Fog Computing Orchestration based on Network Latency

Stojan Kitanov[1] and Toni Janevski[2]
[1]Faculty of Information Systems Multimedia and Animation at University of Information Science and Technology "St. Paul the Apostle"
stojan.kitanov@uist.edu.mk

[2]Faculty of Electrical Engineering and Information Technologies
Ss Cyril and Methodius University of Skopje
Macedonia
tonij@feit.ukim.edu.mk

ABSTRACT: *IoT is found to yield many applications out of which the one that help address the real-time and low latency requirements. This kind of needs develop new challenges while offering service to users. The requirements are currently filled by the present cloud computing applications. The new domain Fog computing can able to address the requirements for the hybrid environment service orchestration. The network edge bring the cloud computing provides more geographic coverage, low latency and higher load balancing. The current developments support the orchestration within the large as well as complex environments and ensure reliable services. This work asses the Hybrid Environment Service Orchestrator (HESO) for resilient and trustworthy Fog Computing services in terms of network latency. This process is aimed to address the real-time big data analysis, 5G networks and IoT.*

Keywords: Cloud, Cloud Computing, Fog, Fog Computing, Mobile Cloud, Mobile Cloud Computing

Received: 5 July 2021, Revised 30 August 2021, Accepted 7 September 2021

DOI: 10.6025/jmpt/2021/12/4/125-131

1. Introduction

The future Internet of Services (IoSs) will become the linkage between extremely complex networked organizations (e.g. telecoms, transportation, financial, health and government services, commodities, etc.), that will provide the basic ICT infrastructure that supports the business processes and the activities of the whole society in general [1]. Frequently, these processes and activities will be supported by orchestrated cloud services, where a number of services work together to achieve a business objective. However, future Internet will exacerbate the need for improved QoS/QoE, supported by services that are orchestrated on-demand and are capable of adapt at runtime, depending on the contextual conditions, to allow reduced latency, high mobility, high scalability, and real time execution. The emerging wave of Internet of Things (IoTs) would require seamless mobility support and geo-distribution in addition to location awareness and low latency. These demands can be only partially fulfilled by existing cloud computing solutions [2].

A new paradigm called Fog Computing, or briefly Fog has emerged to meet these requirements [3]. Fog extends cloud computing and services to the edge of the network. It provides data, computing, storage, and application services to end-users that can be hosted at the network edge or even end devices such as set-top-boxes or access points. The main features of Fog are its proximity to end-users, its dense geographical distribution, and its support for mobility. Fog will combine the study of mobile communications, micro-clouds, distributed systems, and consumer big data. It is a scenario where a huge number of heterogeneous (wireless and sometimes autonomous) ubiquitous and decentralized devices communicate and potentially cooperate among them and with the network to perform storage and processing tasks without the intervention of third parties [4]. These tasks support basic network functions or new services and applications that run in a sand-boxed environment. Users leasing part of their devices to host these services get incentives for doing so.

By deploying reserved compute and communication resources at the edge, Fog computing absorbs the intensive mobile traffic using local fast-rate connections and relieves the long back and forth data transmissions among cloud and mobile devices [5], [6]. This significantly reduces the service latency and improves the service quality perceived by mobile users and, more importantly, greatly saves both the bandwidth cost and energy consumptions inside the Internet backbone. Fog computing represents a scalable, sustainable and efficient solution to enable the convergence of cloud-based Internet and the mobile computing. Therefore Fog paradigm is well positioned for real time big data analytics, 5G network, and IoT.

The move from Cloud to Fog computing brings out several key challenges, including the need for supporting the ondemand orchestration and runtime adaptation of resilient and trustworthy Fog Services. This is essential for the success of the future Internet of Everything (IoE), which a clear evolution of the IoT [7].

This paper provides a model of Hybrid Environment Service Orchestrator (HESO) for resilient and trustworthy Fog Computing services. It is organized as follows. Section II provides an overview of Fog Computing. Section III proposes the HESO model for Fog Computing. Section IV evaluates the HESO model in terms of Round Trip Time (RTT) latency. Finally, Section V concludes the paper and provides future work research directions.

2. Overview of Fog Computing

An overview of three layered Fog Computing architecture is given in Figure 1. The intermediate Fog layer consists of

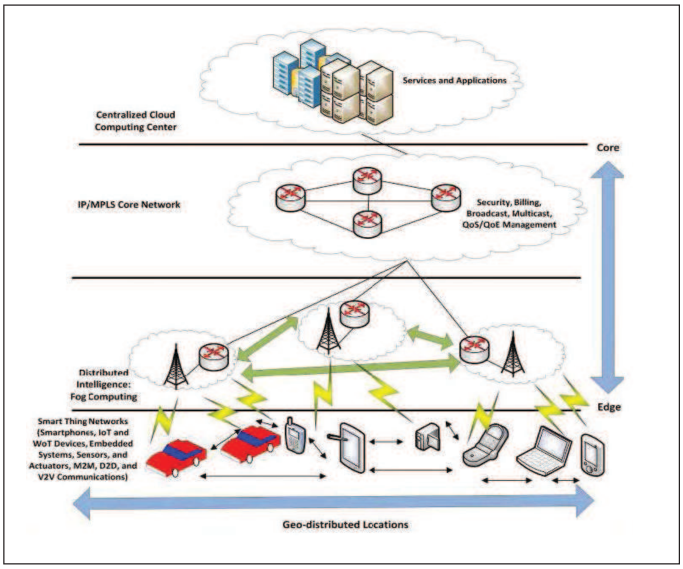


Figure 1. Fog Computing Architecture

geodistributed intelligent Fog Computing servers which are deployed at the edge of networks, e.g., parks, bus terminals, shopping centers, etc. Each Fog server is a highly virtualized computing system and is equipped with the on-board large volume data storage, compute and wireless communication facility [5].

	Fog Computing	**Cloud Computing**
Target Type	Mobile Users	General Internet Users
Service Type	Limited localized information services related to specific deployment locations	Global information collected from worldwide
Hardware	Limited storage, compute power and wireless interface	Ample and scalable storage space and compute power
Distance to Users	In the physical proximity and communicate through single-hop wireless connection	Faraway from users and communicate through IP networks
Working Environment	Outdoor (streets, parklands, etc.) or indoor (restaurants, shopping malls, etc.)	Warehouse-size building with air conditioning systems
Deployment	Centralized or distributed in reginal areas by local business (local telecommunication vendor, shopping mall retailer, etc.)	Centralized and maintained by Amazon, Google, etc.

Table 1. A Comparison between Fog and Cloud

The role of Fog servers is to bridge the smart mobile device things and the cloud. Each smart thing device is attached to one of Fog servers that could be interconnected and each of them is linked to the cloud [6].

The geo-distributed intelligent Fog servers directly communicate with the mobile users through single-hop wireless connections using the off-the-shelf wireless interfaces, such as, LTE, WiFi, Bluetooth, etc. They can independently provide predefined service applications to mobile users without assistances from cloud or Internet. In addition, the Fog servers are connected to the cloud in order to leverage the rich functions and application tools of the cloud.

The existence of Fog will be enabled by the emerging trends on technology usage patterns on the one side, and the advances on enabling technologies on the other side. A comparison between Fog Computing and Cloud Computing is given in [5], and it is summarized in Table 1.

The cloud in 5G network and beyond will be diffused among the client devices often with mobility too, i.e. the cloud will become fog. More and more virtual network functionality will be executed in a fog computing environment, and that will provide mobiquitous service to the users. This will enable new services paradigms such as Anything as a Service (AaaS) where devices, terminals, machines, and also smart things and robots will become innovative tools that will produce and use applications, services and data.

3. Service Orchestration with Fog

The move from cloud to fog brings out several key challenges. This includes the need for supporting the ondemand orches tration and runtime adaptation of resilient and trustworthy fog Services, which is essential for the success of the future IoE, a clear evolution of the IoT.

Traditional service orchestration approaches that have been applied to Cloud services are not adequate to the forthcoming large-scale and dynamic Fog Services, since they cannot effectively cope with reduced latency, high mobility, high scalability, and real time execution. Therefore a new Hybrid Environment Services Orchestrator (HESO) is needed, that will be capable of ensuring the resilience and trustworthiness of open, large scale, dynamic services on the Fog. The Orchestrator will be responsible for the composition of Service Elements available in the Fog environment (e.g. sensing, connectivity, storage, processing, platform services, and software services) into more complex Fog Services (e.g. traffic crowd sensing and trip planning services) to be offered to the users in the Fog environment.

The execution of the Fog Services may involve multiple different components and entities spread in a wide area, increasing the complexity in terms of the decision making process in what regards the resource allocation to achieve acceptable QoS/ QoE levels. To coordinate the execution of the Fog services, Orchestration mechanisms need to synchronize and combine the operation of the different service elements in order to meet the specifications of the composed Fog services, including low latency, scalability and resilience.

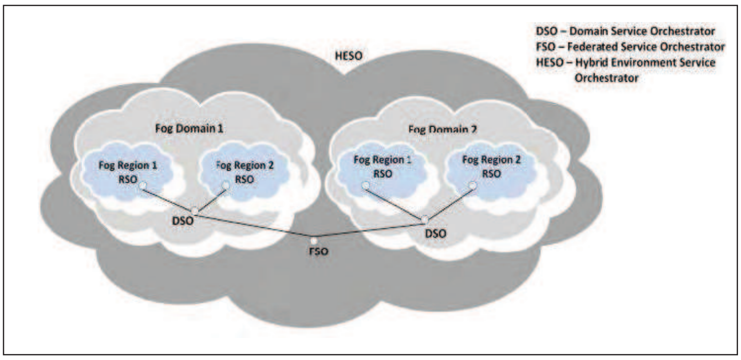


Figure 2. Hybrid Environment Service Orchestrator Model for Fog Computing

The architectural levels of Fog orchestrated services and mechanisms are given in Fig. 2. The HESO in Fog should operate in a loosely coupled mode, resulting in a solution with several levels: Regional Service Orchestrator (RSO), Domain Service Orchestrator (DSO) and Federated Service Orchestrator (FSO).

The RSOs are located at the edges of the Fog environment and they enable semi-autonomous operation of the different Fog Regions. This allows the distribution of the load which provides scalability and a much higher proximity to the end users. Therefore lower latencies can be achieved.

The DSOs is responsible for the Fog domains and supervises the RSOs below. This level will support federation mechanisms to enable intra-domain cooperation between different regions within one domain.

The FSO allows a fruitful interaction between different Fog domains. It is responsible for the management between different Fog domains and, similarly to the DSOs, it should be properly adapted to operate in a federate Cloud environment.

The FSOs will support federation mechanisms to enable cooperation among different Fog Domains (e.g. belonging to different entities or under the administration of different authorities) and the creation of a Multi-Domain Fog Environment able to support service ubiquity.

4. Evaluation of the Hesomodel

The evaluation of the HESO will be explored in terms of the Round Trip Time (RTT) latency. RTT latency is the time it takes for a single data transaction to occur, meaning the time it takes for the packet of data to travel to and from the source to the destination, and back to the source [10]. In the real networks, latency is measured by performing ping tests, that uses ICMP packets. The total size of each ICMP packet is 74 bytes with the headers.

Let the mobile user be located in the Fog Region 1, which s controlled by the Fog Domain 1. And let Fog Domain 1 through a Federated Service Orchestrator be connected with Fog Domain 2. Fog Region 1 may correspond to an LTE/LTE-Advanced Cloud Radio Access Network (CRAN). Fog Domain 1 may correspond to a Cloud Computing Centre in the same region with the CRAN Network, and Fog Domain 2 may correspond to a Cloud Computing Centre in a different region with the CRAN network. Let us assume the mobile user wants to upload and download some file for example a map, movie or similar, or wants to process some data. The RTT latency time required to perform this transaction is equal to:

$$RTT = \frac{T}{R_{UL}} + \frac{T}{R_{DL}} + i_1 t_1 + i_2 t_2$$

Here, T is the packet size travel from the source to the destination, and back to the source, R_{UL} and R_{DL} are the corresponding uplink and downlink data rates of LTE/LTE-A Network. The values of the uplink and downlink data rates [11] vary from the distance d between the mobile user and the CRAN network, and are summarized in Table 2. The binary information coefficients i_1 and i_2 point the location of the data for which the end user is interested. Table 3 summarizes the possible location of the file, as well as the possible values of i_1 and i_2. Finally t_1 and t_2 represent the time for the data file to be received by the Fog Region 1 (LTE Network) from Fog Domain 1 (Cloud Computing Center) or from the Fog Domain 2 (Cloud Computing Center), respectively.

Maximum Downlink Data Rate in Mbps	Maximum Uplink Data Rate in Mbps	Distance between the End user and the LTE-A eNodeB in meters
3000	1500	d < 500
300	100	500 ≤ d < 1000
150	50	1000 ≤ d <1500
50	25	1500 ≤ d

Table 2. Data Rates of an LTE-A/LTE Network

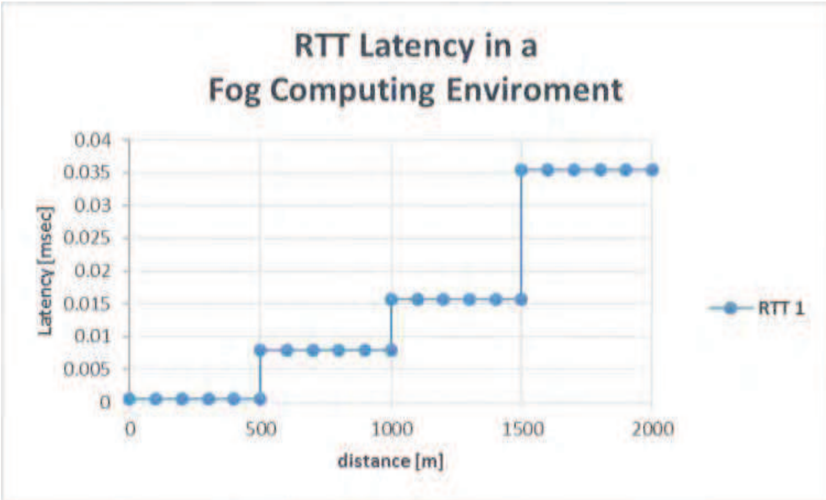
Location of Data File	i_1	i_2
Fog Region 1	0	0
Fog Domain 1	1	0
Fog Domain 2	0	1

Table 3. Possible Values for the Binary Information Coefficients

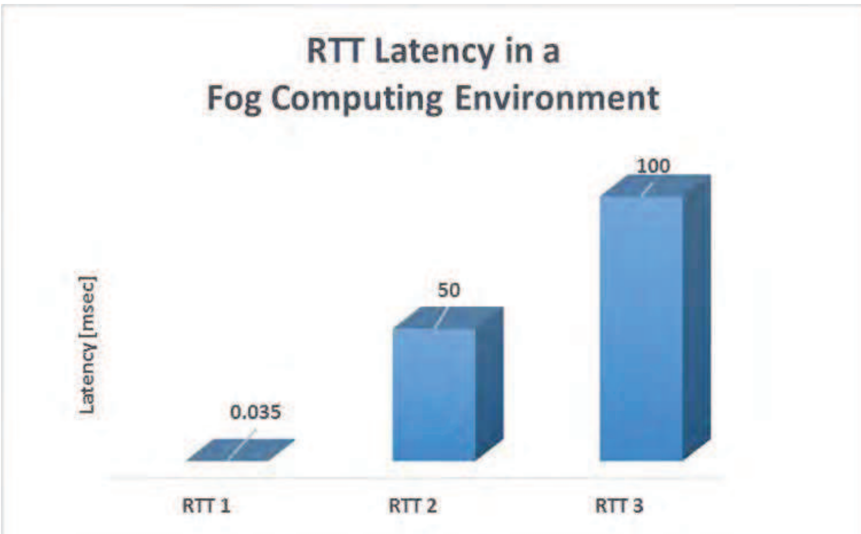
The simulation results are given in Figure 3. The size of a data file is 74 MB, and the values of t_1 and t_2 correspond to 50 msec and 100 msec, respectively. Here RTT1 represent the network latency when data file requested by the mobile user is located in CRAN network (Fog Region 1). RTT2 represent the network latency when data file requested by the mobile user is located in the cloud computing center (Fog Domain 1), which is in the same region with the CRAN network. RTT3 represent the network latency when data file requested by the mobile user is located in the cloud computing center (Fog Domain 2) which is a different region with CRAN network.

Figure 3a shows that the RTT latency increases as the mobile ser moves away from the eNodeB of the CRAN network. Fig. 3b compares the RTT lantencies depending whether the data file requested by the mobile user is located in Fog Region 1, Fog Domain 1, or Fog Domain 2. The lowest network latency is obtained if the data file is located in the CRAN (Fog Region 1). The highest latency is obtained if the file requested by the user is located in the Fog Domain 2, i.e. in a cloud computing center which is in a different region with the CRAN network.

The results demonstrate that latency is significantly reduced from the order of miliseconds to the order of microseconds, which is very important for the delay sensitive applications.



(a)



(b)

Figure 3. RTT Latency in the Fog with a HESO Model

5. Conclusion and Future Work

The purpose of Fog computing is to place a handful of computing, storage and communication resources in the proximity of mobile users, and therefore to serve mobile users with the local short-distance high-rate connections. The move from cloud to fog brings out several key challenges, including the need for supporting the on-demand orchestration and runtime adaptation of resilient and trustworthy fog services, which is essential for the success of the future IoE, a clear evolution of the IoT. This could be solved by the proposed Hybrid Environment Service Orchestrator for resilient and trustworthy Fog Computing services.

The results demonstrate that the latency can be significantly reduced to the order of microseconds by using the HESO model for the Fog Computing, which is very important for the delay sensitive applications. Therefore HESO model with Fog Computing is well positioned for real time big data analytics, 5G network and IoT.

Fog will act as a nervous system of the digital society, economy and everyday people's life. Fog paradigm is well positioned for real time big data analytics, 5G network, and IoT. The cloud in 5G networks and beyond will be diffused among the client devices often with mobility too, i.e. the cloud will become fog. More and more virtual network functionality will be executed in a fog computing environment, and it will provide mobiquitous service to the users. This will enable new AaaS service paradigms, where devices, terminals, machines, and also smart things and robots will become innovative tools that will produce and use applications, services and data. However there are also some aspects that should be addressed in order the Fog approach to be successful. This includes defining hybrid and heterogeneous environments, interaction and integration between the execution managements of each domain, and integration between managements inside one domain. In future we plan to work on solving some of these challenges.

References

[1] Horizon project: A New Horizon to the Internet, 2015, http://www.gta.ufrj.br/horizon/

[2] Zhang, S., Zhang, S., Chen, X., Huo, X. (2010). Cloud Computing research and Development trend, *In*: Proceedings of the 2010 Second International Conference on Future Networks (ICFN '10). *IEEE Computer Society*, Washington, DC, USA, 93- 97.

[3] Bonomi, F., Milito, R., Zhu, J., Addepalli, S. (2012). Fog Computing and its Role in the Internet of Things, *In*: Proceedings of the first edition of the MCC workshop on Mobile Cloud Computing (MCC 2012), ISBN: 978-1-4503-1519-7, doi:10.1145/2342509.2342513, ACM, New York, NY, USA, 13-16.

[4] Vaquero, L. M., Rodero-Merino, L. (2014). Finding your Way in the Fog: Towards a Comprehensive Definition of Fog Computing, *ACM SIGCOMM Computer Communication Review*, 44 (5), doi:10.1145/2677046.2677052, 27-32.

[5] Luan, H. T., Gao, L., Li, Z., Sun, L. X. Y. (2015). Fog Computing: Focusing on Mobile Users at the Edge, arXiv:1502.01815[cs.NI], 2015.

[6] Stojmenovic, I., Wen, S. (2014). The Fog Computing Paradigm: Scenarios and security issues, *In*: Proceedings of the Federated Conference on Computer Science and Information Systems (FedCSIS), ACSIS, 2 (5) doi:10.15439/2014F503, 1-8.

[7] Brech, B., Jamison, J., Shao, L., Wightwick, G. The Interconnecting of Everything. IBM Redbook, http://www.redbooks.ibm.com/redpapers/pdfs/redp4975.pdf.

[8] Brown, G. (2013). Converging Telecom and IT in the LTE RAN, White Paper at Heavy Reading on behalf of Samsung.

[9] Stolfo, J. S., Salem, B. M., Keromytis, D. A. (2012). Fog Computing: Mitigating Insider Data Theft Attacks in the cloud Proceeding of IEEE Symposium on Security and Privacy Workshops (SPW), ISBN: 978-0-7695-4740-4, doi: 10.1109/SPW.2012.19. *IEEE Computer Society, Washington*, DC, USA, 125-128.

[10] Latency Consideration in LTE Implications to Security gateway. (2014). Stoke Inc. White Paper, Literature No. 130-0029-001.

[11] Rosseler, A., Schlienz, J., Merkel, S., Kottkamp, M. (2014). LTE Advanced (3GPP Rel. 12) Technology Introduction, Rhode and Schwarz White Paper 6.2014 - 1MA252 2E.