



HAL
open science

Energy consumption of data transfer: Intensity indicators versus absolute estimates

Gaël Guennebaud, Aurélie Bugeau

► **To cite this version:**

Gaël Guennebaud, Aurélie Bugeau. Energy consumption of data transfer: Intensity indicators versus absolute estimates. *Journal of Industrial Ecology*, 2024, 10.1111/jiec.13513 . hal-04631084

HAL Id: hal-04631084

<https://hal.science/hal-04631084v1>

Submitted on 1 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Energy consumption of data transfer: intensity indicators versus absolute estimates

Gaël Guennebaud¹ and Aurélie Bugeau²

¹Inria, Univ. Bordeaux, Talence, France

²Univ. Bordeaux, CNRS, LaBRI, UMR 5800, IUF, Talence, France

Abstract

The assessment of energy consumption of data traffic for Internet services usually relies on energy intensity figures (in Wh/GB). In this paper, we argue against using these indicators for evaluating the evolution of energy consumption of data transmission induced by changes in Internet usage. We describe a model that estimates global impacts for different scenarios of Internet usages and technological hypotheses, and show that it can overcome some limitations of intensity indicators. We experiment the model on four use-cases: basic usage, video streaming, large downloads, and video-conferencing. Results show that increasing the resolution of videos does increase the total energy consumption while misleadingly decreasing the power intensity indicator at the same time. In other words, a more efficient network does not necessarily mean less energy consumption.

Keywords— Internet absolute energy consumption, Internet usages, Absolute estimates, peak usage, Access network

1 INTRODUCTION

The Information and Communication Technology (ICT) sector represents 2 to 4% of global greenhouse gas emissions [9] and all studies agree on an important growth in future decades. Questioning which levers could limit environmental impacts of this growth is therefore necessary. Part of GHG emissions come from data transmission over the Internet. Existing works studying the impacts on climate change of data transmission usually strive to estimate the overall yearly energy consumption of the Internet (in TWh/year) from which energy intensity estimates of data transmission (in Wh/GB) are extracted. A recent review [6] of these works shows high variability in the results and even inconsistencies between overall energy consumption, energy intensity, and Internet traffic. Those variations are likely explained by differences in Internet modeling, system boundaries, hypotheses, and methodologies.

In this paper, we argue that energy intensity estimates are not adapted to study the evolution of impacts induced by changes in the amount of data transiting through the Internet (section 3). To analyze evolution of absolute impacts, we proposed in [10] a new methodology, summarized in section 4, whose central idea is to consider a given Internet usage at the whole scale of an appropriately chosen territory, and evaluate the impact of this usage, or variants of this usage, on the IT infrastructure. This is accomplished through a parametric bottom-up network modeling of a simplified network infrastructure. Starting from a *reference* infrastructure, our model automatically scales the required hardware according to peak usage scenarios, from which absolute power and energy consumption can be estimated and compared to the reference or other scenarios.

We study four use-cases at the scale of the France territory: basic usage, over-the-top video on demand (OTT VoD), large downloads and video-calls. They are presented in section 5 with the design of the infrastructure. To support the video-calls use-case, we extended the previous model to explicitly account for upstream transfers. All the use-cases imply high data traffic playing an important role on the dimensioning of current infrastructures.

Through our experiments (section 6), we propose an evaluation of the impacts that would result from changes in the use-cases, e.g. limiting video resolution or reducing downloading speed. In total, we evaluate ten scenarios chosen to illustrate some levers on each use-case. Our model enables analyzing which network equipment is most impacted by each parameter of the scenario. The proposed methodology is therefore a first step towards a better understanding of the consequences of political, industrial, or societal decisions on infrastructure sizing. In particular, our results show that using a simple efficiency indicator may lead to misleading decisions.

2 Related works

This section presents related works on energy consumption of data transmission, focusing on network models and alternative approaches to estimating energy intensity of each part of the ICT infrastructure.

baliga2009proposedoneofthefirstbottom-upnetworkinfrastructuremodel.Thetotalpoweristhesumofthepowerofeachequipme and the linear component, which is assumed to be proportional to the current throughput (in bit/s). They observed that for such fixed network elements, this proportional part is small with $P_{idle} > 0.9P_{max}$. This observation is also confirmed by power_{model}.Withsuchanafinepowermodel,allocatingtheproportionalpartboilsdowntoasimplevolume-basedallocation.Toaccommodatefortheidlepower,theyproposeddifferentialallocationstrategies,particularlyonebasedonrelativethroughputprofile.Theidlepowerisfirstequallyspreadtoeachline,andthendistributedtothepotentiallymultipleusersanddevicesusingthisline, and duration-basedforthecustomerpremisesequipment(CPE)andaccessnetworkequipment,andvolume-basedforthecorenetwork.All powerforagivenusageorservice.

To better understand network energy consumption or GHG emissions, several studies focused on a narrow and specific use-case. For instance, schien2012 used traceroute data to estimate the number and type of network devices involved in digital media transmission. coroama2013considereda40MbpsvideoconferencingtransmissionbetweenSwitzerlandandJ

3 Limits of energy intensity estimates

Energy intensity estimates are frequently used to assess the energy consumption, or environmental impacts, of data transmission of a given Internet service such as, for instance, video-streaming. Commonly raised questions are: "What is the energy consumption or the environmental impacts of transferring one GB of data?" or "What is the energy consumption of watching one hour of video streaming?". Answering these questions might be legit to quickly get a rough estimate ahead of a more detailed investigation, or to allocate the overall shared footprint across different usages retrospectively for accounting purposes. However, when the goal is to guide the reduction of these impacts, such intensity indicators exhibit some limitations that we review in this section.

Global averages Firstly, we argue that attempting to answer such a question in a general manner, inevitably leads to severe oversimplifications of the physical reality. It depends on numerous variables that go way beyond distinguishing the core, fixed-access, and mobile-access networks. Some other variable examples include the technological maturity of the considered network (from old energy-intensive equipment to the newest generation) or the actual route taken by the data: a data-intensive service hosted in the US but used in Europe will have a different impact on the network than another service hosted in the same city as its primary users. Choosing levers of action to reduce impacts at the scale of a territory requires going further than averaging.

Poor proportionality Secondly, unlike what their units convey, answers to these questions exhibit a poor proportionality with the physical reality. Indeed, one can quickly come to the false conclusion that, e.g., reducing by two the amount of data of a given usage will reduce by two its impacts. Their use is frequent and misleading both for a short or long-term point of view [18]. On the shorter term because the infrastructure is permanently switched on, and the volume of data passing through it at a given time has little influence on the power consumption of the equipment (especially for fixed network equipment). In the longer term, one could expect a correlation because: the traffic volume increases, the traffic peak is expected to increase too, yielding to an increase in the infrastructure equipment, and thus an increase of the overall consumption of the infrastructure [17]. Conversely, if the traffic is maintained or decreased, oldest equipment might be replaced by smaller and more efficient ones when renewed. However, this long-term correlation is only partial because i) only a subset of network hardware is subject to such correlation with peak demands (because of the equipments required to "just" connect houses and datacenters), ii) the energy efficiency of such equipment improves quickly over time (hence absorbing part of the demand increase), and iii) two identical volumes of transferred data might have very distinct effects on local traffic peaks (because of different bitrates, different routes, or different *burstiness* [11]). Quantifying this amount of correlation is one of the questions our proposal aims to address, and our results provide insights on i) and iii).

Hiding absolute impacts Thirdly, intensity numbers (in Wh/GB or gCO_{2e}/GB) are *efficiency* indicators that hide the true *absolute* energy consumption or absolute impacts. In our context, this observation combined with the aforementioned second point yields a paradox: increasing the total amount of traffic increases load percentage and enables scaling gains. Both lead to an improvement (i.e, a decrease) of those efficiency indicators, while the absolute impacts increase. In contrast, sobriety behaviors are certain to maintain or decrease absolute impacts even though they might degrade those efficiency indicators.

Allocation The Internet network is made of many equipment that are shared by different usages. Attributing the impacts of these equipment to a given use-case with allocation factors is tricky and makes difficult the analysis of the impacts of each use-case over the network infrastructure. The same difficulty arose for end-user terminals such as smartphones or laptops that are manufactured and bought to serve multiple purposes. Indeed, allocating the manufacturing impacts of smartphones to a specific usage such as video streaming in a physically realistic

manner is an impossible task that can only be achieved through arbitrary allocating rules hiding the complexity of the real dynamics.

4 Proposed consequential approach

In this section, we present a general overview of our methodology. We recall that our goal is to estimate the impacts of a given service or use-case through its pressure on the dimensioning of a *hypothetical* infrastructure composed of datacenters, network and customer premises equipment (CPE). This section remains general. Actual boundary choices and use-cases for testing this methodology are later detailed in section 5. Our methodology is based on a bottom-up model that instances and scales-up the infrastructure to satisfy different *scenarios* (i.e., use-cases and hypotheses). We can then estimate its *absolute* operational energy consumption and other environmental impacts. One has to define at least two scenarios, the one we want to assess, and a *reference* one that will be used to compare the absolute impacts estimated for the two scenarios. In practice, one can also compare different hypotheses for the same use-case, hence enabling a better understanding of the consequences of different choices on the different levels of the infrastructure.

Step 1 - Use-case

We first need to define the service or use-case that we aim to model, evaluate, and analyze. This comes with the associated main parameters and variables, and their range of values that will be explored (e.g., video resolutions, number of viewers, server localization, file sizes, frequencies, etc.).

Step 2 - Boundary

This step covers two aspects. First, which parts of the Internet infrastructure are included: datacenters, core, edge, fixed and/or radio network, fiber and/or copper, CPE, etc. Second, which geographical territory: a city, a country, the world? It is important to consider the scale of a large-enough *territory* to be representative of the service/use-case at hand. The choice of a territory might be dictated by the purpose of the evaluation, e.g., one might be interested in evaluating a service for a given country.

Step 3 - Design of the parametric infrastructure model

This step consists in designing the parametric model that generates infrastructures according to some dimensioning variables. To this end, a *base* infrastructure is required, e.g., every home and datacenters of the considered territory must be connected with the capacity to exchange some bits, the radio network must cover 99% of the territory, every user of our scenarios possesses at least one smartphone, tablet or laptop, etc. The model is designed to scale-up and to cover the range of use-cases and boundaries defined previously.

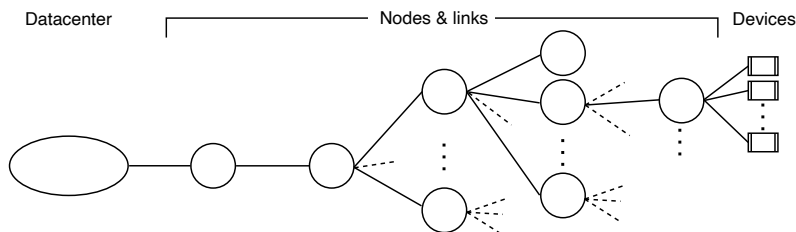


Figure 1: Abstract tree representation of the infrastructure. In this example, the tree goes from a main datacenter to many devices through nodes and links. A node can, for instance, represent an IXP that includes a CDN or routers. The detailed infrastructure for our use-case is presented in Figure 4.

Following [10], we use a simplified tree representation (Figure 1), as a proof of concept. It goes from the home-routers up to the main datacenters hosting the considered service, and passing through nodes that are converging/splitting points of the fixed-access, edge, and core network layers. Congestion might occur in the nodes. Edges represent fiber links which might either include passive-only equipment, but also active equipment for long-distance hops. Its main dimensioning variables are the bandwidth capacities required at the different nodes and links of the tree. At each node and link l , the equipment is scaled up to the minimal quantity enabling:

- i) a connection to every subscribers (households),
- ii) a bandwidth capacity equal or greater to R_l (in bit/s) which has to be adjusted to the peak-access-rate estimated at the node or link l for the given scenario (see Step-4). Throughout this paper, peak-rates are assumed to refer to averaged traffic rates over a few seconds.

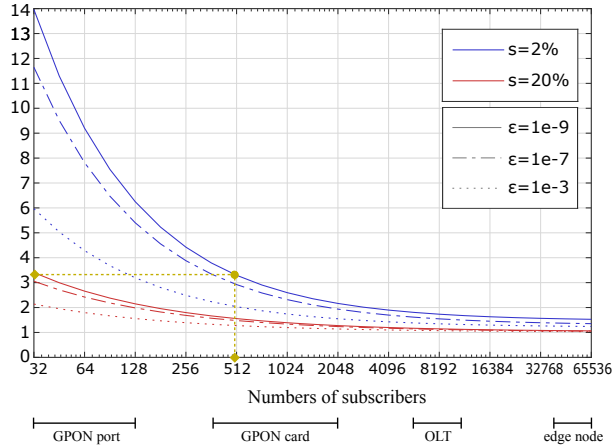


Figure 2: Ratio between the worst-case number of active subscribers and the global average number of active subscribers as a function of the number of subscribers below a given node. Typical ranges are shown for some nodes. It is reported for two different global percentages of active subscribers (2% and 20%) and three different ϵ tolerance values. The same ratio also apply to ratio of active *users*. Numerical values are available in Table “Figure 2” of Supporting Information S1.

This quantity is then multiplied by the redundancy factor η . The global power consumption is estimated as the sum over all equipment and facilities of the infrastructure. To this end, each equipment must be associated with a power consumption profile, i.e., a static (or idle) power (in W) and a power intensity factor (in W/Gbps) for the dynamic power consumption part which is assumed to be proportional to the actual traffic. Because we scale the infrastructure to peak needs, we can better understand the physical reality and bottlenecks behind power consumption.

Step 4 - Scenario evaluation and peak demand modeling

Finally, for evaluation purposes, each scenario needs to be translated to dimensioning variables exposed by the model defined in the previous step. This mainly requires estimating the capacities R_i from peak-access-rates estimated at every node and link of the tree. This step usually embeds a growing margin factor allowing to anticipate exceptional traffic peaks, and for future growth provisioning.

Our scenarios are constructed based on global average statistics such as the percentage of “active” users (e.g., percentage of active subscribers, or the percentage of simultaneous VoD watchers). Whereas globally defined statistic averages would be sufficient when considering a large pool of inhabitants, they cannot be used to reflect worst-case scenario near the leaves of the tree where some equipment are shared by few dozens to a few hundreds of inhabitants only (Figure 1). To be more realistic, we adapt our model to consider a non homogeneous repartition of inhabitants and active users both in space (over shared nodes) and time.

To account for worst-cases, we propose to consider their respective distributions, say d^n , for a given sub-number of inhabitants n . To neglect the most unlikely occurrences, we use the smallest quantity q such that the probability of having an occurrence x greater than q is extremely low, i.e., $d^n(x > q) < \epsilon$. We used $\epsilon = 10^{-9}$. Figure 2 illustrates the relative effect of this *worst-case* approach compared to using global mean quantities for two different averages of active users. For instance, with a global average of 2% of active users within a pool of 500 inhabitants, we have on average 10 active users per group only, but owing to non-homogeneity we expect that some groups will exhibit up to $\times 3.4$ more active users in the worse-case. We refer to ICT4S for more technical details.

5 Design of the infrastructure

We applied the methodology described in the previous section to four uses-cases: *basic* use of Internet (the *reference*), over-the-top (OTT) streaming, large download, and video-calls. For technical details on equipment and their consumption, and the choice of active users, we refer the reader to ICT4S.

5.1 Description of use-cases and boundaries

We now describe the four use-cases and scenarios considered. Table 1 summarizes all parameters chosen for each scenario.

Table 1: Scenario parameters (the OTT, VC and DL use-cases are added on top of the basic use-case).

Use-case	Scenario	Users	Average Peak rate	% content stored in CDN	% content from oversea	ϵ
basic	Reference	2% households	↓10 Mbps	0	33%	10^{-9}
OTT	HD	20% population	↓3 Mbps	80%	100%	10^{-9}
	FHD	20% population	↓5 Mbps	80%	100%	10^{-9}
	UHD	20% population	↓16 Mbps	80%	100%	10^{-9}
	UHD++	20% population	↓27 Mbps	80%	100%	10^{-9}
VC	VC-5%	5% population	↑ 1, ↓1 Mbps	0	10%	10^{-9}
	VC-10%	10% population	↑ 1, ↓1 Mbps	0	10%	10^{-9}
	VC-20%	20% population	↑ 1, ↓1 Mbps	0	10%	10^{-9}
DL	DL-strict	3% households	↓200 Mbps	95%	33%	10^{-7}
DL	DL-relaxed	3% households	↓200 Mbps	95%	33%	10^{-3}

Reference For our *reference* scenario, we considered traffic representing a "minimalist and sober" communication and information-sharing use of the Internet, from which we can assess the full incremental impacts of a given use-case, or, in other words, the *pressure* that a given use-case puts on the dimensioning of the infrastructure. It is by no means meant to represent the current trend of the French infrastructure, but our public web-application¹ does permit users to experiment with a custom reference infrastructure to serve other purposes. At peak hours, the basic traffic is represented as a global percentage of 2% active customers and a per-customer speed rate of 10 Mbps. This pair of values has been arbitrarily chosen to reproduce an average of 200 kbps per-customer as in France in 2013. They are rather high for a "minimalist and sober" use of Internet, but they can be considered as conservative. We assume that the basic peak traffic occurs in addition to the one of other use-cases, i.e. that traffic peaks are fully correlated, which is a conservative choice. Volume-wise, we assume 2GB of data per subscriber per month for this basic use-case.

Over-the-top video on demand (OTT) For video on demand, we limit ourselves to over-the-top (OTT) streaming from a unique service provider. For the sake of simplicity, we chose a narrow boundary. Parts in dark gray in Figure 3 are those that we include, while we ignore steps in light gray ones. The main variables are the video quality R_v and the percentage of inhabitants accessing the service simultaneously during the peak period.

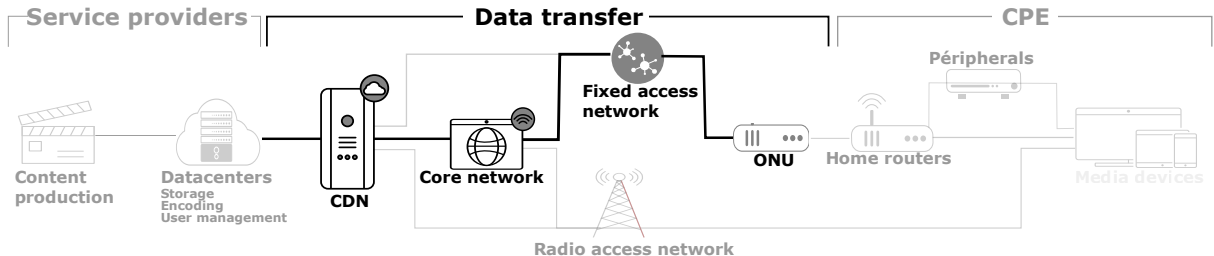


Figure 3: General OTT VoD infrastructure. Steps in black are the ones included in our boundary. ONU stands for Optical Network Unit.

For this use-case, we compared the influence of several video qualities in four scenarios, namely HD (1280×720 at 3 Mbps), FHD (1920×1080 at 5 Mbps), UHD (3840×2160 at 16 Mbps), and a fourth UHD++ scenario with 4K resolution, high-dynamic-range (HDR) color depth, and 60 frame-per-second (at 27 Mbps). The first three bitrates come from previous studies (e.g., [4]), whereas the last one has been measured on Youtube videos. According to audience measuring data for France in 2021 [5], we observe an average of 3h per day with up to 11 million viewers a day for VoD subscribers, and about 4.5 hours of TV content. To account for other OTT platforms and anticipate the ongoing increase of OTT contents, we chose an average of 3.2h of OTT video per day per subscriber with 20 % of watchers at peak for all OTT scenarios.

Video-calls (VC) This use-case includes video-calls between two persons sharing their videos. The main variables are the upload and download peak rates, which we assume both equal to 1 Mbps here², and the percentage

¹web-application

²Typical bandwidth requirements range from 0.6 to 3Mbps for quality ranging from 480p to 1080p (Zoom system requirements).

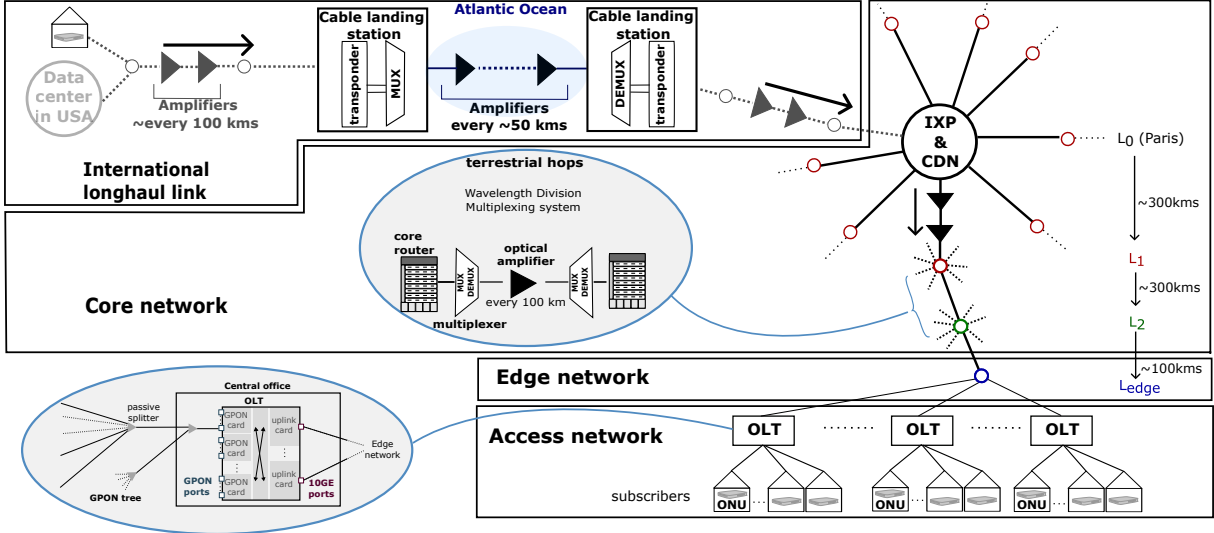


Figure 4: Network modeling. We restrict our boundary to fiber to the home network, a unique datacenter and one CDN used for OTT and DL use-cases.

of inhabitants. We analyzed three scenarios by varying the peak number of persons calling simultaneously, but a common average use of 2 hours of video calls per week per subscriber.

Large downloads (DL) Finally, we considered a download scenario of large files such as OS updates and AAA video games. The latter become larger and larger with the top 26 ranging from 64GB to 200GB at the end of 2021 [16], while generating heavy loads on the network at every release or patch update of the most famous titles. Hence, we used: $R_o = 200$ Mbps and 3% houses, at peak. For this use-case, we relaxed the confidence parameter to $\epsilon = 10^{-7}$ (Section 4-step-4), hence accepting that a few users will likely experience slightly degraded download rates. We also tested a more degraded user-experience using $\epsilon = 10^{-3}$. This use-case assumes an average yearly volume of 25GB per subscriber.

Boundary Figure 4 gives an overview of the infrastructure. We use it for our use-cases. It is the same as the one presented in ICT4S and we refer the reader to this paper for details on equipment specifications. We chose Metropolitan France as a territory, with 65 million inhabitants for about 30 million households, each having an internet connection and being a customer of the OTT, download and video-call service.

For all our use-cases, we ignored content creation and storage, end-user devices, etc., restricting ourselves to data transfer via a fixed network only. We included the core and edge network active equipment, but ignored passive ones. For the access network, we assumed that services are used at home, and considered only a fixed-access network that we assumed to be fully implemented through the GPON (Gigabit Passive Optical Network) FTTH (fiber-to-the-home) architecture.

We assumed that part of the traffic comes from North America at about 900km from the Atlantic submarine cable landing point through an undersea connection. Additionally, we considered that 80% of OTT videos and 95% of DL content were delivered by servers belonging to content delivery networks (CDNs). CDNs reduce data traffic in the core network (and in particular in submarine and longhaul optical fibers) and improve user experience with faster loads. The territory size is adapted to consider a unique CDN, that we naturally located in an Internet exchange points (IXP) in Paris. The CDN servers are partly updated every day during low-traffic periods.

5.2 Scaling and evaluation

Our infrastructure and the number of equipment are scaled according to peak usage, which depends on scenario-specific parameters and other general parameters. In our experiments, the PUE indicator for the network is set to 1.8 [3] for all experiments, while the redundancy factor in case of failure is $\eta = 2$. Our design has been largely inspired by the model of baliga2009, *somepowerconsumptioncomingfromthispaperwithanupdateofenergyintensity(W/Gbps)ofequi*

Owing to the lack of precise power consumption profiles for each of the considered equipment, the previous computations cover static power consumption only, ignoring the dynamic part. To get a rough estimate of what could be the effect of accounting for the dynamic power consumption, we use the average dynamic intensity factor estimated by [14] for a fixed line (i.e., ~ 0.1 Wh/B).

This previous model [10] considered downstream transfers only. However, upstream transfers cannot be ignored in the video call use-case. We make the following three changes, referencing the related equations of the former model [10]. i) We explicitly consider the limited capacity of GPON ports in the upstream direction (1.25 Gbps)

when estimating the maximal number of subscribers per GPON ports (Eq. 6). ii) We assume that the capacity of OLT’s uplink 10 GE ports is limited by the max of the downstream and upstream peak-rate demands (Eq. 8). iii) At all other node levels, we simply assume that both downstream and upstream peak-rates can be summed up to a unique peak-rate demand from which node’s equipment are scaled up.

6 Results on different scenarios

In this section, we compare the reference with other use-cases. We implemented our model as a static web-application³ allowing any user to modify all the parameters and hypotheses of our model, including the capacity and static power of each network element.

Table 2 presents yearly power consumption integrated over one year for each scenario and each network part. To understand the pressure of each usage along the network, this table also reports the peak traffic rate estimated at some key nodes of our tree (omitting growing margins). Finally, to better visualize the dimensioning of the infrastructure, this table also reports the number of GPON and 10GE ports within the OLTs of the access network.

Table 3 presents the global yearly energy consumption as well as global complementary information such as the total yearly traffic volumes together with per subscriber average peak-rate and average rate. This table also reports intensity factors in Wh/GB that will be discussed later.

Table 2: Breakdown statistics. The first part reports annual power consumption in GWh for each network part. The second part reports global peak-rate demands along undersea links (R_u^*), at the national level (R_0^*), and at each central office (R_{hub}^*). The third part reports the number of thousands of instantiated GPON and 10GE ports in OLTs of access network, and in parentheses the increase relative to the reference scenario.

Scenario	ONU Access		National Core+Edge		Int. longhaul	CDN Dynamic Cons.		R_u^* Tb/s	R_0^* Tb/s	R_{hub}^* Gb/s	# GPON ports 10^3	# 10GE ports 10^3
Baseline	667	71	8.8	2.8	0	0.1	2	6.1	2.8	287	6.1	
HD	667	74	24	10	4.1	4.9	7.4	33	13	287 (×1)	12 (×2)	
FHD	667	76	37	15	6.7	8.1	11	51	19	287 (×1)	18 (×3)	
UHD	667	94	107	43	21	26	31	148	55	287 (×1)	55 (×9)	
UHD++	667	149	178	70	35	43	50	246	90	460 (×1.6)	85 (×14)	
undefined	667	246	16	7.2	2.5	5.2	5.1	22	8.7	1011 (×3.5)	15 (×2.4)	
undefined	667	257	69	28	13	17	20	95	35	1011 (×3.5)	37 (×6.1)	
VC 5%	667	71	9.7	3	0	0.4	2.2	9.3	4.9	287 (×1)	6.1 (×1)	
VC 10%	667	71	12	3.2	0	0.4	2.3	13	7	287 (×1)	6.1 (×1)	
VC 20%	667	74	17	3.6	0	0.4	2.6	20	11	287 (×1)	12 (×2)	
DL strict	667	262	140	2.8	32	1	2	189	79	952 (×3.3)	77 (×13)	
DL relaxed	667	137	138	2.8	32	1	2	189	73	444 (×1.5)	67 (×11)	

*1
*3black!10

General observations A first observation from table 2 is the important power consumption from ONUs that are plugged 24/7 in every home. We remark, from this table and total yearly consumption in table 3, that video-calls, even with intensive use, do not add strong pressure over the network. This is not surprising as we consider rather small bitrates, enabling low-resolution videos. At the opposite, large, and fast synchronized downloads have important impacts on the network, with huge pressure on access consumption. In 2022, interconnection traffic in France was 43 Tb/s [2] at the 95th percentile, with a volume of 68 EB [1]. We observe those numbers are close to the FHD scenario, both in terms of volume than peak-rate at the national level (R_0^*).

Materiality From table 2, we see that the increase of the access network consumption is negligible for HD/FHD bitrates, and it remains limited even for the FHD scenario. This is because the configuration of the GPON trees of the basic use-case is enough to handle such bitrates. Only the 10GE uplinks have to be upscaled. The UHD++ scenario, however, yields a higher pressure on the GPON trees, with ×1.6 more GPON ports and ×6 more 10GE ports with respect to the HD scenario. This result illustrates threshold effects.

Proportionality From Table 3, we observe that power consumption is not strictly proportional to bitrates. This is particularly true for part of the networks closer to customers. For instance, whereas the UHD scenario implies a ×24 global peak-rate increase, the energy consumption increases by a factor 3 even without including the consumption of ONUs. Including ONUs, this factor drops down to ×1.2. This is explained by the fact that, whatever the bitrate, a minimum number of equipment is plugged to enable the basic use-case and that the closer

³web-application

Table 3: Global statistics. From left to right: total volume-based traffic, average peak-rate and average rate per subscriber, total yearly power consumption excluding ONUs (with the percentage of increase relative to the reference scenario), and energy intensity factors estimated including or excluding the power consumption of the ONUs.

Scenario	Yearly volume (EB)	Average peak rate (kbps/sub)	Average rate (kbps/sub)	Energy w/o ONUs (GWh)	Efficiency indicators	
					w/ ONUs (Wh/GB)	w/o ONUs (Wh/GB)
Baseline	0.7	200	6.1	82	1026	113
HD	49	1080 (+440%)	406	117 (+42%)	16	2.4
FHD	81	1667 (+733%)	673	143 (+74%)	10	1.8
UHD	257	4893 (+2347%)	2139	290 (+253%)	3.7	1.1
*1 UHD++	433	8120 (+3960%)	3606	475 (+477%)	2.6	1.1
undefined	52	713 (+257%)	432	277 (+237%)	18	5.3
undefined	168	3133 (+1467%)	1395	383 (+365%)	6.3	2.3
VC 5%	3.6	250 (+25%)	30	84 (+2%)	209	23
VC 10%	3.6	300 (+50%)	30	86 (+4%)	210	24
VC 20%	3.6	400 (+100%)	30	95 (+15%)	212	26
DL strict	9.9	6200 (+3000%)	82	437 (+431%)	112	44
DL relaxed	9.9	6200 (+3000%)	82	310 (+277%)	99	31

to customers, the higher the number of equipment and the lesser the traffic. This table also shows that the power consumption is even less proportional to the global volume traffic in GB. For instance, comparing the UHD++ and DL scenarios, we see that both the absolute (389 vs 437 GWh) and incremental (307 vs 355 GWh) energy consumption is less correlated to volume (433 vs 10 EB) than average peak-rates (8.1 vs 6.2 Mbps/sub).

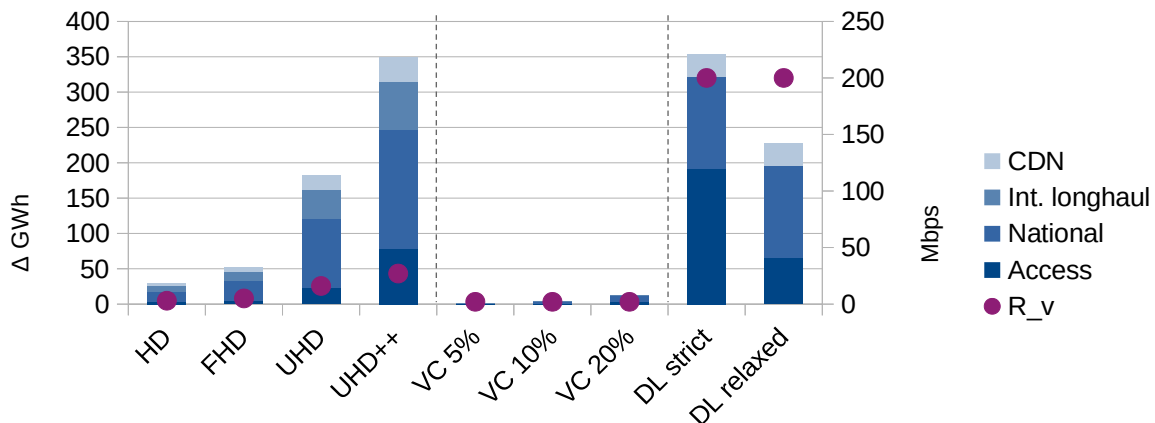


Figure 5: Energy consumption difference between our VoD/DL scenarios and the reference. The respective video bitrates R_v are shown as dots. Numerical values are available in Table “Figure 5” of Supporting Information S1.

Figure 5 shows the energy consumption difference between our OTT/VC/DL scenarios and the reference. One can observe that for the OTT scenarios, for which only the streaming bitrate changes, this difference is almost proportional to the OTT bitrate while absolute consumption is not. However, when including a different use-case as our large-file download scenario, this apparent correlation breaks. We emphasize that those energy consumption differences cannot be naively summed up to account for multiple usages (e.g., UHD + VC 10% + DL relaxed), unless the peak-demands are expected to exactly overlap, which is very unlikely. Instead, one would have to design a scenario combining the different use-cases and estimate the combined peak-demand to feed our model.

Flattening the peak Our two DL scenarios enable studying the influence of peak flattening. The scenario with a lower value of ϵ is expected to yield to a reduced downloading speed at a few moments for some customers, thus limiting the peak by limiting the capacity of the infrastructure. This change has a significant impact on the access network consumption, where the number of GPON tree is considerably reduced.

Illustration of limits of energy intensity indicators Table 3 reports two different energy intensity indicators obtained as the ratio of the total energy consumption (w/ and w/o including the ONUs) and total traffic volume in Wh/GB. In both cases, this clearly shows that increasing the video bitrate enables decreasing the relative data transmission consumption when expressed in Wh/GB, confirming that such an efficiency indicator does not reflect the power consumption increase of more intensive Internet usage. This observation confirms the aforementioned paradoxical situation where increasing the traffic does increase the absolute impacts, but decreases the intensity indicators. Assuming an increase of the traffic over time, this decrease of the intensity factor could be misused to claim energy efficiency gains from the actors whereas this effect is the sole consequence of an increase of the traffic. For the same reason, this effect could easily be confused with a *rebound effect* because an improvement of efficiency indicators is correlated with an increase of the absolute impact. However, here the causal link is reversed, and there is no economical gain that could explain such a direct rebound from the increase of traffic (in contrast to other sectors such as air travel for which increasing the traffic implies reducing the ticket price per passenger). Of course, efficiency gains can also come from technological improvements, and such improvements might reduce the absolute impacts, but they are also subject to rebound effects. In practice, these two different mechanisms coexist. Including or excluding the ONUs yields drastically different intensity values, meaning that conclusions or information that could be extracted from such intensity indicators are expected to drastically depend on the choice of the perimeter.

As an insightful exercise, let us try to anticipate what would be the increase in energy consumption of moving from the FHD scenario to the UHD one using such indicators. This change implies an additional traffic volume of 176 EB. Assuming that the FHD scenario corresponds to the present time, in this exercise, one would observe intensity factors of 9.8Wh/GB or 1.6Wh/GB depending on the perimeter. This leads to an anticipated power consumption increase of 1724 GWh or 282 GWh respectively, whereas, according to our model, the real increase would be about 112 GWh only (with and w/o ONU). Both intensity-based results are thus largely overestimated, meaning that using such intensity indicators might be misleading and could even lead to counterproductive recommendations or actions to limit or reduce the power consumption of networks.

7 Discussions

In this section, we discuss some choices made, limits, and future work for our model.

Conservative hypotheses We emphasize that our observations on the non-proportionality (even for the relative plot of Figure 5) and the sobriety vs efficiency conflict would be exacerbated through less conservative choices. We could for instance include home-routers energy consumption in addition to ONUs one, decorrelate the peak-rates of the basic and other use-cases to better absorb part of use-cases traffic. We could also consider a more realistic non-homogeneous distribution of the population in rural and urban areas, and non-homogenous technologies with newer ones, including renewal rates.

We also assumed that the ONUs are always on. Enabling ONU/home-routers to be switched off when unneeded would result in a significant reduction of the absolute energy consumption. It is thus interesting to question how much a usage is preventing such equipment from being switched off. For the basic use-case, a reasonable assumption would be to assume that they are switched off a few hours over-nights and when the households are empty. An obvious worst-case usage is, however, smart-home equipment that require a permanent connection.

Complexity of the real world Like previous bottom-up models, ours only offer a simplified vision of reality which is more complex. For instance, it does not explicitly account for legacy equipment that yield a huge variability both in terms of capacity and efficiency. A more realistic model should also consider more local cache servers handled by Internet service providers. With many actors deploying network equipment, peak bitrate demand is not the only driver for the increase of the infrastructure, but economic competition and geopolitical strategies also play an important role and lead to overdimensioning. As future work, it would thus be interesting to integrate all those aspects, and in particular mutualization and duplication between operators, in such a model.

Mobile network This paper models FTTH network only, but some use-cases are likely to also occur on mobile network. Modeling the mobile traffic would lead to very different energy consumption.

Restricted perimeters In this work, we have not included user devices, datacenters, nor content creation and encoding. On the datacenter side, maximal video resolution and quality are expected to have a significant effect on the computing (encoding) and storage resources. Those parameters are also expected to play an indirect but important role in accelerating the renewal of end-user equipments, for instance for larger 4K, HDR-enabled TVs, hence increasing the overall electricity consumption, but also manufacturing impacts. Within a consequential approach, multi-purpose equipment (e.g., end-user devices) are classically treated through the *system expansion* mechanism. In our case, it consists of extending the boundary to include most of its usages, hence bypassing tricky allocation issues at the expense of a higher complexity.

Carbon footprint Converting energy consumption to carbon footprint requires knowing emission factors. These factors are country-dependent but also time-dependent as the energy mix depends on the time or season. Moreover, averaged emission factors, even if made temporally varying, are not necessarily correlated to the *consequential* effect of adding or removing a large body of electricity demand. Therefore, in this study, we have omitted this step on purpose.

Life cycle Most studies are limited to the use phase of equipment, omitting other phases of their life cycle (material production, manufacturing, transport, installation, maintenance, end-of-life). The main reason is the high level of uncertainty in estimating the emissions of these other phases. Some works do attempt to include embodied impacts [15, 20] from an average ratio method that computes the scale of embodied emissions compared to the use phase. In this paper, we have focused on the use phase only. Still, we acknowledge the strong importance of including the other phases, as well as accounting for other environmental impacts (e.g., water footprint, human toxicity, abiotic resource depletion, ...). Properly allocating embodied emissions of shared equipment is as tricky as allocating static power consumption. In this regard, since our global approach bypasses the need for arbitrary allocation, we argue that the methodology proposed in this paper is well-suited to be extended to estimate embodied impacts. Our model shall therefore be extended to enumerate all passive equipment that we have neglected so far (cables, shelters, buildings, racks, etc.), as well as installation and maintenance operations.

Recommendations and insights Owing to some of the aforementioned limitations, our implementation cannot readily be used by network planners nor to carry out precise numbers, but we believe it could be used to enable more informed debates between the stakeholders (regulators, content providers, operators, and users) to identify the most impactful recommendations both in terms of actions or bitrate thresholds. For instance, it seems that even a massive use of HD VoD through CDNs puts restrained pressure on a fiber network, in contrast to higher-quality VoD. At such a bitrate, the question of the transmission mean (e.g., OTT versus digital terrestrial television (DTT) broadcast) becomes secondary. A similar reasoning could be used for other raising large-scale and data-intensive use-cases such as cloud gaming or metaverse. Our model also calls for peak-flattening actions (reduced quality, bitrate throttling, or time shifting for large downloads, etc.) that could be endorsed by either content providers and/or end-user with possibly the help of current load indicators from operators. Our results also reveal some use-cases for which simple intensity indicators should be avoided. Those include reporting from service providers or operators to track or claim energy efficiency efforts, and any extrapolation exercise. The last use-case is rather frequent in the literature, such as the LoCaT report [15], where a naive extrapolation from intensity factors would encourage shifting a proportion of VoD traffic to DTT networks through a home caching approach, whereas our model rather reveals a clear increase of the absolute energy consumption of such an approach [10]. Our results also show that intensity indicators should be avoided to report or claim avoided emissions resulting from an action that decreased the amount of transferred data. Such overestimated avoided emissions are often claimed by companies offering “green” services. Targeting end-users, awareness tools (e.g, *impactco2*⁴ or reports (e.g., [4]) based on intensity indicators gives the false impression that carbon emissions are completely proportional to video viewing hours or quality, hence overestimating either additional and avoided emission potentials. Instead, our model and tool could be used to qualitatively assess the increase or decrease of the network pressure by exploring “what if” scenarios where all citizens would be users of a given service or concerned by a given action.

8 Conclusion

In this paper we have extended and further analyzed the methodology from [10] to assess the relationship between a given usage and fixed network power consumption. By looking at global energy consumption rather than attempting to arbitrarily allocate power between the different usages, we avoid the classical pitfalls. In addition, this methodology enables to observe the consequences over the network of a given use-case, and to illustrate threshold effects. Our results confirmed that classical efficiency indicators do not reflect the power consumption increase of more intensive Internet usage, and might even lead to misleading conclusions. The bottom-up parametric network model we presented has the notable property of translating global average statistics to local smaller pools of inhabitants. This theoretical network model is, however, necessarily imperfect and this paper discussed many future work opportunities such as variation of the density of population, variability of equipment, broader boundary, adding a broadband radio access network, and modeling other use-cases. To enlarge the application range of our model, it would be useful to investigate how to extend our methodology to properly account for multiple use-cases whose peak demands are expected not to overlap. At last, we believe our model could also be extended to simulate some higher-order effects [12] such as obsolescence or direct rebound-effects. Many of the suggested additions would require to extend it to account for the temporal dimension beforehand.

⁴<https://impactco2.fr/usagenumerique/streamingvideo>

References

- [1] ADEME. *Evaluation de l’empreinte environnementale de la fourniture d’accès à internet en France*. 2024. URL: <https://bibliothèque.ademe.fr/produire-autrement/6789-evaluation-de-l-empreinte-environnementale-de-la-fourniture-d-acces-a-internet-en-france.html>.
- [2] ARCEP. *Baromètre de l’interconnexion de données en France*. 2023. URL: <https://www.arcep.fr/cartes-et-donnees/nos-publications-chiffrees/linterconnexion-de-donnees/barometre-de-linterconnexion-de-donnees-en-france.html#c10040>.
- [3] Joshua Aslan et al. “Electricity Intensity of Internet Data Transmission: Untangling the Estimates”. In: *Journal of Industrial Ecology* 22.4 (2018), pp. 785–798.
- [4] Carbon Trust. *Carbon impact of video streaming*. 2021. URL: <https://www.carbontrust.com/resources/carbon-impact-of-video-streaming>.
- [5] CNC. *Audiometry data from CNC*. 2021. URL: <https://www.editionmultimedia.fr/wp-content/uploads/2022/03/Observatoire-CNC-de-la-VOD-17-12-21.pdf>.
- [6] Vlad Coroamă. *Investigating the Inconsistencies among Energy and Energy Intensity Estimates of the Internet, Metrics and Harmonising Values*. Bern, Switzerland: Swiss Federal Office of Energy SFOE, June 2021.
- [7] Vlad C. Coroamă et al. “The Direct Energy Demand of Internet Data Flows”. In: *Journal of Industrial Ecology* 17.5 (2013), pp. 680–688.
- [8] Vlad C. Coroamă et al. “The Energy Intensity of the Internet: Home and Access Networks”. In: *ICT Innovations for Sustainability*. Springer International Publishing, 2015, pp. 137–155.
- [9] Charlotte Freitag et al. “The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations”. In: *Patterns* 9 (2021). ISSN: 2666-3899. DOI: <https://doi.org/10.1016/j.patter.2021.100340>. URL: <https://www.sciencedirect.com/science/article/pii/S2666389921001884>.
- [10] Gaël Guennebaud, Aurélie Bugeau, and Antoine Dudouit. “Assessing VoD pressure on network power consumption”. In: *ICT4S - International Conference on Information and Communications Technology for Sustainability*. Rennes, France, June 2023. URL: <https://hal.science/hal-04059523>.
- [11] R. Guerin, H. Ahmadi, and M. Naghshineh. “Equivalent capacity and its application to bandwidth allocation in high-speed networks”. In: *IEEE Journal on Selected Areas in Communications* 9.7 (1991), pp. 968–981.
- [12] Lorenz M Hilty and Magda David Hercheui. “ICT and sustainable development”. In: *What kind of information society? Governance, virtuality, surveillance, sustainability, resilience*. Springer, 2010, pp. 227–235.
- [13] Kerry Hinton, Fatemeh Jalali, and Ashrar Matin. “Energy Consumption Modelling of Optical Networks”. In: *Photonic Netw. Commun.* 30.1 (Aug. 2015), pp. 4–16. ISSN: 1387-974X.
- [14] Jens Malmodin. “The power consumption of mobile and fixed network data services - The case of streaming video and downloading large files”. In: *Electronics Goes Green 2020+* (2020).
- [15] William Pickett et al. *Quantitative study of the GHG emissions of delivering TV content*. The Locat Project, 2021. URL: <https://thelocatproject.org/>.
- [16] Jacob Ridley. *The biggest games by install size, real mighty storage hogs*. PC-Gamer, <https://www.pcgamer.com/biggest-game-install-sizes/>. 2021.
- [17] Daniel Schien, Paul JS Shabajee, and Chris W Preist. “Rethinking Allocation in High-Baseload Systems: A Demand-Proportional Network Electricity Intensity Metric”. In: *IETF Internet Architecture Board workshop on Environmental Impact of Internet Applications and Systems*. 2022.
- [18] Daniel Schien et al. “Help, I shrunk my savings! Assessing the Carbon Reduction Potential for Video Streaming from Short-Term Coding Changes”. In: *2023 IEEE International Conference on Image Processing*. 2023.
- [19] Daniel Schien et al. “The Energy Intensity of the Internet: Edge and Core Networks”. In: *ICT Innovations for Sustainability*. Ed. by Lorenz M. Hilty and Bernard Aebischer. 2015, pp. 157–170.

- [20] Noel Ullrich et al. “Estimating the resource intensity of the Internet: A meta-model to account for cloud-based services in LCA”. In: *Procedia CIRP* 105 (2022). The 29th CIRP Conference on Life Cycle Engineering, April 4 – 6, 2022, Leuven, Belgium., pp. 80–85. ISSN: 2212-8271.
- [21] Anson Wu and Paul Ryan. *Intelligent Efficiency for Data Centres and Wide Area Networks*. Tech. rep. International Energy Agency (IEA), 2019.