Draft Manuscript

The environmental footprint of a distributed cloud storage

Lorenzo Posani

Abstract Every time we access a file on the Cloud, a chain of routing servers is activated to transfer the payload from the storage facility to the user's device, and back. The data center, being densely packed and highly-performant, is hardly optimized in terms of energy consumption and environmental impact. The ICT (information communication technologies) ecosystem is estimated to be responsible for the 10% of the full worldwide energy consumption - equivalent to Germany and Japan total energy demand. However, being a fast-inflating and almost unregulated market, the environmental footprint caused by data storage and transfer on the internet shows no signs of slowing down. In this paper, we analyze a reversal paradigm for cloud storage (implemented by the startup Cubbit) in which data are stored and distributed over a network of p2p-interacting single board computers. We compare Cubbit to the traditional server-based solution in terms of environmental footprint and power/usage efficiency. We demonstrate how a distributed architecture is beneficial for impact reduction on both sides of communication and storage. First, being virtualized and distributed upon small single-board computers, storage is far more efficient than data centers racks and does not need any additional cooling. Secondly, the distributed architecture can leverage proximity of files to the user (e.g. within the same city) to minimize the consumption due to the powered route from the virtualized storage facility to the user's device. We hereby quantify both these effects using the public internet model published by Baliga et al. (Proceedings of the IEEE 99.1, 2011), and compare the estimation with the same model applied to a server-based cloud storage (e.g. Dropbox). Results show that a remarkable reduction in terms of energy consumption is obtained on both storage (down to 8% of consumptions) and transfers (c.a. 50% of consumptions).

Lorenzo Posani
· Laboratoire de Physique Statistique, Ecole Normale Supérieure and CNRS UMR 8550, PSL Research, Paris Sorbonne UPMC, 24 rue Lhomond, 75005 Paris, France
· Cubbit, Via San Vitale 39, 40125, Italy - www.cubbit.io

For correspondence: lorenzo.posani@cubbit.net

1 Introduction

Over the last decades, the rise of environmentalism and the acknowledgment of climate change by the general audience has driven the law regulations of most western countries towards an increasing awareness of consumptions and efficiency boundaries. The average per-device consumption of household appliances (fridge, cooling, etc.) has consistently decreased in the last 20 years [2]. Despite the constant improvement of the average household efficiency, there is an underestimated factor that sensibly contributes to the environmental impact of our daily lives: our use of the Information-Communication Technology (ICT) ecosystem or, in other words, our digital life. A middle-range estimation of the total impact of the ICT ecosystem approaches 1500 TWh of annual consumption [8, 9], which roughly amounts for the 10% of the world energy consumption, more than total energy demand of Germany and Japan, taken together.

The computation of the pro-capita consumption shows that the fact of owning and using a smartphone constitutes, without considering the charging costs, an equivalent energy consumption of owning an additional fridge [9]. Contrary to the electronics market, however, the footprint caused by our online life is much less tangible and, as a consequence, much less opposed. That, combined with the fast-increasing trend of online presence and internet-accessing devices pro capita, results in an increasing and mostly uncharged environmental impact that shows no signs of slowing down [1].To gain an intuition of the impact of the digital life, one has to consider that any time a video is streamed from Youtube servers to an iPad, or a photo is accessed on Google Photos or Dropbox, the whole infrastructure that separates the final user to the corporate data center, and the data center itself, has to be powered to reliably transmit information in both directions.

From this point of view, the process of transmitting information can be orders of magnitude more demanding, in terms of energy consumption, than the storage itself, depending on the relative location of the exchanging nodes. In this document, we will analyze, using a model inspired by the one described in [7], the energy consumption and consequent environmental impact of the cloud storage services, and compare it to an alternative setup where data are stored on low-consumption devices in users houses, locally distributed in order to minimize the distance between communicating nodes.

	Equipment	Capacity	Consumption
Data Center storage rack	HP 8100 EVA	604.8 Tb	4.9 kW
Data Center gateway router	Juniper MX-960	660 Gb/s	5.1 kW
Ethernet Switch	Cisco 6509	160 Gb/s	3.8 kW
BNG	Juniper E320	60 Gb/s	3.3 kW
Provider Edge	Cisco 12816	160 Gb/s	4.21 kW
Core router	Cisco CRS-1	640 Gb/s	10.9 kW
WDM (800 km)	Fujitsu 7700	40 Gb/s	136 W/channel

Table 1: **Equipments**

2 Analysis of Server-based Cloud consumptions

The energy consumption of a cloud storage service can be divided into two main factors:

(a) the cost of storing the data, i.e. powering and cooling the data center (Storage consumption)

(b) the cost of sending the data from the user to the server and back (Transfer consumption)

While the first can be easily estimated from technical specifications of storage equipment, the second needs a more detailed analysis that takes into account the public internet infrastructure and the distance between user and server locations. For both these estimations we refer to [7]. Briefly, the consumption model is composed of several addends and coefficients which account for the multiplicity of involved devices as well as other factors, such as redundancy, cooling, overbooking (see below).

To delineate the calculation, we start from the storage consumption, i.e. the average power, expressed in W/GB, necessary to store the payload in *hot storage*. This is estimated from the specs resumed in table 1, considering a factor 1.5 for cooling and a factor 2 for redundancy of stored data, as specified in [7]:

$$\mathrm{P}^{\mathrm{storage}}_{\mathrm{server}} = 1.5 \times 2 \times \frac{4.9\ \mathrm{kW}}{604.8\ \mathrm{Tb}} = 0.195\ \frac{\mathrm{W}}{\mathrm{GB}} \tag{1}$$

Similarly, we compute the transfer energy, expressed in J/GB, following the public internet model of [7]. The analysis relies on the definition of the consumption per bit, which is computed by dividing the operating power (W) for the total transfer capacity (Gb/s), resulting in a Joule/bit measure, then converted in J/GB.

These units are taken from the manufacturer specs sheet, shown in table 1. These quantities are combined with a set of coefficients that reflect the redundancy of the packet transmission, the under-operating regime of the infrastructure and the cooling energy, as well as the multiplicity of some devices in a single transmission (e.g. two ethernet switches at entry points plus another one inside the data center). The average distance between core routers on the network is estimated of c.a. 800Km. For full description of coefficients and estimations we refer to [7]

$$\mathrm{E}_{\text{server}}^{\text{transfer}} = 6 \times \left(3\frac{P_{es}}{C_{es}} + \frac{P_{bg}}{C_{bg}} + \frac{P_g}{C_g} + 2\frac{P_{pe}}{C_{pe}} + 18\frac{P_c}{C_c} + 4\frac{P_w}{C_w} \right) \simeq 23.9\ \frac{\mathrm{kJ}}{\mathrm{GB}} \tag{2}$$

Where the factor of six accounts for redundancy ($\times$ 2), cooling and other overheads ($\times$ 1.5), and the fact that today's network typically operate at under 50% utilization ($\times$ 2); the addends represent, in order, the ethernet switch, the broadband gateway, the data center gateway, the provider edge router, the core network, and the relay optical fiber transmission. The detailed analysis of prefactors can be found in [7]. Briefly, the factor 3 in the ethernet switch accounts for the 2 routers involved in the acces to the public internet and for the router in the data center; the factor 18 in the core network accounts for an average of 9 hops (2 baseline + 7 for the 800km distance between core nodes) of internet packets from source to destination, times 2 for the redundancy.

3 Analysis of Distributed Cloud consumptions

The distributed architecture of the Cubbit network relies on the same public internet infrastructure delineated in the previous chapter. In the distributed paradigm, there are two key differences with the server-based cloud storage:

(a) The low energy consumption of un-cooled ARM devices (Raspberry Pi)

(b) The geographical locality of uploaded data

We consider a network of ARM devices (raspberry pi 3) each connected to an hard disk of 1TB and placed in a user house with ISP internet connection. The files are considered as already uploaded on the network, with a redundancy storage factor of 1.5 (Reed Solomon erasure coding with 24+12 redundancy shards [10, 11]). As done in the previous chapter, we analyze the consumption in terms of storage (W/GB) and transfer (J/GB) energy.

For what concerns the storage maintenance, the raspberry pi 3 has a baseline idle consumption of 1.4 W [5], and a single-core peak usage of 2.4 W [5]. We therefore can make a conservative assumption and estimate an average power of

2 W. We then consider an average consumption of a 1TB hard disk of 8W. The storage energy consumption of the Cubbit network is therefore

$$\mathrm{P}_{\mathrm{cubbit}}^{\mathrm{storage}} = 1.5 \times \frac{2+8\ \mathrm{W}}{1\ \mathrm{TB}} = 0.015\ \frac{\mathrm{W}}{\mathrm{GB}} \tag{3}$$

Thanks to the locality of stored data, i.e. shards of the distributed payloads are located within 80 km from the user's access point (distribution of the payload is under the control of AI optimization routines [11]), we can assume an average number of 2 packet hops in core network routers. This lowers down the factor 18 in Eq. 4 to a factor 4, accounting for two core hops and the redundancy of the packets on the network (factor 2). For the same reason, the 800km-relay consumption P_w is not taken into account. With respect to Eq. 4 we also ignore all the data-center-specific terms: one of the three ethernet switches and the data center gateway. We however need to consider an additional BNG, since we are considering a p2p transfer between endpoints that are located behind an ISP network. The transfer energy therefore reads

$$\mathrm{E}_{\mathrm{cubbit}}^{\mathrm{transfer}} = 6 \times \left(2\frac{P_{es}}{C_{es}} + 2\frac{P_{bg}}{C_{bg}} + 2\frac{P_{pe}}{C_{pe}} + 4\frac{P_c}{C_c} \right) \simeq 11.9\ \frac{\mathrm{kJ}}{\mathrm{GB}} \tag{4}$$

4 Comparison between server-based and distributed architectures

The difference between a distributed and a centralized architecture in terms of environmental impact can be visualized by computing the difference of the two total consumptions (storage and transfer), in a year, for different use types, see Fig. 1.

As a first estimation, let us consider a consumer user with 1 TB of cloud storage and very limited daily access (less than 1 Gb per day). This setup is typical of, for example, a backup on Dropbox pro. The easy computation shows that a server-based architecture for a 1TB plan has an annual consumption of 1700 kWh. On the contrary, the same storage provided by a distributed architecture has a limited cost of 131 kWh per year. All considered, it means that, for a typical 1 TB storage plan with little daily usage, choosing a distributed cloud architecture saves ca 1570 kWh per user per year (see Fig. 2), which is grossly equivalent to the annual consumption of **a small (1 pearson) apartment** [4].

Likewise, we observe that the difference is even more evident when data are often accessed and updated. Thanks to the locality of data in the distributed architecture, we can estimate an annual saving, for a small video-streaming service that stores 10 TB and accesses 1 TB of data every day (2000 views of a 500MB movie), of c.a. 17000 kWh, grossly equivalent to the full annual consumption of a small office [3]. Finally, if we speculate about the overall data volume of a global

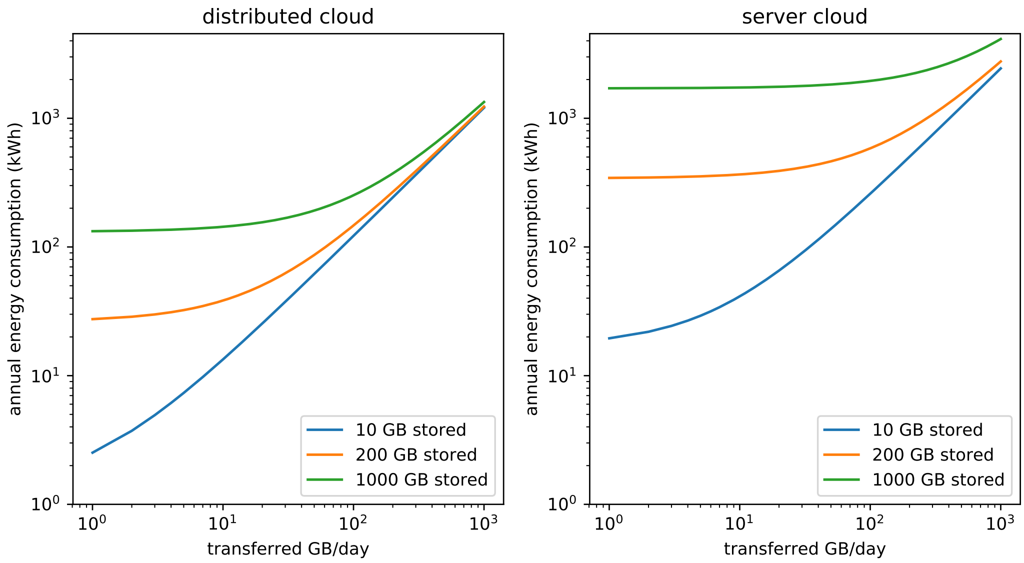

Fig. 1: Annual energy consumption of distributed (left) and server-based (right) cloud storage service.

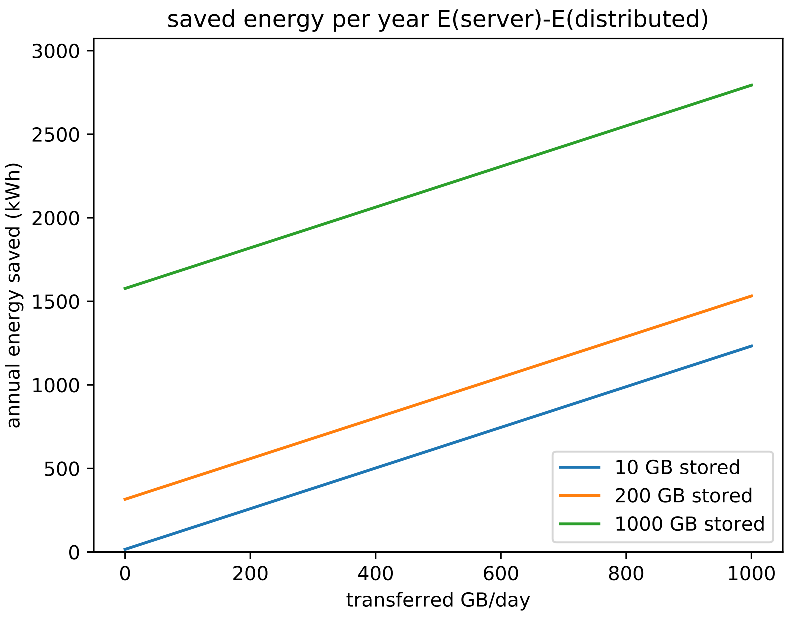

Fig. 2: Annual saved energy choosing a distributed cloud storage service over a server-based one

cloud service like Dropbox, values rise dramatically. Of course such interpolations have to be taken with due caution, since estimations are based on undisclosed values. As an example let us consider the Dropbox which can roughly be estimated to consist of 600 Millions users. The average conversion rate from free to pro plan is around 3%. This results in a theoretical data volume of ca. 19.2 Millions of stored TB. Considering a factor 5 of overbooking, it gives an estimation of 3.8 millions of TB of effective storage. As an average daily usage, we can estimate that each user accesses 10 Mega Byte of files from the cloud. That gives ca 6000 Tera Bytes of daily-accessed data. If we plug these estimations in our model, we obtain a total saved annual energy, using a distributed architecture instead of a centralized one, of ca 6,000,000,000 kWh, equivalent to the full annual electricity consumption of the entire **Luxemburg** [6].

References

1. Cisco vni global fixed and mobile internet traffic forecasts. https://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/index.html#cloud-forecast.
2. Energy efficiency and energy consumption in the household sector. https://www.eea.europa.eu/data-and-maps/indicators/energy-efficiency-and-energy-consumption-5/assessment.
3. G20 energy efficiency indicators: Unit electricity consumption of services per employee. http://g20-energy-efficiency.enerdata.net/indicators/unit-electricity-consumption-of-services-per-employee.html.
4. How much energy do my household appliances use? https://www.energuide.be/en/questions-answers/how-much-energy-do-my-household-appliances-use/71/.
5. Raspberry pi power consumption. https://www.pidramble.com/wiki/benchmarks/power-consumption.
6. Wikipedia list of countries by electricity consumption. https://en.wikipedia.org/wiki/List_of_countries_by_electricity_consumption.
7. Jayant Baliga, Robert WA Ayre, Kerry Hinton, and Rodney S Tucker. Green cloud computing: Balancing energy in processing, storage, and transport. Proceedings of the IEEE, 99(1):149–167, 2011.
8. Greenpeace International. How clean is your cloud? catalysing an energy revolution. Technical report, 2012.
9. Mark P Mills. The cloud begins with coal. Digital Power Group, 2013.

10. Matteo Monti, Steen Rasmussen, Marco Moschettini, and Lorenzo Posani. An alternative information plan. Technical report, Working paper Santa Fe Institute, 2017.

11. Lorenzo Posani, Marco Moschettini, and Alessio Paccoia. Cubbit: a distributed, crowd-sourced cloud storage service. Invited Talk at CERN Workshop on Cloud Storage Services (CS3 2018), 2018.