we describe in particular five research questions we plan to address.

Leveraging existing devices. Many devices today serve one particular purpose, yet their overall resource utilization is often low. Looking at the example of home routers and network middleboxes, existing potential often remains untapped. We therefore plan to examine ways to upgrade these devices so that they can provide general-purpose computation capabilities. In doing so, care must be taken to not impede the normal functioning of these devices. Closely related techniques we could employ are virtualization, micro operating systems and software defined networking. We will also study what kind of contribution can be expected from leveraging these devices, e.g. by analyzing router coverage in a typical city.

Enabling user mobility. In-network processing has to take into account user mobility. A user might only be connected to an edge gateway for a short period of time and wireless connectivity is often intermittent. Therefore, to enable permanent access to storage and computation at the edge, we need to develop handover and replication mechanisms when users are on the move. For instance, the results of a computation might be available on a cloudlet at a time when the user is no longer connected to it. In a mobile cloudlet architecture, the desired data and computations move with the user to a new, cloudlet-enabled location. This requires mechanisms for task migration and result delivery across different in-network clouds.

Operator placement and adaptation. To enable new applications that run inside the network, processing components should be placed dynamically on nodes in the network that act as cloudlets. Naturally, the question arises on where to place those processing components. Even in the most basic version, this is a variant of the well-known task assignment problem, which is NP-hard. Several approaches [10], [11], [12] have already addressed the issue of placing operators. Most of these approaches are either targeted at specific applications, operate in the context of huge data centers, or do not take into account the dynamics of the network or tasks. In our context of in-network clouds, we are faced with a multitude of operators that are required to carry out tasks and have different requirements and priorities. Besides changes in the network underlay, the tasks and operators are also subject to changes. Taking into account those factors, an optimal solution cannot be computed in polynomial time, and therefore, efficient heuristics will be required. Before executing changes and migrating operators we also need to take into account the costs that can occur when doing so.

Analyzing and adapting communication patterns. In a dynamic in-network processing scenario, nodes in the network may communicate using different communication patterns, such as a direct point-to-point connection or via publish/subscribe overlays. These patterns are important when relaying data between operators or reuse it for multiple tasks in the network. Contrary to the semantics of most publish/subscribe systems, a node might also have the choice of where to publish a message, i.e. select one recipient among many. This can be used as a mechanism to balance load in the network. We plan to analyze which patterns are most beneficial in a given scenario and propose a structure to dynamically switch between communication patterns. We also need loose coordination mechanisms for publishers and subscribers of data, something not implemented in today's publish/subscribe systems that mostly rely on a centralized broker. We also plan to analyze how this problem and the one of operator placement influence each other.

Management and orchestration of resources. With a large number of independent, decentralized cloudlets, we are faced with the problem of how to manage and orchestrate them. The objectives are to provide load balancing, energy efficiency, and a reduction of the operational costs. A management and orchestration platform should provide mechanisms for those independent cloudlets to collaborate, exchange context and state information in order to have more valuable information in the network. Management functionalities are also required to adapt the placement and enable the colocation of operators.

IV. Conclusion

In this paper, we outlined our vision of a new paradigm for edge computing that shifts power from distant, centralized clouds to local, in-network processing architectures using cloudlets, thus providing low-latency access to users for processing and storage. Based on this vision, we defined five research questions we would like to address in the future.

Acknowledgement

This work has been cofunded by the German Federal Ministry for Education and Research (BMBF, Software Campus project DynamicINP, grant no. 01IS12054).

References

[1] E. Borgia, "The internet of things vision: Key features, applications and open issues," *Computer Communications*, vol. 54, pp. 1–31, 2014.
[2] J. Jin, J. Gubbi, S. Marusic, and M. Palaniswami, "An Information Framework for Creating a Smart City Through Internet of Things," *IEEE Internet of Things Journal*, vol. 1, no. 2, pp. 112–121, 2014.
[3] R. Stoenescu, V. Olteanu, M. Popovici, M. Ahmed, J. Martins, R. Bifulco, F. Manco, F. Huici, G. Smaragdakis, M. Handley, and C. Raiciu, "In-net: In-network processing for the masses," in *Proceedings of the Tenth European Conference on Computer Systems*, ser. EuroSys '15. ACM, 2015, pp. 23:1–23:15.
[4] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog Computing and Its Role in the Internet of Things," *Proceedings of the first edition of the MCC workshop on Mobile cloud computing*, pp. 13–16, 2012.
[5] S. Yi, C. Li, and Q. Li, "A Survey of Fog Computing: Concepts, Applications and Issues," *Proceedings of the 2015 Workshop on Mobile Big Data - Mobidata '15*, pp. 37–42, 2015.
[6] A. Chandra, J. Weissman, and B. Heintz, "Decentralized edge clouds," *IEEE Internet Computing*, vol. 17, no. 5, pp. 70–73, 2013.
[7] H. Chang, A. Hari, S. Mukherjee, and T. V. Lakshman, "Bringing the cloud to the edge," *Proceedings - IEEE INFOCOM*, pp. 346–351, 2014.
[8] M. Satyanarayanan, P. Bahl, R. Cáceres, and N. Davies, "The case for VM-based cloudlets in mobile computing," *IEEE Pervasive Computing*, vol. 8, no. 4, pp. 14–23, 2009.
[9] J. Mineraud, O. Mazhelis, X. Su, and S. Tarkoma, "A gap analysis of internet-of-things platforms," *Computer Communications*, 2016.
[10] P. Pietzuch, J. Ledlie, J. Shneidman, M. Roussopoulos, M. Welsh, and M. Seltzer, "Network-aware operator placement for stream-processing systems," in *Proceedings of the 22Nd International Conference on Data Engineering*, ser. ICDE '06. IEEE Computer Society, 2006, pp. 49–.
[11] G. T. Lakshmanan, Y. Li, and R. Strom, "Placement strategies for internet-scale data stream systems," *IEEE Internet Computing*, vol. 12, no. 6, pp. 50–60, 2008.
[12] V. Cardellini, V. Grassi, F. Lo Presti, and M. Nardelli, "Optimal operator placement for distributed stream processing applications," *Proceedings of the 10th ACM International Conference on Distributed and Event-based Systems - DEBS '16*, pp. 69–80, 2016.