Efficient Green Solution for a Balanced Energy Consumption and Delay in the IoT-Fog-Cloud Computing

Adila Mebrek , Leila Merghem-Boulahia, and Moez Esseghir
ERA/ICD - UMR CNRS 6281, University of Technology of Troyes,12 rue Marie Curie
CS 42060, 10004 TROYES CEDEX, France
{adila.mebrek, leila.merghem_boulahia, moez.esseghir}@utt.fr

***Abstract*—This paper introduces a study of the fog computing suitability assessment as a solution for the increasing demand of the IoT devices. In particular, we focus on the energy consumption and the Quality of Service (QoS) as two important metrics of the performance of the fog. Therefore, we present a modeling of these two metrics in the fog. Then, we express the problem as constrained optimization and solve it efficiently using Evolutionary Algorithms (EA). Our approach stands out as an energy-efficient solution.**

***Index Terms*—Energy Efficiency, Fog Computing, Cloud computing, Internet of things (IoT), Service latency.**

I. Introduction

The cloud offers an efficient computing models where resources such as online applications, computing power, storage and network infrastructure can be shared as services through the internet [1]. In recent years, cloud systems provided a solution to the IoT. The IoT is a concept that refers to the digital interconnection of everyday objects with the Internet [2]. Currently, IoT applications are increasing in various fields. However, the number of devices connected to the Internet worldwide (therefore accessing cloud services) is in continuous growth. This means that the communication latency and the power consumed during the communications with the cloud turn down the expected advantages of this technology.

Fog Computing [3] is a solution proposed by Cisco in 2012, which subdues the shortcomings of cloud computing. It is a highly distributed platform, with nodes located at the edge of the network. These nodes offer resources such as computing, storage, and networking to the applications operating under this infrastructure. Recently, many works such as [4] have investigated the benefits of the fog in the context of IoT applications. Works similar to [5] have also explored issues in the fog computing such as security, privacy, and resources allocation. With the increase in the number of users demanding latency-sensitive services, the energy consumption in the fog computing is receiving an important attention by the research community. Several efforts have been made to build an energy models, managing the workload fluctuation and trying to achieve an efficient trade-off solutions between the QoS and energy consumption in the fog computing. The energy consumption of all devices in home-based fog computing environment and the energy-delay trade-off in different levels of the fog-cloud interaction have been highlighted widely in researches, using tools from stochastic optimization [6]. Since the devices/sensors are energy-constrained, works like [7] have focused on some issues like residual battery lifetime, energy-characteristics of the communications in these devices. Differently from the previous works, our models formulate the energy consumption and the delay for the entire execution cost rather than focusing only on the devices. Our main contribution is to investigate the problem of energy consumption of the fog computing in the context of IoT applications, and proposing a balanced energy-delay solution based on Evolutionary Algorithms approach.

The rest of the paper is organized as follows. In the section II, we present an overview of the fog-cloud computing system and a model of the performances metrics. The problem formulation and the proposed solution IGA (Improved Genetic Algorithm) is also described in Section II. The assessment of our solution is covered in Section III, and the Section IV concludes the work.

978-1-5386-1465-5/17/$31.00 ©2017 IEEE

II. Description of the problem and the proposed solution

We consider the fog architecture described as follows: We suppose that the cloud has one DC attached to the network core. The network at the edge includes several fog nodes called Fog Instances (FI). These fog instances can cooperate (by sharing network, computational and storage) to provide an optimal service for IoT objects. Considering this settled environment, we define below the performance metrics. The main parameters used throughout the paper are summarized in Table I.

A. Energy consumption

To mathematically schematize the energy consumed when an object sends a request N for a service located in the fog, we split it into three components: (a) the energy

TABLE I
SUMMARY OF NOTATIONS

Symbol	Description
i	Fog or Cloud identifier.
K	The size of the request N (bytes).
R	The size of the answer to the request (bytes).
$E^i_{network}$	E^{fog}_{edge} or E^{Cld}_{core}.
E^{fog}_{edge}	The energy consumed to transmit one data byte within the edge network (J).
E^{Cld}_{core}	The energy required to transfer one byte of data from IoT object to the cloud data center (J).
U	The utilization rate of a network element, within either the fog or the cloud.
Ci_{idle}	The average idle capacity of the network element, either the fog or the cloud (bytes/s).
Ci_{max}	The average maximum capacity of the network element, either the fog or the cloud (bytes/s).
Pi_{idle}	The idle power of network element (kw).
Pi_{max}	The maximum power consumption of a network element (kw).
DX^i	The transmission delay of one data byte form an object to fog instance or the cloud DC (s).
T^i_θ	The time to process an instruction in the fog or the cloud (s).
θ^i	The size of an instruction in the fog or the cloud (bytes).

consumed in the transport network for transmission of the object's request. This includes the energy consumed by the access technology at the edge network (routers and switches between the object and the fog instance) and the energy consumed by the access technology within fog instance ($E^{fog}_{Acs,N}$); (b) the energy consumed to process and store the object's request ($E^{fog}_{Prc-Str,N}$); and (c) in some cases the request may require the intervention of the cloud computing, thus the fog instance sends the request to be processed to the cloud ($E^{fog}_{Fwd,N}$). The total energy consumed can be expressed as:

$$E_{fog} = E^{fog}_{Acs,N} + E^{fog}_{Prc-Str,N} + \beta E^{fog}_{Fwd,N} + E^{fog}_{Acs2,N} \tag{1}$$

$E^{fog}_{Acs2,N}$ is the energy consumed in the fog to send the answer to the object, and $\beta = \frac{N_{sent}}{N_{total}}$ is the ratio of the number of requests sent from the fog to the cloud.

In the case of cloud computing, the energy consumption consists of two components: (a) the energy consumption of the transport network between the object requesting data and the DC hosting the data ($E^{Cld}_{Acs,N}$); and (b) the energy consumed by the cloud to process and store data($E^{Cld}_{Pr-Str,N}$). It can be expressed as:

$$E_{Cld} = E^{Cld}_{Acs,N} + E^{Cld}_{Pr-Str,N} + E^{Cld}_{Acs2,N} \tag{2}$$

As for the fog, $E^{Cld}_{Acs2,N}$ is the energy consumption of the cloud to answer the object.

1) Access energy: The access energy consumed can be defined as:

$$E^i_{Acs,N} = K * E^i_{network} + K * h_i * (\frac{P^i_{idle}}{C_{idle}} + \frac{P^i_{max}}{U * C^i_{max}}) \tag{3}$$

2) Data Processing and Storage : Let $E^i_{process}$ and E^i_{store} be the average energy consumption to process and store one data byte in the fog (respectively in the cloud). Thus, the total energy can be expressed as:

$$E^i_{Pr-Str,N} = K * E^i_{process} + K * E^i_{store} \tag{4}$$

we express $E^i_{process}$ as:

$$E^i_{process} = P^i_{idle} * \frac{T_N}{T_{tot}} + P^i_{max} * \frac{T_{act,N}}{T_{tot}} \tag{5}$$

3) Forwarding to the cloud: We define the energy consumed by the fog instance due to the transmission to the cloud as:

$$E^{fog}_{Fwd,N} = K * E^{fog}_{core} \tag{6}$$

with E^{fog}_{core} is the energy consumed for the transmission of one data byte between the fog instance and the DC.

B. Service latency

The service latency of a request N sent by an IoT object represents the response time to be served in the fog, and is computed as the sum of (a) the transmission latency ($D^{fog}_{Tr,N}$), (b) the processing latency of the request, and in the case where a fog instance sends the request to be processed in the cloud ($D^{fog}_{Pr,N}$), we include (c) the delay consumed due to forwarding the request to the cloud ($D^{fog}_{Fwd,N}$).

The total delay is computed as:

$$D_{fog} = D^{fog}_{Tr,N} + D^{fog}_{Pr,N} + \beta D^{fog}_{Fwd,N} \tag{7}$$

In the case of cloud computing, the latency of service consists of two components: (a) the delay for the transport of the request between the object and the DC hosting the data ($D^{Cld}_{Tr,N}$); and (b) the delay of the cloud to process and store the data ($D^{Cld}_{Pr,N}$).

The total delay of the cloud can be expressed as:

$$D_{Cld} = D^{Cld}_{Tr,N} + D^{Cld}_{Pr,N} \tag{8}$$

1) Transmission latency: The delay for transmission can be computed as :

$$D^i_{Tr,N} = K * DX^i + R * DX^i \tag{9}$$

2) Processing latency: We express the processing delay as:

$$D^i_{Pr,N} = K * \frac{T^i_\theta}{\theta^i} \tag{10}$$

3) Forwarding to the cloud latency: The transmission latency to the cloud can be expressed as:

$$D^{fog}_{Fwd,N} = K * DX^{Cld} + K * \frac{T^{Cld}_\theta}{\theta^{Cld}} + R * DX^{Cld} \tag{11}$$

C. Problem formulation

Towards the power consumption and delay trade-off in fog-cloud computing, we provide an overview of the fog based optimization model. We model the fog architecture as an undirected graph $G(V, E)$, with the set V containing nodes and E the set of links. The set $F = \{f_1, f_2, \ldots, f_m\} \subset V$ contains fog instances and $O = \{o_1, o_2, \ldots, o_n\} \subset V$ is the set of objects. Each node in O is a candidate to be assigned to a node in F. Each link $e = (i, j)$ is between an object $i \in O$ and a fog instance $j \in F$. We assume that the amount of energy consumed and latency in the links and the fog depend on the distance between i and j and the workload of the fog instance.

Our objective is to find an optimal assignment of x_e that minimizes the total energy consumption of the mobile. A first constraint is needed to guarantee that each object will be assigned to exactly one fog instance. A second constraint have to be added to ensure that total execution time must be below acceptable delay D_j. This problem can be modeled as follows: Given a set of objects and a set of fog instances, each fog instance is selected by a subset of objects to provide a service. If an object i selects a fog instance j, it will have an effect on the energy consumed by all the m fog instances. Let $p_{i,j}$ be the amount of energy consumed by j to serve i, noted p_e. Considering $d_{i,j}$ the delay taken by j to satisfy the request of i. The mathematical representation is as follows:

$$minimize \sum_{e=0}^{|E|} p_e * x_e \tag{12}$$

where $X = \{x_e | e = 1, \ldots, |E|\} \in \{0, 1\}^{|E|}$, x_e is the binary decision variable to indicate whether the link $e = (i, j)$ is selected so that the object i will be served by fog instance j.

subject to the following constraints:

- Ensure that each object is assigned to one fog instance. That is,

$$\forall i \in O : \sum_{j=0}^{m} x_{i,j} = 1 \tag{13}$$

- The selected fog must meet the delay constraint

$$\sum_{i=0}^{n} d_{i,j} * x_e \leq D_j \quad \forall i \in O, \forall j \in F, (i, j) = e \in E \tag{14}$$

where $d_{i,j}$ represents the delay of the fog j to serve the object i, and Dt is the delay threshold for the fog instance j, defined as

$$D_j = \min_{f \in F} \max_{o \in O} \{d_{o,f}\} \tag{15}$$

D. Proposed solution

The main idea of the proposed solution (IGA) is that according to the number of objects present in the system, we adopt two different optimization methods, BIP algorithm and the improved genetic algorithm, to solve the problem presented in Section II.c. The process is defined as followed: First, we determine the number of objects, then we decide whether to use BIP algorithm or GA algorithm.

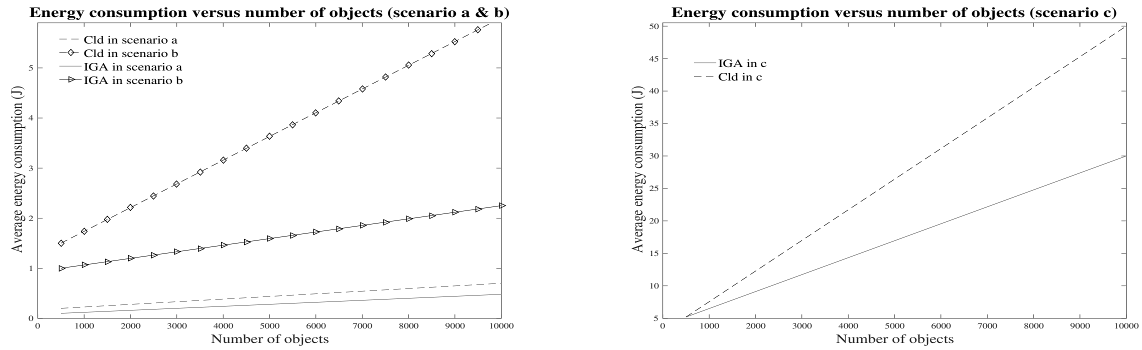
Our genetic algorithm is inspired from Genetic Algorithms, and it is described as follows: Each object (respectively fog instance) is encoded by its respective index. A chromosome corresponds to a distribution of all the assignment objects-fog instances possible. The fitness of a chromosome represents the total amount of the energy consumed by the fog instances used in the solution, with regard to the same constraints described before. According to the selected assignment object-fog instance, it will determine the performance of our solution. We start by encoding the objects and FIs, so that we can generate a population including all the genes (all the possible assignments object-fog). If the best solution is not replaced after $l \leq |F| * |O|^2$ iterations, the process is ended by taking the best member as the optimal solution. The selection of the parent members is done by using the roulette method. An improuved cross over process is applied to the parents then. We replace a gene in the new members with the set mutation probability.

III. Evaluation

We consider a system with various FIs connected to a single DC. Smart objects are uniformly distributed in an area (7.7 x 3.5 km^2). The machine instruction size is assumed to be 64 $bits$. Processing speed of the devices at the fog computing and the cloud data centers are taken as 1256 $MIPS$ $(ARMCortexA5)$ and $124,850$ $MIPS$ $(IntelCorei74770k)$ respectively.

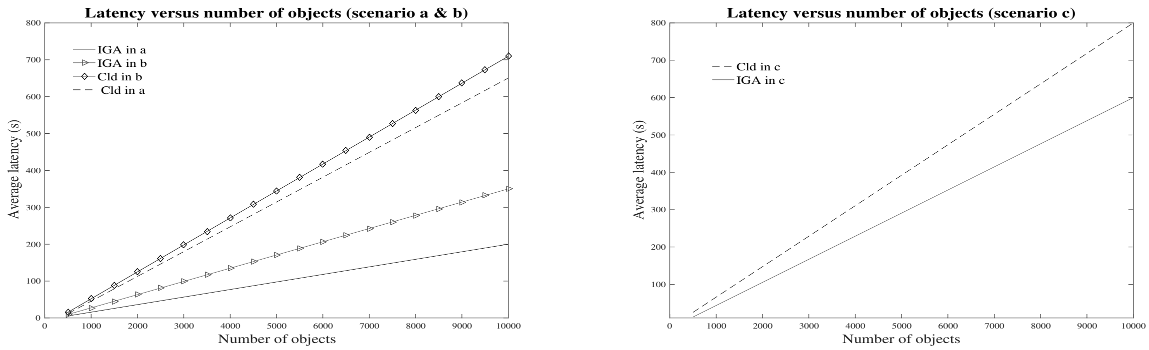
To validate our work, we formulate three different types of scenarios : (a) services for which the data source is primarily in fog instances with static content; (b) services for which the source of data is primarily in fog computing with dynamic content such as video surveillance; (c) services for which the source of data is not created in fog instances but must be pre-downloaded to fog instances from the cloud DC.

Fig.1 shows the results of three set of experiments to investigate the behavior in terms of energy consumption of IGA algorithm as the number of objects and the number of fog instances increase. IGA is used to create a preference list for pairing IoT object-Fog instances. In this set of experiments, we study the performance of our algorithm compared to the traditional cloud solution. From Fig. 1a and 1b, we can observe that for a small or medium number of IoT objects the performance of our algorithm and the cloud are quite the same in terms



(a) (b)

Fig. 1. Energy consumption against number of objects



(a) (b)

Fig. 2. Analysis of service latency

of energy consumption, in particular for scenario (a). Thus, the IoT devices do not take fully advantage of the fog resources. As the number of objects increases, the fog utilization rises up. The fog computing architecture improves the energy consumption compared to the performance of the cloud.

As shown in Fig. 2a and 2b, with the increase of the number of objects, the average on latency rises up with a linear slope. As the percentage of applications routed towards the DC grows up (scenario (c)), the transmission latency is observed to increase. However, the simulations show the efficiency of our solution over the cloud with the increase of IoT objects, even with the augmentation of the number of requests processed in the cloud DC.

IV. Conclusion

In this work, we presented a model to investigate the power consumption and delay in a fog-cloud computing and in the traditional cloud computing. Then, we proposed a solution inspired from Evolutionary Algorithms approach to resolve the trade-off problem. Simulations and numerical results have shown the practical relevance of our approach with a large number of real-time, low latency IoT applications.

Acknowledgment

This work is partly supported by CD10 and the SOLOTEC project which is financed by the European Union and Grand-Est region.

References

[1] Q. Zhang, L. Cheng, and R. Boutaba, Cloud computing: state-of-the-art and research challenges, J Internet Serv Appl, vol. 1, no. 1, pp. 718, May 2010.

[2] Evans, Dave. "The internet of things: How the next evolution of the internet is changing everything." CISCO white paper 1.2011 (2011): 1-11.

[3] P. Varshney and Y. Simmhan, Demystifying Fog Computing: Characterizing Architectures, Applications and Abstractions, arXiv:1702.06331 [cs], pp. 115124, May 2017.

[4] S. Shekhar and A. Gokhale, Dynamic Resource Management Across Cloud-Edge Resources for Performance-Sensitive Applications, in Proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, Piscataway, NJ, USA, 2017, pp. 707710.

[5] R. Deng, R. Lu, C. Lai, T. H. Luan, and H. Liang, Optimal Workload Allocation in Fog-Cloud Computing Toward Balanced Delay and Power Consumption, IEEE Internet of Things Journal, vol. 3, no. 6, pp. 11711181, Dec. 2016.

[6] V. Cardellini et al., A game-theoretic approach to computation offloading in mobile cloud computing, Math. Program., vol. 157, no. 2, pp. 421449, Jun. 2016.

[7] X. Chen, L. Jiao, W. Li, and X. Fu, Efficient Multi-User Computation Offloading for Mobile-Edge Cloud Computing, IEEE/ACM Transactions on Networking, vol. 24, no. 5, pp. 27952808, Oct. 2016.