Power Consumption and Delay in Wired Parts of Fog Computing Networks

Bartosz Kopras, Filip Idzikowski, Paweł Kryszkiewicz
Faculty of Electronics and Telecommunications, Poznan University of Technology, Poland

***Abstract*—In the last decade Cloud computing has seen a surge of popularity. Clouds, with their scale and high functionality, are used to outsource various infrastructure, platform, and software services. However, relaying solely on distant Cloud Data Centers (DCs) can be inefficient for many applications concerning mobile devices and Internet of Things (IoT) in general. A more decentralized Fog computing paradigm has been proposed to augment Cloud availability and execution.**

This work addresses latency and power consumption in Fog computing networks. Models for power consumption and delay are proposed. Performance of Fog computing is estimated using parameters setting based on real-world equipment and traffic. Our results tackle the balance between Fog and Cloud. Applications requiring heavy computations (relative to size of offloaded data) are best served by Cloud DCs, while it is faster (and more power-efficient) to compute "lighter" requests in the Fog Nodes (FNs). However, where is the trade-off between power consumption and delay in the context of Fog and Cloud? We answer this question modeling multiple architectures and using various network scenarios.

***Index Terms*—power, energy, latency, fog, cloud, green**

I. Introduction

IoT is rapidly expanding as the number of connected devices continues to grow [1]. This creates new challenges for the Information and Communication Technology (ICT) sector. On one hand, the "Things" often have limited storage, computational power, and battery life related to their small sizes. These limitations hinder their ability to process gathered data. On the other hand, sending all of this data to be processed at a remote Cloud would introduce unprecedented amount of traffic [2] in the backbone of the Internet, which would influence the total energy consumed by the Internet. Moreover, some applications such as video surveillance [3] or augmented reality [4] require low delays. Tasks performed for these applications are often too complex (and battery-draining) to be processed by Mobile Devices (MDs). Offloading these tasks to a Cloud DC could introduce unacceptable delays.

Fog computing [5] addresses these problems. The main idea behind Fog computing is to augment the Cloud by providing computational, networking, and storage resources closer to end users. This is achieved by introducing the Fog tier between the Cloud tier and the Things tier [6] as shown in Figs. 1 and 2. Various types of devices are present in the lowest (Things) tier. They include sensors, security cameras as well as "smart" devices commonly associated with IoT (vehicles, home appliances, personal equipment). The Fog tier consists of

The presented work has been funded by the National Centre of Research and Development, Poland within the FAUST project (no. PL-TW/V/3/2018).

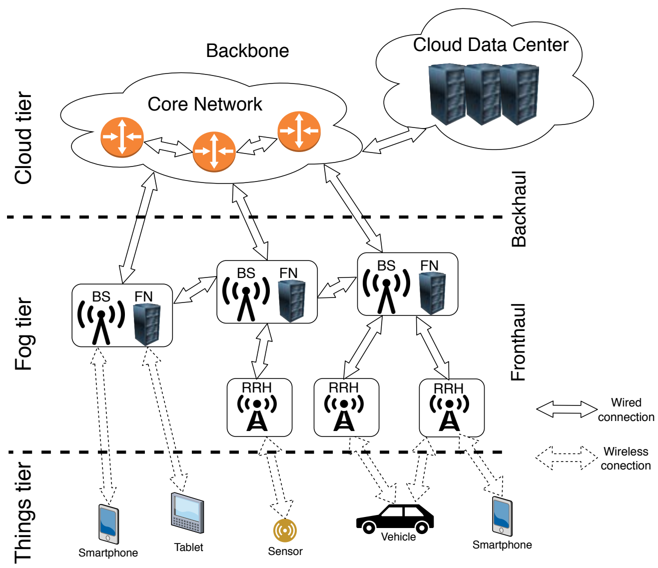


Fig. 1: Example architecture of Fog computing network with fronthaul, backhaul, and backbone parts captioned.

elements called FNs located near the end users, "on the edge" of the network. Thanks to this fact, data transmission between "Things" and FNs is quicker and less energy-consuming than alternative Thing-to-Cloud communication. "Things" can also communicate with FNs via Remote Radio Heads (RRHs). FNs can be interconnected and share offloaded workload. A detailed description of a FN, including its role in a network can be found in Sec. 5.5 of [6]. The Cloud tier contains DCs organized in multiple specialized clusters. Their high computation performance allows for extensive data analysis and storage. DCs are highly virtualized so multiple Virtual Machines (VMs) on a single physical cluster (let alone a single DC) can simultaneously serve multiple different applications.

We focus on the wired parts of the Fog computing network consisting of fronthaul, backhaul, and backbone as depicted in Fig. 1. The concept of fronthaul originates from Cloud Radio Access Networks (C-RANs) [7] and Fog Radio Access Networks (F-RANs). There is little agreement between the researchers on what constitutes a fronthaul. We define it in this work as wired links between RRHs and FNs. Consequently, backhaul consists of the links connecting FNs to the core network (Fig. 1). While the interconnections within the core network (including connections to the Cloud Data Centers) are regarded as backbone.

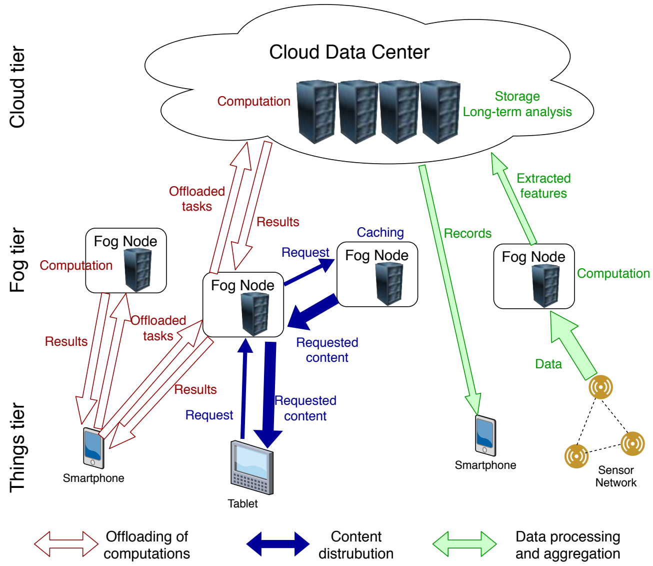


Fig. 2: Tasks performed by Fog computing network.

Three broad types of services provided by Fog network are shown in Fig. 2. They include (1) offloading, (2) content distribution, and (3) data aggregation. The idea of offloading data/computation/task from a device with limited capabilities to a more powerful one is not new. Othman *et al.* [8] show how to increase battery life of a laptop by putting some of its computational load to a nearby Base Station (BS). This idea is reused in the context of Fog computing. An overview of the offloading schemes is provided in [9]. Fog computing is based on moving resources closer to the edge of a network, i.e., Fog can be regarded as an "extension" of the traditional Content Delivery Network (CDN). An overview of content distribution methods used to save energy in Fog is also provided in [9]. Eventually, extensive data collected by IoT devices can be reduced at Fog nodes. E.g., processing raw Electrocardiogram (ECG) data in a nearby FN reduces the data sent to the Cloud by over 90% [10].

Starting from a survey of articles estimating power consumption and latency, we propose power consumption and latency models of a network using Fog computing. We consider costs related to both communication and computation in and between devices belonging to the Fog and the Cloud tiers of the network. Differently from other works, we take a holistic approach for the wired parts of the network, i.e., we take into account power spent by each device through which the data flows. Furthermore, we provide an extensive comparison of efficiency of utilizing Fog/Cloud or a mix of them using realistic network scenarios.

Our work is structured as follows. An overview of related work is provided in Section II. Mathematical models describing Fog computing are developed in Section III. In Section IV the performance of Fog computing is tested under various conditions using models from Section III. Finally, Section V concludes the work.

II. Related Work

In this section models used by researchers for estimation of power consumption and latency in Fog computing networks are discussed.

Table I provides an overview of power consumption and latency models used in related work on Fog computation networks. In all of the surveyed works there is a distinction between power/time spent on computation and on communication. In Table I entries in columns labeled "End devices", "Fog nodes", and "Cloud data centers" mean that these devices spend time/power performing computations. The remaining columns contain information regarding communications between devices. We point out that the focus of this work is set on power and latency occurring in the wired part of the network, i.e., in the Fog and Cloud tiers (Fig. 1).

The works [11] and [12] examine this problem from the point of view of a MD. They only consider the power consumed by these devices (on both computation and transmission to FNs). For latency, they consider contributions from the whole network.

On the other hand, in [13], the offloaded tasks are considered as they enter the Fog tier of the network. Computations in the FNs and Cloud DCs consume both power and time, while the communication between FNs and DCs introduces only delay.

The authors of [14] use an unorthodox network architecture, where all computational and storage resources are located in the lowest tier (each end user has a personal FN – a nano DC). Latency is not considered. However, they develop detailed models for power consumption related to communication (which include idle power of devices and number of nodes data flows through). These ideas are further developed in our work.

Extensive modeling is performed in the works of Sarkar *et al.* [15], [16]. Power consumption and delay is caused by both communication and computation in multiple tiers of the network. It coincides with the topic of this paper. However, their models are heavily skewed towards showing that processing data in the Fog tier is more efficient than doing it in the Cloud which yields dubious results (e.g., they show that processing 25% of requests in the Fog tier and 75% in the Cloud tier results in over 40% lower power consumption than processing 100% of requests in the Cloud tier).

As shown in Table I, our work has different focus from [11], [12], [14], i.e., we focus on power and latency due to FNs and Cloud DCs as well as communication between them. We contribute with a model of latency and power consumption of a Fog network including its interaction between the Fog and Cloud tiers. Differently from [13], we cover power consumption needed for communications between FNs and Cloud DCs. Differently from [15], [16], our model covers complexity of computation tasks and power spent on data transmission through each single node in the network. The concept of computational complexity is studied regarding its impact on offloading efficiency. Parameter values from credible sources are used to describe devices used in our models.

TABLE I: Overview of power consumption and latency models. Which parts of network are considered?

Research work (Year)	End devices	Comm. between end devices & FNs	Fog nodes (FNs)	Comm. between FNs	Comm. between FNs & Cloud DCs	Cloud data centers (DCs)
This work	not considered	power & latency* (incl. downlink)	power & latency	negligible	power & latency (incl. downlink)	power & latency
Dinh *et al.* [11] (2017)	power & latency	power & latency (incl. downlink)	latency	not considered	latency (incl. downlink)	latency
Liu *et al.* [12] (2018)	power & latency	power & latency	latency	not considered	latency	latency
Deng *et al.* [13] (2016)	not considered	negligible	power & latency	negligible	latency	power & latency
Jalali *et al.* [14] (2016)	not considered	power	power	not considered	not considered	not considered
Sarkar and Misra [15] (2016)	not considered	power & latency (incl. downlink)	power & latency	negligible	power & latency (incl. downlink)	power & latency
Sarkar *et al.* [16] (2018)	not considered	power & latency*	power & latency	negligible	power & latency	power & latency

* – gateways between end devices and FNs are considered rather than end devices themselves

III. Network Model

This section defines models used for estimating delay and power consumption related to the offloading of computations in Fog computing network. Computational requests are modeled in Section III-B. The power consumption of the network is modeled in Section III-C, and delay is modeled in Section III-D.

A. Network Description

In the bottom tier of the network there are end devices (e.g., smartphones, sensors) which may require offloading computational tasks. These tasks can be processed in either the Fog tier (consisting of a set $\mathbb{F}$ of FNs) or the Cloud tier (set $\mathbb{C}$ of DCs). Data is sent to the FNs through the RRHs as shown in Fig. 1 and Fig. 3. Then the results are transmitted back to the MD.

The FNs are capable of sharing the computational load between themselves. To simplify the model and following calculations, they do so with inducing neither additional delays nor power consumption. These costs have also been left out by other researchers [13], [15], [16].

B. Offloaded Tasks – Computational Requests

Let there be a total of N requests offloaded during analyzed time period T. $\mathbb{R}^F$ and $\mathbb{R}^C$ are the sets of requests offloaded to Fog tier and Cloud tier respectively. Let L^{R_i} be the size (in bits) of the *i-th* request R_i. Total amount of offloaded data L, amount of data offloaded to the Fog tier X and Cloud tier Y are therefore:

$$L = X + Y,\ X = \sum_{R_i \in \mathbb{R}^F} L^{R_i},\ Y = \sum_{R_i \in \mathbb{R}^C} L^{R_i} \quad (1)$$

Concepts discussed in this section are illustrated in Fig. 3. It is important to note that, in this work, the requests are examined as they enter the Fog tier of the network. The cost of transmitting a single request from MDs to RRH/FN is not considered.

The volume of offloaded data is not the only thing that matters when it comes to processing data. For certain applications,

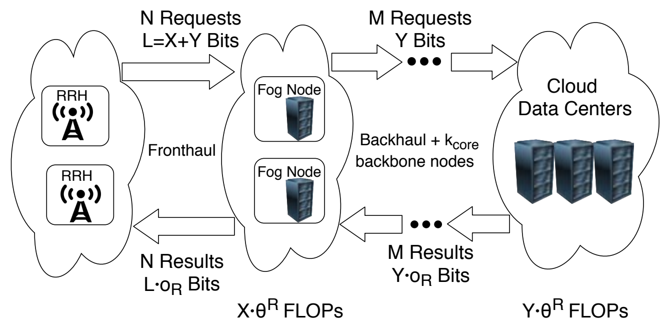


Fig. 3: Diagram explaining the flow of data through the considered Fog computing network.

the amount of required computations can be disproportionate to the size of an input, e.g., a Portable Game Notation (PGN) file representing a chess game [17] is minuscule ($<$ 1 kB) compared to the amount of computation used to analyze the game. Let θ^{R_i} be the computational complexity of task R_i defined as a ratio of single precision Floating Point Operations (FLOPs) required to process this task to its size in bits.

When the data has been successfully processed, the results are transmitted back to the MD. Let o^R be the average ratio of size of the result to the size of the offloaded task. If $o^R = 0$, then there is no transmission of the results.

C. Power Consumption

The power consumption model is divided into two parts: communication (transmission and reception of the data) and computation (processing of data). In this work, the power consumption of end devices is not considered.

1) Communication: For the power consumption of networking equipment, the linear model from [14] is used which includes idle power P_{idle} and active power that scales with load C (in bits/second) by parameter γ_b (in Joules/bit):

$$P(C) = P_{idle} + C\frac{P_{max} - P_{idle}}{C_{max}} = P_{idle} + C\gamma_b \quad (2)$$

where P_{idle} and P_{max} denote idle and maximum power consumption respectively. C_{max} is maximum load. Energy-per-bit cost of transmitting data through the core, backhaul, and fronthaul networks is equal to the number of devices through which data flows (k_{core}, k_{bck}, k_{fr} respectively) multiplied by the average γ_b parameter, i.e., $\overline{\gamma_b}$, of devices in this part of network:

$$\gamma_{b-core} = k_{core}\overline{\gamma_b},\ \gamma_{b-bck} = k_{bck}\overline{\gamma_b},\ \gamma_{b-fr} = k_{fr}\overline{\gamma_b} \quad (3)$$

The number k_{core} can vary from 1 if the Cloud DC is directly connected to the Fog tier of the network to 5-20 in a more realistic scenario of Cloud being few hundred/thousand kilometers away [18]. For backhaul transmission it is assumed that there is only a single hop/node from the FNs to the core network. k_{fr} is equal to 2 (RRH and FN).

Total energy E_{act}^{comm} used for "active" transmission is equal:

$$E_{act}^{comm} = \gamma_{b-fr} L\left(1+o_R\right) + \left(\gamma_{b-core} + \gamma_{b-bck}\right) Y\left(1+o_R\right). \quad (4)$$

Eq. (4) clearly shows that L bits go through the fronthaul to the Fog and, from these L bits, Y bits are transmitted further (via backhaul and core) to the Cloud.

We also consider idle power consumption of devices in the Fog tier of the network. For fronthaul communication it is assumed that there is one networking device in each of F FNs and H RRHs. Idle power consumption P_{idle}^{front} of devices in fronthaul is defined as:

$$P_{idle}^{front} = \left(F+H\right) P_{idle}^{f} \quad (5)$$

If there is a fixed number of B networking devices in the backhaul (e.g., one per FN), then the backhaul idle power consumption P_{idle}^{back} is given as:

$$P_{idle}^{back} = B \cdot P_{idle}^{b} \quad (6)$$

P_{idle}^{f} and P_{idle}^{b} denote idle power of a single networking device in fronthaul and backhaul respectively.

Total power P_{comm} spent on transmission is given as:

$$P_{comm} = \frac{E_{act}^{comm}}{T} + P_{idle}^{front} + P_{idle}^{back} \quad (7)$$

Respective values γ_b, P_{idle}, and C_{max} are taken from [19] and [20]. For IP routers we calculate $\gamma_b = \frac{P_{max}-P_{idle}}{C_{max}}$ assuming that P_{idle} is equal to 90% of P_{max} [20].

2) Computation: Different power consumption models for Cloud DCs and FNs should be used as their scale of operation is different. FNs are utilized solely by the tasks from within examined Fog computing network while Clouds are full of computational requests from various devices across the Web. For FNs we assume that they consume P_{act} watts of power when performing computations and P_{idle} when in idle-state. Values of P_{act} and P_{idle} depend on type and model of the device as well as parameters such as clock frequency.

$\mathbb{R}^{F_i}$ (subset of $\mathbb{R}^{F}$) is the set of requests processed in the FN i. Assuming the clock frequency f^{F_i} of FNs is fixed while performing calculations, then FN i is actively computing for:

$$t_{act}^{F_i} = \frac{\sum_{R_i \in \mathbb{R}^{F_i}} L^{R_i}\theta^{R_i}}{f^{F_i} s^{F_i}}, \quad (8)$$

where s^{F_i} is the FLOPs per cycle parameter of FN i. $t_{act}^{F_i}$ and $t_{idle}^{F_i}$ are the time-lengths in which FN i is in active and idle state respectively, $T = t_{act}^{F_i} + t_{idle}^{F_i}$. Therefore total power consumption P_{cp}^{fog} spent on computations in Fog equals:

$$P_{cp}^{fog} = \frac{\sum_{F_i \in \mathbb{F}} \left(t_{act}^{F_i} P_{act}^{F_i} + t_{idle}^{F_i} P_{idle}^{F_i}\right)}{T} \quad (9)$$

For Cloud DC, it is not feasible to estimate direct impact of offloaded workload on the clock frequency (and number of running servers). Instead, a simpler measure called Floating Point Operations per Second (FLOPS) per watt is used, this is a benchmark that is often used in comparing performance of supercomputers [21]. Let β^{C_j} be the FLOPS per watt value characterizing DC j. $\mathbb{R}^{C_j}$ (subset of $\mathbb{R}^{C}$) is the set of requests processed in the Cloud DC j. Power $P_{cp}^{C_j}$ consumed in Cloud DC j is calculated as follows:

$$P_{cp}^{C_j} = \frac{\sum_{R_i \in \mathbb{R}^{C_j}} L^{R_i}\theta^{R_i}}{\beta^{C_j}} \quad (10)$$

Power P_{cp}^{cld} consumed in whole Cloud tier is equal to the sum of power consumption of each Cloud DC:

$$P_{cp}^{cld} = \sum_{C_j \in \mathbb{C}} P_{cp}^{C_j} \quad (11)$$

Total power P consumed in the network is the sum of power consumption spent on transmission of data (P_{comm}) and on performing computations in Fog – P_{cp}^{fog} and in Cloud – P_{cp}^{cld}.

$$P_{tot} = P_{comm} + P_{cp}^{fog} + P_{cp}^{cld} \quad (12)$$

D. Latency

The delay model is also divided into two parts: communication and computation. The wireless communication channel between end devices and FNs/RRHs is not considered. Wireless transmission from MDs is pivotal to other research works [11], [12].

1) Communication: For delays associated with communication the Round-Trip Time (RTT) [22] metric is used. 7.5μs/km is chosen as a numerical value assigned to RTT based on [18]. It can be seen that for data computed in the Fog tier the RTT value is negligibly low as FNs are meant to be located close to the end users. On the other hand, when it comes to offloading computation to the Cloud tier, the RTT value is significant as the Cloud can be a few hundred (thousand) kilometers away from the source of the request.

The average-per-request transmission delay in the fronthaul D_{comm}^{front} is defined as:

$$D_{comm}^{front} = \frac{L\left(1+o_R\right)}{N \cdot r_{b,front}} \quad (13)$$

where L/N is the average size (in bits) of a computational request sent to be processed in either the Fog tier or the Cloud tier and $r_{b,front}$ is the bitrate of a fronthaul link between a RRH and a FN. RTT estimated by the length of the fronthaul link is assumed to be negligible.

Requests sent to the Cloud traverse multiple links and hops in the network. It is assumed that the fronthaul link is the slowest in terms of bitrate out of all these links. Only the fronthaul delay caused by packet size is considered as it is seen as the "bottleneck" in this network scenario. However, backbone/backhaul network can introduce significant "distance" delay – RTT. Let $\overline{d}$ be the average distance between Cloud DC and FNs which forward requests to this DC. The average-per-request transmission delay in the backhaul and the backbone equals

$$D_{comm}^{back} = \overline{d} \cdot 7.5\mu s/km \tag{14}$$

Total (average-per-request) communication delay D_{comm} is calculated based on the premise that all N requests go through the fronthaul while M requests are transmitted through the backhaul and backbone network:

$$D_{comm} = \frac{N \cdot D_{comm}^{front} + M \cdot D_{comm}^{back}}{N} \tag{15}$$

2) Computation: For estimating computational delays in the Fog tier of the network we assume that there is a queueing system. In [13] each FN is modeled as an M/M/1 queue. As it is assumed that FNs can balance the load between themselves, the queueing for each FN should not be independent – as long as there is an unutilized node the request should not "wait in queue" of another node. Therefore, in this work, the entire Fog tier is modeled as an M/M/n queue. $\lambda^{fog} = \frac{(N-M)}{T}$ is the average request arrival rate. $n = F$ is equal to the number of FNs. The service rate μ^{fog} is calculated as the ratio of FLOPS performance of an FN[1] and an average number of FLOPs that are needed to process a request sent to the Fog tier:

$$\mu^{fog} = \frac{f^{F_i} s^{F_i}}{\frac{\sum_{R_i \in \mathbb{R}^F} L^{R_i} \theta^{R_i}}{N-M}} \tag{16}$$

Then, the average delay D_{cp}^{fog} of a request caused by computing (and queueing) in the Fog tier equals:

$$D_{cp}^{fog} = \frac{C(n,\lambda^{fog}/\mu^{fog})}{n \cdot \mu^{fog} - \lambda^{fog}} + \frac{1}{\mu^{fog}} \tag{17}$$

where $C(n,\lambda^{fog}/\mu^{fog})$ is the Erlang C formula.

For computations in a Cloud DC, it is assumed that the computational resources are vast and there is never a need for queueing tasks. It is modeled as an M/M/∞ queue. So the average latency D_{cp}^{cld} depends only on the computation and is equal to $\frac{1}{\mu^{cld}}$, where:

$$\mu^{cld} = \frac{f^{cld} s^{cld}}{\frac{\sum_{R_i \in \mathbb{R}^C} L^{R_i} \theta^{R_i}}{M}} \tag{18}$$

is analogous to Eq. (16). It is worth noting that by assuming infinite computational resources at each Cloud DC the number of DCs in the network does not impact the performance. Total (average-per-request) computational delay D_{cp} is equal:

$$D_{cp} = \frac{(N-M) \cdot D_{cp}^{fog} + M \cdot D_{cp}^{cld}}{N} \tag{19}$$

Total (average-per-request) delay D_{tot} is calculated as a sum of computational and communicative delays:

$$D_{tot} = D_{cp} + D_{comm} \tag{20}$$

[1]All FNs are modeled to have the same frequency and FLOPS measure to allow the use of the M/M/n model.

IV. Results

We look at power consumption and delay in various scenarios for Fog without Cloud and for full Fog computing. Evaluation has been performed using GNU Octave [23].

A. Fog Without Cloud

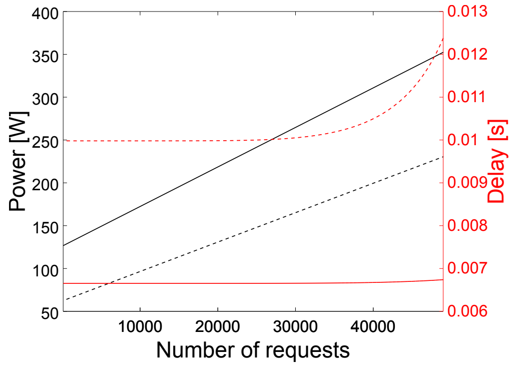
First, let us consider a network with $H = 20$ RRHs, $F = 10$ FNs and no possibility of offloading to Cloud DCs. Each request has size 1 MB, and values $\theta^R = 10$ and $o^R = 0.1$. The length of examined time period T is set to 60 s.

1) Computations: Let us examine the computational delay and power consumption in the FNs. We assume each FN is equipped with *Intel Core2 Duo E6850* as its Central Processing Unit (CPU). $s^{F_i} = 4$ as that is the maximum number of double precision FLOPs that can be calculated in a single cycle of this processor [24]. P_{act} and P_{idle} are parametrized using the power consumption model from [25]:

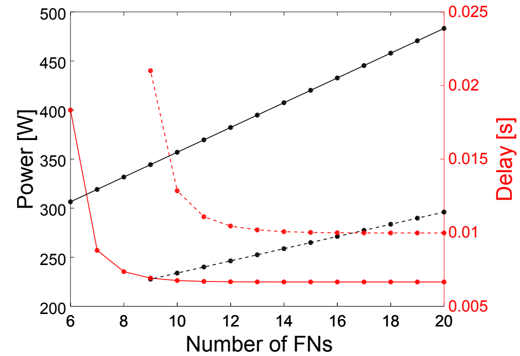
$$P_{cpu} = 8.4503\nu_{proc}V_{cpu}^2 f_{cpu} + 36.3851V_{cpu} - 33.9503 \tag{21}$$

where ν_{proc} is the total utilization of the processor and the units of P_{cpu}, V_{cpu}, and f_{cpu} are W, V, and GHz respectively. It is assumed that for the idle state $\nu_{proc} = 0$ and for the active state $\nu_{proc} = 1$. We consider power consumption of a FN at two different frequency-voltage levels measured in [25]: $f_{cpu} = 3.006$ GHz, $V_{cpu} = 1.28$ V later called as '3 GHz' and $f_{cpu} = 2.004$ GHz, $V_{cpu} = 1.104$ V later called as '2 GHz'. At 3 GHz $P_{act}^{F_i} = 54.241$ W and $P_{idle}^{F_i} = 12.6226$ W. At 2 GHz $P_{act}^{F_i} = 26.859$ W and $P_{idle}^{F_i} = 6.2189$ W. These values are then inserted into Eq. (9) to calculate the computational power consumption of FNs. Delay and power consumption of FNs are plotted against the number of computational requests in Fig. 4a. It shows that decreasing the CPU frequency lowers power consumption, but increases delay. Results in Fig. 4a are calculated for the number of FNs $F = 10$. Fig. 4b plots power consumption and delay against the number of FNs for number of requests $N = 50000$. It can be seen that higher number of FNs decreases delay. There are diminishing returns as the delay caused by queueing quickly disappears and processing delay does not change with number of FNs. Moreover, utilizing more FNs increases power consumption.

2) Communications: The communications delay in the fronthaul depends on the bit rate of fronthaul links and the size of the requests (and responses) as in Eq. (13). The term *home gateway* is used in [19] to describe access interface composed of several components, namely a processor plus memory, a Wide Area Network (WAN) interface and several Local Area Network (LAN) interfaces. We use it to model interfaces



(a) Dependency on number of requests, number of FNs $F = 10$.



(b) Dependency on number of FNs, number of requests $N = 50000$.

Fig. 4: Power consumption and delay related to computation in the Fog tier (dashed line – 2 GHz, solid line – 3 GHz).

TABLE II: Power consumption of networking equipment.

Equipment	Capacity	P_{act}	P_{idle}	γ_b
Gateways [19]				
1G EPON	1 Gb/s	3.3 W	3.0 W	0.3 nJ/bit
10/10G EPON	10 Gb/s	5.5 W	3.5 W	0.2 nJ/bit
Core routers [20]				
Juniper T1600	640 Gb/s	6572 W	5915 W	1.03 nJ/bit

of RRHs and FNs. Parameters of Ethernet Passive Optical Network (EPON) home gateways are detailed in Table II.

3) Overall performance: The values calculated in Sections IV-A1 and IV-A2 are added and presented in Fig. 5. Fig. 5 shows total (stemming from computations and transmission) power consumption and delay in a „Fog without Cloud" network. The following conclusions can be drawn. The average delay is low – hovering around 10-20 ms. There is also a visible trade-off between power consumption and delay – higher clock frequency and interface bandwidth results in shorter delay and greater power consumption. With that in mind, a combination of FNs running at 2 GHz frequency and 10 Gbit fronthaul outperforms FNs running at 3 GHz and 1 Gbit fronthaul in both estimated metrics at given traffic conditions.

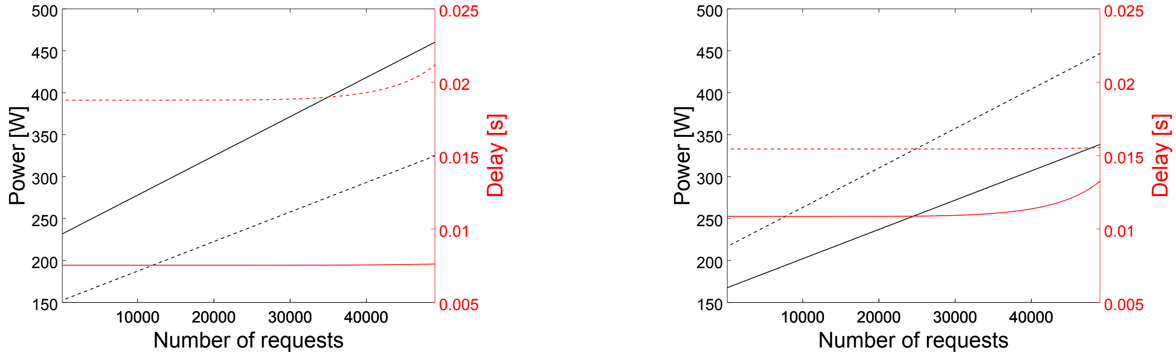
B. Full Fog Computing Network

Let us consider a network with $H = 20$ RRHs, $F = 10$ FNs and one Cloud DC. RRHs and FNs are connected through 1G EPON. FNs are connected through the 10G EPON backhaul to the core network. For estimating delays and power consumption occurring in the core network the following three distance scenarios are chosen. In *near* scenario the Cloud DC is located 100 km and 6 nodes away from the Fog tier of the network. The other scenarios are *medium* – 2000 km, 12 nodes and *far* – 8000 km, 18 nodes. There is a Juniper T1600 router (Table II) installed at each node. Cloud computational power efficiency is assumed to be 1 GFLOPS/W, its processors are running at 1.5 GHz frequency, and can calculate up to 32 FLOPs in a single cycle ($s^{cld} = 32$). FNs are assumed to work at 2 GHz with $s^{F_i} = 4$ FLOPs per cycle as in the previous subsection. Each request has size 1 MB, and values $\theta^R = 10$ and $o^R = 0.1$. The length of time period remains $T = 60$ s.

Let us vary the number of requests M which are sent to Cloud DC while the total number of offloaded requests stays the same ($N = 50000$). Power consumption and average delay are plotted in Fig. 6 against the fraction of requests $\frac{M}{N}$ sent to Cloud. There are a few interesting observations. First, the distance (physical and logical) to the Cloud DC plays a key role in determining whether offloading data all the way to the Cloud DC is beneficial. In the *near* scenario utilizing Cloud reduces both latency and power consumption. More interesting results can be seen in the *medium* scenario: performing computations in Cloud decreases power consumption while performing them in Fog results in lower average latency. Figs. 7a-d show how network performance changes when values of certain parameters are modified.

Power efficiency of Cloud DCs contributes significantly to total power consumption. As shown in Fig. 7a, if power efficiency is decreased (from 1 to 0.5 GFLOPS/W), then sending requests to Cloud is more expensive than computing them in Fog for all distance scenarios. On the other hand, at 10 GFLOPS/W (world-class efficiency [21]), sending requests to Cloud consumes less power (Fig. 7b). Delay does not depend on power efficiency.

The impact of computational complexity of requests θ^R on network performance is also studied. The results for $\theta^R = 1$ and $\theta^R = 100$ are shown in Figs. 7c and 7d. Corresponding share of total delay and power consumption spent on computation and communication is plotted in Fig. 8. For $\theta^R = 1$ computing requests in the Fog tier is faster and consumes less power than processing them in the Cloud tier. The delay and power consumption caused by data sent through the core network outweighs the fact that Cloud DCs are more efficient at computing. On the contrary, when the offloaded requests require heavy computations, it is beneficial to send these requests all the way to the Cloud as high θ^R means lower communication cost compared to computation cost.



(a) Dashed line: f=2 GHz, r=1 Gbps; solid: f=3 GHz, r=10 Gbps. (b) Dashed line: f=3 GHz, r=1 Gbps; solid: f=2 GHz, r=10 Gbps.

Fig. 5: Total power consumption and delay in "Fog without Cloud" network. f – FN clock frequency, r – fronthaul bit rate.

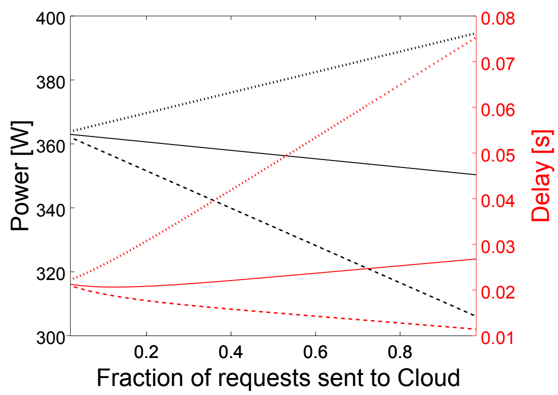


Fig. 6: Total power cons. and delay, full Fog computing. Dashed, solid, dotted lines – *near*, *medium*, *far* scenarios resp.

V. Conclusion

This work surveys several research articles which model the power consumption and delay in the Fog computing network. The common concept is that the computation and communications are modeled separately. Following up on the survey, the models for total power consumption and average delay in the Fog computing network are developed. The concept of computational complexity is introduced to differentiate offloading requests depending on how computationally intensive they are to process.

Impact of different parameters on power consumption and delay in the Fog computing is examined. It is observed that there is often a trade-off between delay and power consumption. E.g., Fog Nodes (FNs) working at higher frequency consume more power, but provide lower latency. High computational complexity of offloaded requests favors (as expected) processing in Cloud. Delay and power consumption caused by transmitting data through the core network is offset by the high processing speed and computational power efficiency. Conversely, requests which require relatively few operations to process are best served by nearby FNs.

Future work includes analyzing wireless transmission between Mobile Devices (MDs) and the Fog tier of the network. We are going to examine under what conditions it is beneficial to offload computations from the MDs. Moreover, we would like to confront our results with a small scale testbed.

References

[1] P. Cerwall *et al.*, "Ericsson mobility report," Ericsson, Tech. Rep., June 2018.

[2] Cisco, "Cisco global cloud index: Forecast and methodology, 2016–2021 white paper," Cisco, Tech. Rep., 2018.

[3] N. Chen, Y. Chen, Y. You, H. Ling, P. Liang, and R. Zimmermann, "Dynamic urban surveillance video stream processing using fog computing," in *Proc. BigMM*, Apr. 2016.

[4] J. K. Zao, T. T. Gan, C. K. You, S. J. R. Méndez, C. E. Chung, Y. T. Wang, T. Mullen, and T. P. Jung, "Augmented brain computer interaction based on fog computing and linked data," in *Proc. IE*, June 2014.

[5] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the Internet of Things," in *Proc. MCC*, Aug. 2012.

[6] OpenFog Consortium, "OpenFog reference architecture for fog computing, OPFRA001.020817," 2017.

[7] M. Peng, S. Yan, K. Zhang, and C. Wang, "Fog-computing-based radio access networks: issues and challenges," *IEEE Network*, vol. 30, no. 4, pp. 46–53, July 2016.

[8] M. Othman and S. Hailes, "Power conservation strategy for mobile computers using load sharing," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 2, no. 1, pp. 44–51, Jan. 1998.

[9] C. Mouradian, D. Naboulsi, S. Yangui, R. H. Glitho, M. J. Morrow, and P. A. Polakos, "A comprehensive survey on fog computing: State-of-the-art and research challenges," *IEEE Communications Surveys Tutorials*, vol. 20, no. 1, pp. 416–464, Q1 2018.

[10] T. N. Gia, M. Jiang, A. Rahmani, T. Westerlund, P. Liljeberg, and H. Tenhunen, "Fog computing in healthcare Internet of Things: A case study on ECG feature extraction," in *Proc. WNM*, Oct. 2015.

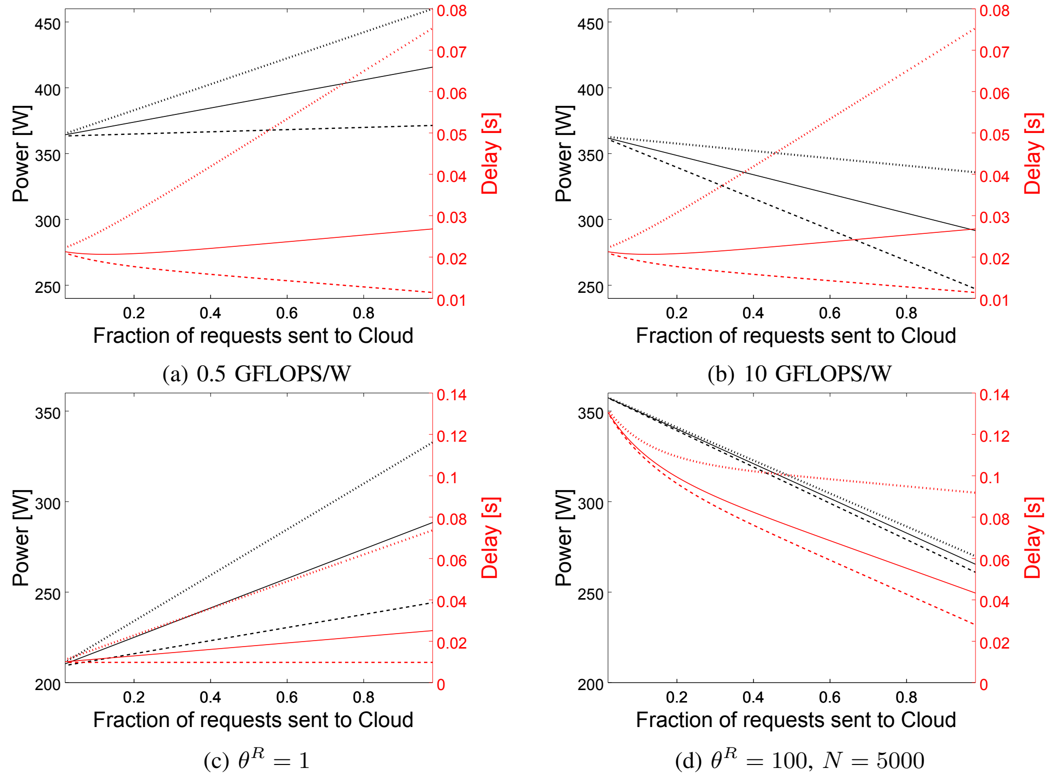
[11] T. Q. Dinh, J. Tang, Q. D. La, and T. Q. S. Quek, "Offloading in mobile edge computing: Task allocation and computational frequency scaling," *IEEE Transactions on Communications*, vol. 65, no. 8, pp. 3571–3584, Aug. 2017.

[12] L. Liu, Z. Chang, and X. Guo, "Socially aware dynamic computation offloading scheme for fog computing system with energy harvesting devices," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1869–1879, June 2018.

[13] 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. 1171–1181, Dec. 2016.

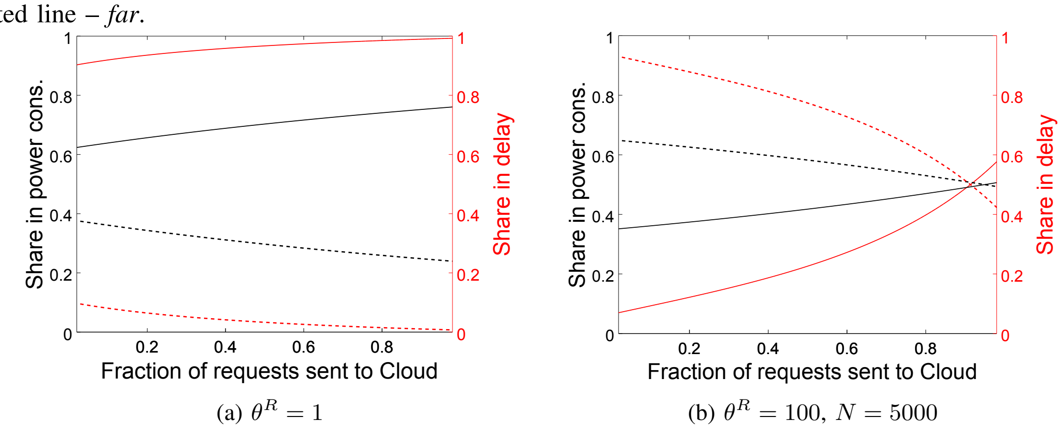
[14] F. Jalali, K. Hinton, R. Ayre, T. Alpcan, and R. S. Tucker, "Fog computing may help to save energy in cloud computing," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1728–1739, May 2016.

[15] S. Sarkar and S. Misra, "Theoretical modelling of fog computing: a green computing paradigm to support IoT applications," *IET Networks*, vol. 5, no. 2, pp. 23–29, Mar. 2016.



(a) 0.5 GFLOPS/W
(b) 10 GFLOPS/W
(c) $\theta^R = 1$
(d) $\theta^R = 100,\ N = 5000$

Fig. 7: Total power consumption and delay in full Fog computing network depending on fraction of requests sent to Cloud, Cloud power efficiency (GFLOPS/W) and computational complexity of requests θ^R. Dashed line – *near* scenario, solid line – *medium*, dotted line – *far*.



(a) $\theta^R = 1$
(b) $\theta^R = 100,\ N = 5000$

Fig. 8: Share of computation (dashed line) and communication (solid line) in power consumption and delay in full Fog computing network, *medium* scenario.

[16] S. Sarkar, S. Chatterjee, and S. Misra, "Assessment of the suitability of fog computing in the context of internet of things," *IEEE Transactions on Cloud Computing*, vol. 6, no. 1, pp. 46–59, Jan. 2018.

[17] S. J. Edwards, S. D. Forsyth, J. Stanback, and A. Saremba, "Standard: Portable Game Notation specification and implementation guide," 1994, http://www.saremba.de/chessgml/standards/pgn/pgn-complete.htm.

[18] M. Olbrich, F. Nadolni, F. Idzikowski, and H. Woesner, "Measurements of path characteristics in PlanetLab," Technical University Berlin, Tech. Rep. TKN-09-005, July 2009.

[19] P. Bertoldi, "EU code of conduct on energy consumption of broadband equipment: Version 6," 2017.

[20] W. Van Heddeghem *et al.*, "Evaluation of power rating of core network equipment in practical deployments," in *Proc. OnlineGreenComm*, Sep. 2012.

[21] W. Feng and T. Scogland, "Green500 list for November 2018," 2018. [Online]. Available: https://www.top500.org/green500/lists/2018/11/

[22] G. Almes, S. Kalidindi, and M. Zekauskas, "A round-trip delay metric for IPPM," 1999, RFC 2681.

[23] "GNU octave." [Online]. Available: www.gnu.org/software/octave/

[24] R. Dolbeau, "Theoretical peak FLOPS per instruction set: a tutorial," *The Journal of Supercomputing*, vol. 74, no. 3, pp. 1341–1377, Mar 2018.

[25] S. Park, J. Park, D. Shin, Y. Wang, Q. Xie, M. Pedram, and N. Chang, "Accurate modeling of the delay and energy overhead of dynamic voltage and frequency scaling in modern microprocessors," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 32, no. 5, pp. 695–708, May 2013.