Multimedia Microservice Placement in Hierarchical Multi-tier Cloud-to-Fog Networks

Fillipe Santos
Institute of Computing (IC)
University of Campinas
Campinas, Brazil
fillipe@lrc.ic.unicamp.br

Roger Immich
Metropolis Digital Institute (IMD)
Federal University of Rio Grande do Norte
Natal, Brazil
roger@imd.ufrn.br

Edmundo Madeira
Institute of Computing (IC)
University of Campinas
Campinas, Brazil
edmundo@ic.unicamp.br

***Abstract*—The demand for multimedia services in mobile networks has increased in the last years. The high quantity of users mobile, both consuming and producing multimedia content to and from the Cloud can outpace the available bandwidth capacity. Notwithstanding the many benefits of Cloud Computing (CC), it has been noticed that it does not provide adequate latency in areas with high demands for multimedia content. Furthermore, using Fog Computing (FG) it is possible to improve on the above-mentioned issues, being especially useful in latency-sensitive applications such nodes are physically much closer to devices if compared to centralized data centers. The main goal of this work is twofold, first, it proposed a method to design/create a hierarchical multi-tier Cloud-to-Fog network. Second, it introduced a novel multimedia microservices placement algorithm for multi-tier Fog nodes. The performance assessment was composed of two months of real-world mobile network traffic data from Milan, Italy. The obtained results showed that our algorithm selects the nodes closer to the user to meet their demands. This decision improves the services delivered to end-users, for example, a local Fog node can instead be responsible for the video stream and is far quicker than offloading the processing to a centralized cloud platform.**

***Index Terms*—Cloud-to-Fog networks, Multimedia services, Microservices placement**

I. Introduction

During recent years, there has been a rapid proliferation of real-time multimedia services and applications, driven by a proliferation of connected devices [1]. Already, multimedia applications and services such as video streaming, sharing, and on-demand HD video require gigabit-per-second throughput and low latency. This type of communication will represent 82% of all Internet traffic by 2021 [2]. Incidentally, the next generation of telecom networks will allow this growth to be even greater, where networks can serve communication needs for billions of connected devices due to its high bandwidth capacity and low latency.

Despite the many benefits that CC has, such as high availability, scalability, and interoperability, multimedia services require low latency. Adapting them to this environment is a nontrivial task [3], [4]. New solutions are required to meet the demand of these latency-sensitive services [5], [6]. FG and Edge Computing (EC) present themselves as a joint solution to improve on this issue.

The main idea is to allocate Cloud-like resources physically closer to end-users [7]–[9]. Both have similar benefits compared to CC, including a reduction in latency to milliseconds and network congestion [10]. Thus, Fog nodes are infrastructures that can provide resources for services that can be executed in a distributed and independent way as microservices, available closer to end-users [11]. Additionally offloading to the Cloud environment and offering low latency communications, FG and EC have been recognized as an important platform to provide location-aware cloud services. This implies that the data is processed locally for more immediate response. Therefore, considering these factors, it is possible to place microservices in the most suitable Fog nodes through microservices placement mechanisms [8]. In doing that, it is possible to further reduce the latency as well as provide both high availability and resilience.

Placing microservices in the most appropriate locations of the FG can be related to the Capacitated Facility Location Problem (CFLP) [5]. This is also known as location analysis, as it deals with the optimal placement of facilities (resources) to minimize the cost of satisfying some set of demands (of the clients) while considering a set of constraints like the distance between facilities and clients or competitors' facilities [12].

The algorithms to the Multimedia Microservices Placement Problem (MMPP) can be evaluated in Cloud-Fog hierarchical environments. In these environments, nodes are hierarchically organized in tiers, from the Edge to the Cloud. The non-availability of these environments makes the evaluation of these algorithms a challenge.

In order to improve on the aforementioned issues, this work proposes a hierarchical multi-tier Cloud-to-Fog network along with an algorithm to MMPP. The main contributions are twofold:

- The design, implementation, and assessment of a new process to build a hierarchical multi-tier Cloud-to-Fog network for multimedia services distribution.
- An novel algorithm to the MMPP modeled as CFLP. The goals are to select the minimum number of nodes, considering their hardware capacities for providing multimedia services in such a way that the latency for servicing all the demands is minimized.

The performance assessment was conducted in a simu-

lated environment on MultiTierFogSim[1]. The environment simulated is based on real data witch representing workload variation in a metropolitan area occupied by mobile users. The results show that our algorithm can achieve a good balance among the Fog nodes' geographical location along with their hardware capacity and the users' location.

The remainder of this paper is organized as follows. Section II gives an overview of the main related work. Section III presents the system model and the problem formulation. Section IV details the experimental method and results. Finally, Section V presents the conclusion and future work.

II. Related Work

Several research efforts are being done to address the issue of reducing latency to deliver multimedia services in the context of CC, FG, and EC. They range from changes in the Cloud architecture [13] to the use of Fog/Edge nodes to reduce network delay and improve Quality of Experience (QoE) [14], [15].

The optimal deployment of Cloud-assisted video distribution services is addressed in [13]. The authors jointly minimize operational costs and latency. The joint minimization issue is handled by an offline algorithm modeled based on the Nash bargaining solution. Also, a solution is proposed to play down operational costs and under-provisioning of resources. The outcomes show that the proposed solution achieves good harmony among multiple objectives and definitely optimizes both operational costs and QoE. However, regarding ordinary data service in the Cloud (e.g., a short text and image service), latency tends to be negligible so that there is relatively little motivating to provide the latency announcement services.

A solution to support the Quality of Service (QoS) requirements of applications, such as multimedia services, was proposed by [16]. The technique combines the Cloud-Fog operations and can accomplish high system capacity whereas granting low latency for requested services. The problem is modeled as an Integer Linear Programming (ILP) model for latency optimization. The authors concluded that there are benefits of service distribution among multi-tier Fog nodes because avoid the high delay access on the cloud layer. Nevertheless, the effect on time overhead created by the service distribution with an expansive number of Fog nodes for mobile users is not considered. In normal conditions, this may be analyzed, however, in an unusual situation this could be a problem. For example, resources can be consumed by a big group of users.

The services allocation problem in Fog Radio Access Networks (FRAN) is also addressed [14]. They developed both centralized and distributed transmission aware cache placement strategies to minimize users' average download delay while meeting the fog storage capacity constraints. The authors concluded that the proposed algorithms improve the users' cache hit probability and provide more flexible cooperative transmission opportunities for the users. Nevertheless, the services stored in a single Fog node can not provide adequate QoE in areas with high demands, due to their storage capacity constraints. A collaborative resource strategy in multi-tier Fog nodes receives more attention. Also, the user preference model was not validated and assumed all the users' requests for multimedia services.

In light of the aforementioned issues, it is proposed a hierarchical multi-tier Cloud-to-Fog network and an algorithm to select the minimum number of nodes. One advantages is, the service distributions in the network and close to users allows processing and storage close to the data source, without the need to send all of these services to the remote Cloud or other centralized systems. It is taken into consideration their hardware capacities for providing multimedia services in such a way that the latency for servicing all the demands is minimized.

III. Cloud-to-Fog Networks and Multimedia Microservice Placement

This section introduces our hierarchical multi-tier Cloud-to-Fog network and provides an ILP formulation for MMPP.

The Cloud-Fog environment nodes work collaboratively, in which the Cloud is able to migrate multimedia services to the multi-tier Fog nodes and the other way round, whenever necessary. In such a hierarchy, it is expected a strongly connected and fully Fog-enabled scenario, where the highest tier (Cloud Tier) has more hardware resources than the lowest tier (Base station Tier). In contrast, the lowest tier supports latency-sensitive applications not fully-attainable by CC.

The environment considered in this work assumes real-world characteristics, where the end-users have mobility and can request several types of multimedia services, such as Video on Demand (VoD), interactive video 3D, high-definition, or even Ultra High Definition Video (UHD) (that includes 4K UHD and 8K UHD) video streaming. These represent services with different latency requirements. Also, the demands for these services vary over time and region.

A. The design of Hierarchical Multi-tier Networks

This section introduces the proposed method to create/design a hierarchical multi-tier Cloud-to-fog network for multimedia services distribution. The method uses a bottom-up approach, starting from a set $BS = \{bs_1, bs_2, ..., bs_{bs}\}$ of base stations. The steps are:

---

1 - Define a graph $\mathcal{G} = (BS, \mathcal{E})$.
2 - Detect communities in $\mathcal{G}$.
3 - For each community, an upper tier node is added in $\mathcal{G}$ which communicates with all the base station nodes that the community belongs to. This step ends with removing the edges between all base stations.
4 - Add edges between all nodes in the current tier.
5 - Detect communities and remove edges between nodes of different communities.
6 - For each subgraph, add an upper tier node with edges between the node and the subgraph.

[1]https://github.com/fillipesansilva/MultiTierFogSim.git

7 - Go back to step 4.

Stopping criterion: Number of subgraph is **N**.

Then, add an upper tier node that connects to the lower tier nodes.

This method was applied to actual cellular network data collected by Telecom Italia in the region of Milan [17], but it can also be applied to other datasets. This dataset contains two months of network traffic data (November/2013 to December/2013). The geographical area is composed of a 100 $\times$ 100 grid, with a size of about 235 $\times$ 235 meters each. Every time a mobile user requests services to a telecommunication provider, a Call Detail Record (CDR) is recorded. This information is then compiled into 10-minute intervals. Furthermore, a base station set $BS = \{bs_1, bs_2, ..., bs_{bs}\}$ was obtained from CellMapper[2], which consists of the locations and coverage areas of active base stations observed in the two months periods.

Fig. 1 shows the map view of the scenarios studied. Fig. 1(a) depicts the seven communities over the base stations (colored dots), which are represented by the letters A to G. Each community portrays a region. It is possible to notice that many regions (e.g., C, D, F, and G) are composed of an urban and also a suburban segment. This indicates that the base stations in these areas are potentially complementary due to traffic patterns. To find the community set S, we use the Louvain heuristic. This is a fast algorithm $O(n+m \cdot logn+m)$, where n and m are the numbers of vertices and edges, respectively, used to detect communities in large-scale networks based on modularity optimization [18]. This method is aimed at finding partitions (structures composed of communities) that maximize the density of intra-group connections concerning the density of inter-group connections, and thus at finding dense optimal sub-graphs in large graphs.

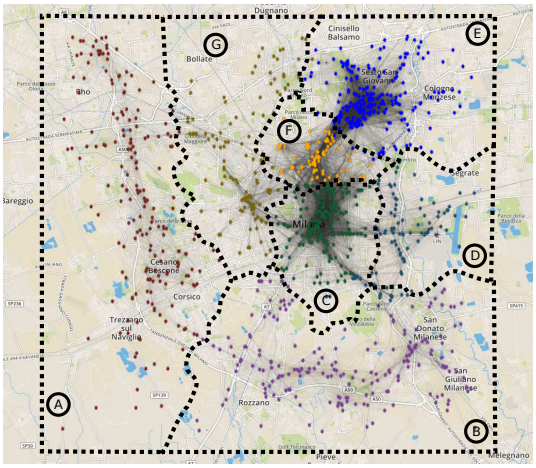
In this work, the stopping criterion is **N** = 2, but could be any value. Nodes added of each tier from steps **(2)** to **(6)** are labeled cloudlet (**CL1**, **CL2**, **CL3**, **CL4**, **CL5**, **CL6** and **CL7**), regional cloud (**RC1** and **RC2**) and cloud (**CL**), respectively. Fig. 1(b) illustrates the final scenario of the Cloud-Fog hierarchical environment.

To measure the hardware capacity and latency between fog nodes, two tools were adopted, namely CoLisEU [19] and EmuFog [20]. The former is a management tool for network infrastructures, which is capable of measuring the latency between the nodes. The latter is an extensible emulation framework tailored for FG scenarios, which can assess the hardware capacity of the nodes. It is worth noticing that both latency and Fog nodes' hardware capacity can dynamically change over time.

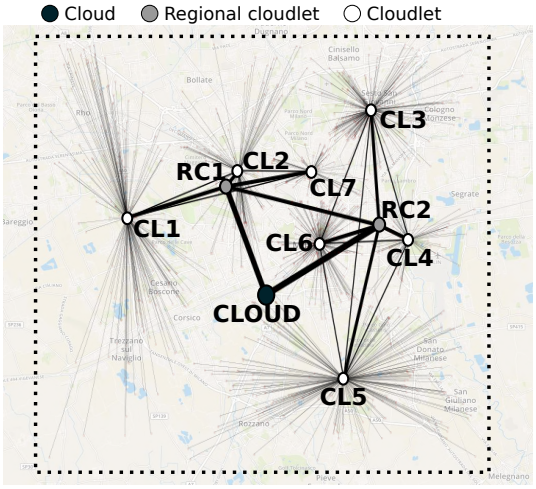
Table I shows the number of nodes and average coverage area per tier.

The grids were mapped to the coverage areas of the base stations and aggregate the CDR amount per base station. It is considered that there are multimedia services requests in a grid if its CDR amount is above the average. The multimedia services requests are aggregated by region.

[2]https://www.cellmapper.net/map



(a) Detecting communities by Louvain clustering algorithm based on base stations.



(b) Map view of the network topology

Fig. 1: Map view of the scenarios studied. Figure best viewed in colors.

TABLE I: Number of nodes and average coverage area per tier.

Nodes	Number of nodes	Coverage area (m^2 per node)
Base station	1150	492.46
Cloudlets	7	9072.67
Regional Cloud	2	31754.37
Cloud	1	552250

B. The design of MMPP

The solution proposed for the MMPP is modeled as CFLP, where (i) Fog nodes are the potential facility sites, (ii) the user's multimedia service requests are the demand, (iii) Fog nodes storage capacity and user's demand are the constraints,

(iv) and the multimedia services correspond to the kind of service. The set of Fog nodes capable of processing and storing multimedia microservices is given by $\mathcal{L}$. This includes Cloud, regional Cloud, and Cloudlets nodes, each one of them with capacity $c_{max(\ell)}$, where $\ell \in \mathcal{L}$. Sets C, RC, CL, and BS represent the Cloud, regional Cloud, Cloudlets, and base station nodes, respectively. The set of regions is represented by R, each one of them with a set of demands D_r, where $r \in R$. c_{ℓ,d_r} is also part of the input and represents latency from node ℓ to serving d_r. Further, variable $x_{\ell,d_r} \geq 0$ represents the fraction of the demand d_r filled by Fog node ℓ, and binary variables $y_w^{ms} = 1$ indicates if multimedia service ms is installed at node ℓ, $y_w^{ms} = 0$ otherwise. Table II gives the notation used in the model.

An integer-optimization model for MMPP can be specified as follows:

Minimize

$$\sum_{\ell \in \mathcal{L}} y_\ell^{ms} + \sum_{\ell \in \mathcal{L}} \sum_{r \in R} c_{\ell,d_r} \cdot x_{\ell,d_r} \quad (1)$$

subject to

$$\sum_{\ell \in \mathcal{L}} x_{\ell d_r} = d_r \qquad \forall r \in R \quad (2)$$

$$\sum_{r \in R} x_{\ell,d_r} \leq c_{max(\ell)} \cdot y_\ell^{ms} \qquad \forall \ell \in \mathcal{L} \quad (3)$$

$$x_{\ell,d_r} \geq 0 \qquad \forall \ell \in \mathcal{L} \; and \; \forall d_r \in R \quad (4)$$

$$y_\ell^{ms} \in \{0,1\} \qquad \forall \ell \in \mathcal{L} \quad (5)$$

The objective function 1 selects the minimum number of Fog nodes considering their storage capacity to deploy multimedia microservices, in such a way that the latency for meeting all the demands is minimized. The constraint in Eq. 2 requires that each region's demand r for multimedia services must be satisfied. The capacity of each node ℓ is limited by the constraint in Eq. 3, that is, if node ℓ is not activated, the demand satisfied by ℓ is zero. Otherwise, its capacity restriction is observed. Finally, the constraints in Eqs. 4-5 set the minimum values for the decision variables. The linear programming model was coded using the Gurobi Optimizer solver [21]. Gurobi is a commercial mathematical programming solver. It is possible to implement shared-memory parallelism, which is an efficient way to exploiting any number of processors and cores per processor. The solver uses an iterative process to converge on an optimal solution.

TABLE II: Notation used in the MMPP.

Input Parameters	
Notation	**Description**
C	Set of Clouds
RC	Set of regional Clouds
CL	Set of Cloudlets
BS	Set of base stations
$\mathcal{L}$	Set of Fog nodes where the multimedia services can be deployed.
R	Set of regions
D_r	Set of demands of region r, where $r \in R$
$c_{max(\ell)}$	Storage capacity of node ℓ, where $\ell \in \mathcal{L}$
c_{ℓ,d_r}	Cost of transportation from Fog node ℓ to serving d_r
Decision variables	
y_ℓ^{ms}	**1** if multimedia service ms is deployed at node ℓ. **0** if not.
x_{ℓ,d_r}	Fraction of the demand d_r filled by Fog node ℓ.

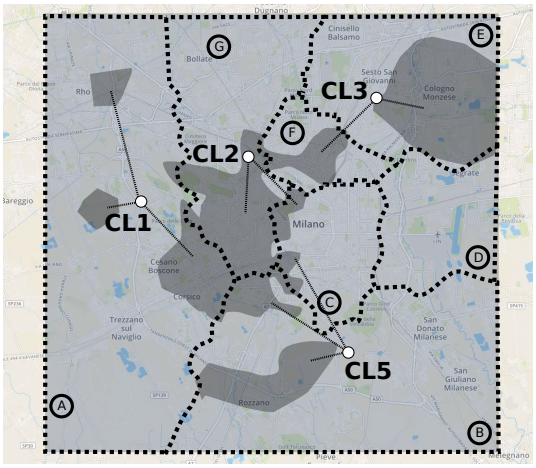
IV. Assessment results

Figs. 2-4 show the instantaneous multimedia service requests fragmentation measured in the Milan region for each snapshot. Each one of these figures presents two plots. Plots labeled "(a)" relate the snapshot of the multimedia service requests in the scenario as well as the Fog nodes selected to receive the multimedia microservices. In these plots, every gray spot corresponds to one demand $d_r \in D_r$, where $r \in R$, for multimedia services that must be provided. The circles represent the Fog nodes enabled to the placement of multimedia microservices. Plots "(b)" show the Fog nodes' hardware capacity (x-axis) and latency (y-axis) at that moment. The resulting figures give a rough, yet intuitive, idea of the Fog nodes selected in different possibilities. It is considered that one multimedia service request takes between 300 and 800 Millions of Instructions Per Second (MIPS) to be processed [10]. Additionally, the maximum acceptable delay to deliver multimedia services is less than 0.1 seconds [22].

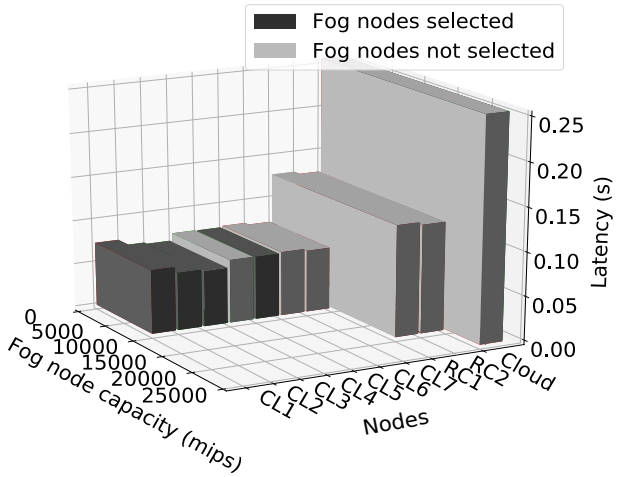
Fig. 2(a) shows the first snapshot (Nov/17/2013 at 06:00). It exhibits a low traffic intensity (≈40.000 MIPS) during dawn on weekends. Based on this, our algorithm selected the CL1, CL2, CL3, and CL5 nodes. Fig. 2(b), depicts that these Fog nodes have an appropriate hardware capacity and the lowest latency to meet this demand. Also, they are positioned geographically close to these regions, reducing the latency and enhancing user experience.

Fig. 3(a) shows the second snapshot (Dez/10/2013 at 09:30). It exhibits a medium traffic intensity (≈63.000 MIPS) in the morning hours. In this case, all the Fog nodes have a maximum acceptable delay to deliver multimedia services, i.e., less than 0.1 seconds. Also, all Cloudlet nodes are low on hardware capacity (they may be running other services, for example). Therefore, our algorithm selected the CLOUD, RC1, and RC2 nodes. Fig. 3(b) shows their characteristics.

Finally, Fig. 4(a) shows the sixth snapshot (Dez/28/2013 at 14:30). This is based on high traffic peaks (≈180.850 MIPS) during the working hours of weekdays. Fig. 4(b) shows that all Fog nodes have a maximum acceptable delay ($\leq$ 0.1 s), but the total hardware capacity is ≈144.880 MIPS. In this special case, all the Fog nodes are selected to meet as much of this demand as possible, ≈ 144,880 MIPS. Thereby, some regions will not be served or some users will have their video-rate adapted, delivery with poor QoE due to the low nodes' hardware capacity.

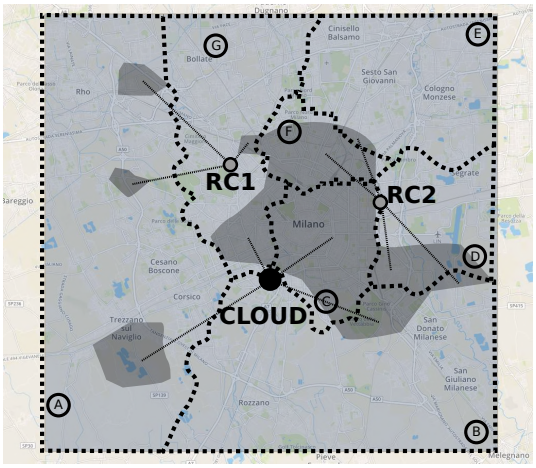


(a) Multimedia requests and selected nodes.

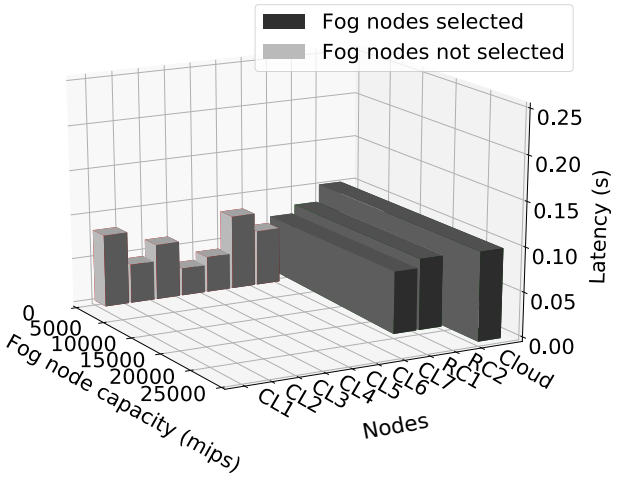


(b) Nodes' hardware capacity and latency

Fig. 2: Low traffic intensity



(a) Multimedia requests and selected nodes.



(b) Nodes' hardware capacity and latency

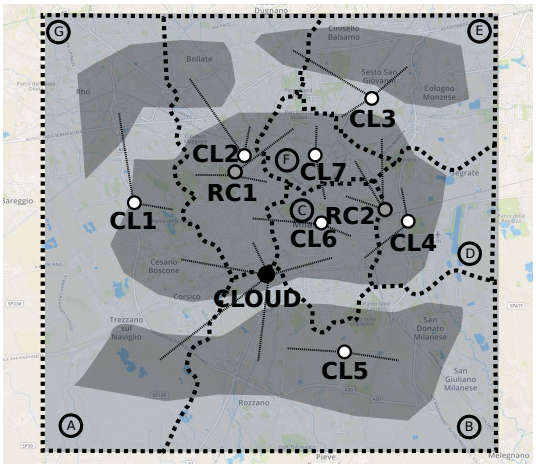
Fig. 3: Medium traffic intensity

Taking everything into consideration, it is possible to infer that the Fog nodes' storage capacity and their geographical location, number of nodes able to process tasks, amount demand for multimedia services, and network latency are all paramount factors to decide which Fog nodes can deploy multimedia services. Also, selecting these Fog nodes closer to the user using a distributed strategy may reduce the bandwidth which results in lower costs and improves efficiencies of the network, guaranteeing that most users who depend on the Fog are served, and improving the network deployed to mitigate provider costs.

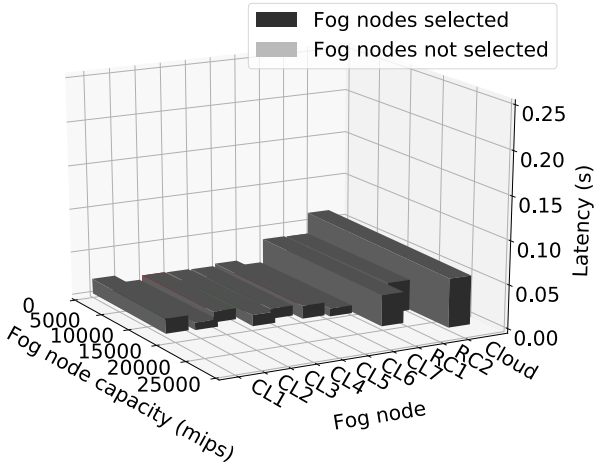
V. Conclusion

The constant communication technology's advancements and the great availability of streaming video services bring the need for new methods to ensure the quality for end-users. To improve on this issue, first, this work presented a process to design/create a hierarchical multi-tier Cloud-to-Fog network for multimedia services distribution. Moreover, this work also proposed an algorithm that selects the minimum number of nodes able to deliver multimedia services with low latency in multi-tier Fog nodes architecture. The performance assessment was carried out using two months of real-world mobile network traffic data in Milan, Italy.

The results showed that our proposed algorithm is able to select the Fog nodes closer to the users to meet their requests. Hence, this solution enhances the QoE, since the response time is lower in comparison to the Cloud tier, reducing the latency. Moreover, reducing the demand on the Cloud allows turning off servers in the data center to save energy. As has been noted, the proposed solution can be used in a real-world network to cope with future challenges in providing seamless and, at the same time, high-quality multimedia services in hierarchical multi-tier Fog nodes. As future work, we intend to extend our hierarchical multi-tier Fog nodes utilizing new methods, such as Density Based Spatial Clustering of Application with Noise (DBSCAN) to find communities and analyze the energy consumption and network usage in our simulation Also, we will adopt more dynamic evaluation scenarios to prove the benefits of the algorithm.



(a) Multimedia requests and selected nodes.



(b) Nodes' hardware capacity and latency

Fig. 4: High traffic intensity

Acknowledgments — This research is part of the INCT of the Future Internet for Smart Cities funded by CNPq proc. 465446/2014-0, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001, FAPESP proc. 14/50937-1, and FAPESP proc. 15/24485-9.

References

[1] R. Immich, L. Villas, L. Bittencourt, and E. Madeira, "Multi-tier edge-to-cloud architecture for adaptive video delivery," in *2019 7th International Conference on Future Internet of Things and Cloud (FiCloud)*, Aug 2019, pp. 23–30.

[2] C. V. networking Index, "Forecast and methodology, 2016-2021, white paper," *San Jose, CA, USA*, vol. 1, 2016.

[3] L. Bittencourt, R. Immich, R. Sakellariou, N. Fonseca, E. Madeira, M. Curado, L. Villas, L. DaSilva, C. Lee, and O. Rana, "The internet of things, fog and cloud continuum: Integration and challenges," *Internet of Things*, vol. 3-4, pp. 134 – 155, 2018.

[4] C. Quadros, E. Cerqueira, A. Neto, A. Riker, R. Immich, and M. Curado, "A mobile qoe architecture for heterogeneous multimedia wireless networks," in *2012 IEEE Globecom Workshops*, 2012, pp. 1057–1061.

[5] C. Da Silva, A. Rodrigo, S. da Fonseca, and L. Nelson, "On the location of fog nodes in fog-cloud infrastructures," *Sensors*, vol. 19, no. 11, p. 2445, 2019.

[6] R. Immich, E. Cerqueira, and M. Curado, "Towards a qoe-driven mechanism for improved h.265 video delivery," in *Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net)*, June 2016, pp. 1–8.

[7] T. X. Tran, P. Pandey, A. Hajisami, and D. Pompili, "Collaborative multi-bitrate video caching and processing in mobile-edge computing networks," in *Wireless On-demand Network Systems and Services (WONS), 2017 13th Annual Conference on*. IEEE, 2017, pp. 165–172.

[8] T. Taleb, S. Dutta, A. Ksentini, M. Iqbal, and H. Flinck, "Mobile edge computing potential in making cities smarter," *IEEE Communications Magazine*, vol. 55, no. 3, 2017.

[9] E. S. Gama, R. Immich, and L. F. Bittencourt, "Towards a multi-tier fog/cloud architecture for video streaming," in *2018 IEEE/ACM International Conference on Utility and Cloud Computing Companion (UCC Companion)*, 2018, pp. 13–14.

[10] R. Mahmud, R. Kotagiri, and R. Buyya, "Fog computing: A taxonomy, survey and future directions," in *Internet of everything*. Springer, 2018, pp. 103–130.

[11] C. F. C. Solutions, "Unleash the power of the internet of things," *Cisco Systems Inc*, 2015.

[12] R. Z. Farahani and M. Hekmatfar, *Facility location: concepts, models, algorithms and case studies*. Springer, 2009.

[13] J. He, D. Wu, Y. Zeng, X. Hei, and Y. Wen, "Toward optimal deployment of cloud-assisted video distribution services," *IEEE transactions on circuits and systems for video technology*, vol. 23, no. 10, pp. 1717–1728, 2013.

[14] J. Liu, B. Bai, J. Zhang, and K. B. Letaief, "Cache placement in fog-rans: From centralized to distributed algorithms," *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7039–7051, 2017.

[15] J. Kharel and S. Y. Shin, "Multimedia service utilizing hierarchical fog computing for vehicular networks," *Multimedia Tools and Applications*, pp. 1–24, 2018.

[16] V. B. C. Souza, W. Ramírez, X. Masip-Bruin, E. Marín-Tordera, G. Ren, and G. Tashakor, "Handling service allocation in combined fog-cloud scenarios," in *2016 IEEE international conference on communications (ICC)*. IEEE, 2016, pp. 1–5.

[17] G. Barlacchi, M. De Nadai, R. Larcher, A. Casella, C. Chitic, G. Torrisi, F. Antonelli, A. Vespignani, A. Pentland, and B. Lepri, "A multi-source dataset of urban life in the city of milan and the province of trentino," *Scientific data*, vol. 2, p. 150055, 2015.

[18] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.

[19] M. A. Marotta, L. R. Faganello, M. A. K. Schimuneck, L. Z. Granville, J. Rochol, and C. B. Both, "Managing mobile cloud computing considering objective and subjective perspectives," *Computer Networks*, vol. 93, pp. 531–542, 2015.

[20] R. Mayer, L. Graser, H. Gupta, E. Saurez, and U. Ramachandran, "Emufog: Extensible and scalable emulation of large-scale fog computing infrastructures," in *2017 IEEE Fog World Congress (FWC)*. IEEE, 2017, pp. 1–6.

[21] G. Optimization, "Inc.,"gurobi optimizer reference manual," 2015," 2014.

[22] E. Liotou, D. Tsolkas, N. Passas, and L. Merakos, "Quality of experience management in mobile cellular networks: key issues and design challenges," *IEEE Communications Magazine*, vol. 53, no. 7, pp. 145–153, 2015.