Standalone Data-Center Sizing Combating the Over-Provisioning of the IT and Electrical Parts

Manal Benaissa, Georges da Costa, Jean-Marc Nicod

Standalone Data-Center Sizing Combating the Over-Provisioning of the IT and Electrical Parts

Manal Benaissa[1,2], Georges Da Costa[1], and Jean-Marc Nicod[2]

[1] *IRIT Institute*, Université de Toulouse, CNRS, Toulouse, France,
georges.da-costa@irit.fr
[2] *FEMTO-ST Institute*, Univ. Bourgogne Franche-Comté, CNRS, ENSMM, Besançon, France,
manal.benaissa,jean-marc.nicod@femto-st.fr

November 2, 2022

Abstract

During the two last decades, sustainability of IT infrastructures like datacenters became a major concern for computer giants like Google or Amazon. Datacenters powered with renewable energy have been proposed. But because of the intermittency of these power alternatives, these platforms remains connected to the classical power grid. IT structure and electrical constraints were often questioned separately, leading to a non-efficient global system. In this paper, an energy self-sufficient green datacenter is modeled and designed, by proposing an electrically autonomous infrastructure including wind turbines, solar panels and its own short and long-term storage devices mainly based on batteries and hydrogen system. Existing sizing approaches limit themselves to a perfect QoS leading to an overestimation of the needed equipment. In this paper we show how to reduce and combat this overprovioning by questioning its impact on the QoS and needed equipment: decreasing the number of computing or storage elements (servers and batteries). As an example, decreasing the targeted QoS from 100 to 95% more than halves the needed number of servers, while a decrease of 30% of the battery capacity has a negligible impact on the electrical infrastructure.

Green Datacenter, Sizing reduction, Scheduling, Renewable energy, Quality of Service.

1 Introduction

Approximately two-thirds of the world's population will have access to the internet by 2023 [1] [2]. This represents almost 5.3 billion people using all services available on the internet. Datacenters form a large group of computing resources linked by a fast network designed to run the most diverse types of applications that demand a huge amount of resources [3] [4]. One of the main concerns of datacenters is their major impact on global electricity consumption. The entire IT sector wasted 4.7 % of the worldwide electricity consumption in 2012, while datacenters alone consumed 1.4 % [5] [4]. Renewable energy utilizes self-renewing resources such as wind and sunlight, which provide clean energy. These two renewable sources are suitable options due to their competitive installation cost and their abundance [6]. Using only renewable energy like in the DataZero Project [7] imply the use of storage systems to manage the intermittent nature of renewable energy production.

Renewable energy seems the energy future, mainly to large datacenters, but it also introduces some uncertainty that demands new strategies. Datacenters are modeled to deal with workload peaks leading to over-provisioning, generating under utilization most of the time [8]. Hence, there are two main problems in the datacenter size definition. The first problem is to define the appropriate sizing to a datacenter that minimizes under/over utilization. This definition needs to meet the workload demand by providing accurate resources. On the other hand, the electrical side needs to determine the components to meet the datacenter power requirement. Our approach is in two steps. First, we propose an IT model to define the right amount of servers according to workload demand and a given utilization rate. This model outputs the number of servers applying a scheduling algorithm which provides a more accurate way to predict the datacenter size according to certain metrics. Then, we define the electrical components needed to supply the power envelope required by the datacenter using only renewable energy. The novelty of our ap-

proach is the link between the electrical side using only green energy with the computing side sizing.

Therefore, our contribution to mitigate overprovisioning during sizing is twofold:

- We explore the balance between QoS and number of servers (and their utilization);
- We explore the number of days the storage elements are used at full capacity, hinting on the impact of their possible reduction.

This paper is organized as follows: Section 2 starts with an overview of the state of the art. Then Section 3 defines the problem we address, including the models and the metrics. Section 3.1 describes the algorithms to size both the IT and electrical components for a QoS of 100 % while Section 4 proposes strategies to mitigate oversizing of the datacenter. Section 5 describes and discusses the experimental results. Section 6 concludes the paper.

2 Related work

Until now, the holistic sizing problem in the context of renewable-powered datacenter was not really addressed before. Most research on such sizing actually focus on a part of the whole infrastructure. In [9], authors focus only on the sizing of the electrical system. Even more precisely, in [10], authors focus on the storage sizing, and propose to use the storage not only for the datacenter itself, but as a tool for the overall grid stability also. The choice of renewable energy was mainly based on the availability of wind and solar energy across the world, and in [11], the combination of both sources seems to lead to better results. But solar and wind energy are known to be intermittent, and compensating this problem with adequate storage system is a prerequisite.

Only few works address the case of full datacenter sizing, from computing elements to electrical one. In [12], authors choose to provide a perfect datacenter, i.e., able to answer to all job requests. It leads to an oversizing of both computing and electrical infrastructure due to high variability for both renewable energy production and job workload.

3 Problem statement

The sizing problem consists in designing both the IT and electrical parts of the datacenter by choosing which component and how many one need to give the most appropriate infrastructure to address metrics that have to be optimized. Multiple configurations exist and the sizing process has to give these solutions which associated metrics to help the decision maker to make his choice depending on the future datacenter usage. The IT components are servers, while the electrical ones are photovoltaic panels (PV), wind turbines (WT), batteries (Bat), electrolyzers (EZ) and fuel cells (FC). The last two respectively consume or produce H_2 in exchange for electricity. We consider an extreme case where the datacenter does not have access to the classical power grid.

We assume that we are able to known what kind of workload $\mathcal{W}$ has to be executed within the datacenter with a given quality of service (QoS) and a given scheduling algorithm. We also assume that weather data is available for the planning period for the location where the datacenter is to be built. The problem is to find the appropriate number of servers to run tasks of $\mathcal{W}$ and what electrical platform is needed to supply the IT part of the datacenter.

We propose here to question this sizing process that has been described with a perfectly guaranteed Quality of Service in [12]. The proposed approach consists in measuring the influence of reducing some component size such as the number of servers or the capacity of batteries on the datacenter performance on specific metrics. In the following, we show the impact of reducing these elements on the overall QoS provided by the datacenter. This aims at combating datacenter over provisioning and at providing to datacenter operators' tools to select the most relevant sizing for their own use case.

3.1 Models

An usual sizing process is divided into two steps: first sizing the IT part and then the electrical part of the datacenter. The first step consists in computing the number of adapted servers given a workload $\mathcal{W}$ to be executed on a time horizon $\mathcal{H}$. Then, executing $\mathcal{W}$ on these servers implies a power demand that has to be supplied on site by the electrical part of the datacenter. The number of wind turbines, the area of photo-voltaic panels (producing renewable energy from wind and sunlight) and the capacity of the storage devices (for mitigating the intermittency of these renewable energies) are linked to the IT-provided power demand. Several configurations are possibles considering the balance between the number of wind turbines and the surface of solar panels.

The decision horizon $\mathcal{H}$ is defined as a set of decision steps. Indeed, $\mathcal{H}$ is discretized into K indivisible time slots Δt such that $\mathcal{H} = K\Delta t$ and where it is possible to take decisions. For simplicity reasons, we assume that Δt is one unit of time ($\Delta t = 1\, u.t.$).

In the rest of the paper, one unit of time will be one hour.

Notations and models are now given to describe the addressed problem and solutions provided. These models are based on a previous study initially described by the authors in 2021 in [12]. However this study does not evaluate the over sizing of the datacenter infrastructure as we propose in what follows.

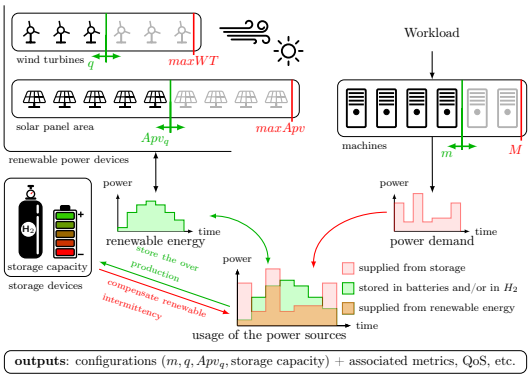
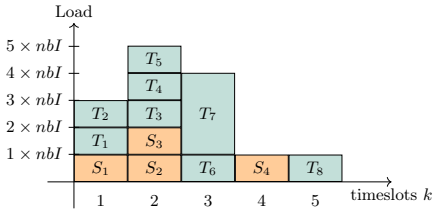


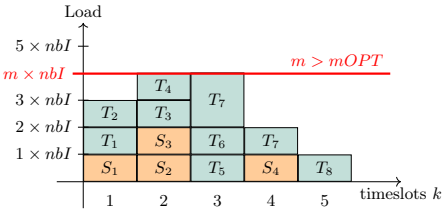
Figure 1: Electrical sizing: the configuration of the renewable devices (q, Apv_q) is the number of WT and the corresponding surface of PV required to meet the datacenter demand with its m computing servers. Storage devices are designed to store or deliver energy to mitigate the renewable energy intermittency. Note that $(0, maxApv)$ and $(maxWT, 0)$ are valid configurations.

3.1.1 IT infrastructure

The amount of work that has to be processed by the datacenter is a workload. This workload $\mathcal{W}$ consists of a set of jobs that are either services $S_i \in \mathcal{S} = \{S_1, \ldots, S_r\}$ or tasks $T_i \in \mathcal{T} = \{T_1, \ldots, T_n\}$. Each job has respectively a release date rs_i or rt_i and an number of instructions $ws_{i,k}$ and wt_i that has to be proceed, depending whether it is a service or a task. We assume that unlike tasks, services are unable to be delayed (interactive applications or video streaming for instance). Moreover, each service has to complete its workload in due time (i.e., $ws_{i,k}$ must be processed during the k^{th} time-window). The task execution model is quite different. Each task T_i is submitted at time rt_i – released date at the beginning of the time slot k $(k = 1, \ldots, K)$, $rt_i = (k-1)$– and its earlier completion time is $rt_i + 1$ whatever the requested amount of work $w_{i,k}^{req}$ to process. When scheduled, T_i is assigned onto processors in different time slots such that its actual completion time is less than $rt_i + \delta$ where δ is the flexibility of $\mathcal{T}$. As a consequence, the amount of work $wt_{i,k}^{sch}$ of T_i at time slot k is not the same after the scheduling process of the workload. The amount of work to process within a time slot is expressed in million of instructions [MI].



(a) Gantt chart of $\mathcal{W}$ $(\mathcal{S} \cup \mathcal{T})$ at their submitted release date. Time slot $k = 2$ (the busiest one) defines $maxM = 5$. Time slots $k = 3$ and 5 define $minM = 0$.



(b) Gantt chart of $\mathcal{W}$ being scheduled on $m = 4 > mOPT$ servers.

Figure 2: Motivated example of a workload $\mathcal{W}$ to be scheduled thanks to EDF algorithm considering the same flexibility $\delta = 2$ time slot for each task (each task T_i can be completed at most 2 time-slots latter than c_i^{req} $(c_i^{req} = rt_i + 1 \leq c_i^{sch} \leq c_i^{req} + \delta)$. The red horizontal line represents the overall load in [MI] the datacenter is able to run.

The jobs are scheduled with the EDF algorithm (Earliest Deadline First). When the available resources are not available a task might not be executed before its deadline. In this case, we reject this task. The QoS is given by the ratio between the number of executed tasks and the total number of tasks.

On the IT infrastructure side, the operating room of the datacenter consists of a set of m homogeneous servers $M_j \in \mathcal{M} = \{M_1, \ldots, M_m\}$. Each server M_j consumes at most $p_j = p$ Watts for the computation of at most $nbI_j = nbI$ MI (Millions of instructions).

Considering the number m of servers of the IT hardware architecture of the datacenter, it is possible to evaluate the power demand $\mathcal{D}^m = \{D_1, \ldots, D_K\}$ of the datacenter time slot k by time slot k as explained in [12]. PUE (Power Usage Effectiveness) is a coefficient greater than one that allow to integrate the datacenter facilities, as the cooling system, into the power demand.

The power demand is needed to provide the electrical sizing of the datacenter. But to obtain the power envelope requested by the IT part, the work-

load scheduling has to be processed first and the value for m has to be set. The IT sizing process consists in finding the minimal number of servers $m = mOPT$ for a given QoS. The minimum number of servers $mOPT$ is found using a binary search algorithm between $minM$ and $maxM$, where each step corresponds to a schedule attempt.

At the end of the IT sizing process, $mOPT$ and the power profile $\mathcal{D}$ are given.

3.1.2 Electrical infrastructure

The electrical infrastructure consists of primary sources, such as wind turbines and photovoltaic panels, and secondary sources such as batteries and hydrogen system to mitigate the effect of intermittency of the primary sources. This intermittency has several origins as day and night alternation (short-term) and seasonal variations (long-term). Given the storage capacity and the efficiency involved by these storage devices, we assume to dedicate batteries and hydrogen systems for managing respectively short-term and long term-storage.

The electrical sizing process aims to determine the number of each component of the electrical infrastructure, to satisfy the IT power demand $\mathcal{D}$. Basically, considering a workload $\mathcal{W}$, and characteristics of individual elements (one server, one wind turbine, one PV, one battery, one fuel cell), the sizing:

1. Selects a number m of servers and produce the associated power demand $\mathcal{D}^m$ needed to run $\mathcal{W}$;

2. Determines the maximum number of wind turbines $maxWT$ to meet alone the power demand $\mathcal{D}^m$ over $\mathcal{H}$. Then the number of configurations to explore is $maxWT + 1$, each with q WT with $0 \leq q \leq maxWT$;

3. Determines the surface of photovoltaic panels Apv_q to meet $\mathcal{D}^m$ thanks to the aggregated combined renewable power production using also q WT;

4. Simulates over $\mathcal{H}$ the behavior of the whole system (IT and electrical parts) and determines the needed short-term and long-term storage capacities (resp. BC and LOH) and power of each storage device.

Considering any number of servers m chosen to run $\mathcal{W}$, a power demand D^m is requested to the electrical part of the datacenter. Weather data over one year ($\mathcal{H} = 1$ year) allows us to determine which electrical infrastructure should be required in term of WT and surface of PV to meet $\mathcal{D}^m$ and to maintain the same level of H_2 at the end as at the beginning of the time horizon ($LOH_0 = LOH_K$). As mentioned before, $maxWT + 1$ configurations are valid. The principle is to set q ($0 \leq q \leq maxWT$) and to find the corresponding surface Apv_q of PV using a binary search approach to satisfy the constraint $LOH_0 = LOH_K$. This level is determined using the rules of the game of the storage management, time slot by time slot, as initially described in [12].

4 How to combat the oversizing?

The sizing process, as explained above, provides a set of configurations (m, q, Apv, BC, LOH) for both the IT and the electrical parts of the datacenter. In the initial sizing, the configuration obtained fully met the demand, without consideration of possible variations in it. As mentioned before, the electrical sizing is designed to supply the power demand of the IT part of the datacenter to reach a QoS of 100%. Thus the sizing of both IT and electrical parts is designed by a worth case approach. Indeed, the number of servers is defined by the time slot that needs to schedule the maximum amount of work. In the same way, the capacity of the batteries is defined by the day where we need the largest amplitude to allow the batteries to compensate for the alternation of day and night in terms of renewable power production, allowing the datacenter to be used even if the renewable energy is missing. In the both cases, if the sizing is defined by an epiphenomena, the sizing of the datacenter can be arbitrary large. In this section, we propose an approach to combat the oversizing by questioning both the initial number of servers and the battery capacity. Moreover we show how to compensate the battery capacity limitation without changing the datacenter power supply. Section 5 is then dedicated to the analysis of this approach.

4.1 Mitigate the IT oversizing

The number of servers needed to satisfy a given workload with a QoS of 100% is the minimum number of servers to complete each job before its deadline. To question this number of servers, the QoS becomes a parameter. Thus, if the QoS is 100%, the proposed configuration does not tolerate any deadline violation. Conversely, a lower QoS admits a certain number of job cancellations and then allows a lower number of servers for the datacenter.

In this way we adopt the same binary search approach as described before to obtain the optimal number of servers for different QoS targets. Section 5 shows which number of servers is needed for QoS levels of 100%, 99%, 95% and 90% respectively.

4.2 Mitigate the battery capacity

To question the battery capacity, we choose BC into a set of values that are less than the one optimally computed by the sizing process in other to respect the rules of the game and the datacenter QoS. Then we supply the datacenter within the energy demand using a greedy strategy of using the battery as much as possible and the hydrogen system to store renewable energy only when the battery is full or to generate electricity only when the battery is empty. This strategy aims to minimize energy waste since the efficiency of the battery is higher than that of the hydrogen system. Since BC is less than its initial value, this strategy put more pressure on the hydrogen system. After this one year simulation process, it is easy to appreciate the lack of hydrogen at the end of the year because we have to consume more hydrogen as initially expected by the initial sizing. Indeed, we loose energy because the hydrogen system is less efficient than the batteries.

The question then becomes how to compensate for this lack of energy in order to supply the datacenter as required by the IT demand and to maintain a QoS of 100%. A valid option is to produce this missing energy from renewable energies by oversizing one of the electrical components, such as additional solar panels. To validate this process of improving the electrical infrastructure, we have to measure whether limiting the battery capacity and adding solar panels, even if it means slightly breaking the rules of the game as mentioned before, is an advantage for different metrics or not.

The surface of solar panels to be added can be optimally computed using a binary search algorithm. This algorithm consists in increasing the surface of the solar panels if the level of hydrogen LOH_K at the end of the time horizon $\mathcal{H}$ is less than its level LOH_0 at the beginning, or decreasing the surface otherwise. During the period, as BC is fixed for the battery capacity, a greedy algorithm is implemented to know which storage device to use hour by hour. Since the efficiency of the battery is higher than the one of the hydrogen system, we use the battery whenever we can. This simulation step of our approach is performed for different values of the battery capacity BC in order to evaluate the benefit of this under-sizing process.

5 Experiments

5.1 Inputs

For the IT experiments, 50 workloads of 8760 hours were generated, following distributions law of Google trace workloads [13]. The average number of job arrival per hour is 50, when each job size is between 10^{10} and 10^{15} MI.

The considered IT resources are 16 cores servers, where each server can compute 5.29×10^{13} MI and consume 350W (only one medium frequency is considered here). The considered value for PUE is 1.3 (Power Usage Effectiveness).

Electrical sizing requires the characteristics of solar panels, wind turbines and storage infrastructure. These include the charging and discharging efficiencies of batteries and hydrogen system. These characteristics are listed in Table 1.

P_r	$400\,kW$	V_r	$14\,m/s$	V_{ci}	$4\,m/s$
V_{co}	$25\,m/s$	η_{fc}	0.6	η_{ez}	0.6
η_{ch}	0.8	η_{dch}	0.8	η_{pv}	0.15

Table 1: Devices behaviors for the electrical sizing process

5.2 Simulation results when combating the oversizing

5.2.1 IT oversizing

To illustrate the combating IT oversizing process, we consider the previous presented workloads as an input considering a given QoS target. On the particular configuration described as an example, the minimum number of servers required to run the entire workload is $mOPT = 194$ as shown in Table 2.

$minM$	135 servers
$maxM$	342 servers
$mOPT$	194 servers
Number of jobs	254420 jobs
max power demand $\max_{1 \leq k \leq K}(D_k)$	67.9 kW

Table 2: Variables during a IT sizing process when QoS is 100% on one of the tested workload.

Based on this result, a linear search is done between $minM$ and $mOPT$ to identify which tasks are decisive in the computation of W. Starting from $mOPT$ and decreasing the number of servers to $minM$, the problematic tasks are listed. A task becomes problematic when its deadline is violated or when it remains no space in any timeslot to be scheduled. Several strategies are studied when such

a task is found, but only one is selected: The Reject Strategy. This strategy consists in rejecting a task when its schedule is no longer possible without violating its deadline. It is then discarded and the scheduling process restarts without the rejected task.

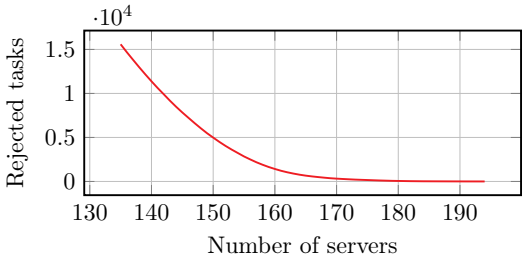


Figure 3: Number of rejected tasks, for a load per server at 1000 MI. The linear search starts with $mOPT = 194$ on the bottom right and finishes with $minM = 135$ on the left.

The Reject Strategy impacts on the number of rejected tasks is shown in Figure 3. The experiment starts on the bottom right with $mOPT$ servers, where no rejection is expected. The number of servers decreases until $minM$, showing in the same way that the number of rejected tasks increases drastically.

This first step allows us to build a first intuition of the behavior of sizing at reduced QoS. In the next step, other sizing are proposed, respectively with a QoS of 99%, 95% and 90%. In order to propose such configurations, the same linear search is performed, but with another lower bound: For a given workload, its rejection limit is calculated from the total number of tasks in the workload, as well as the targeted QoS. Thereafter, the number of servers decreases from the $maxW$ at QoS of 100%. The linear search stops when the limit of rejections is reached.

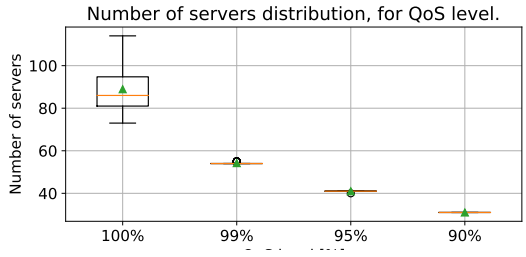


Figure 4: Distribution of the servers over 50 workloads with an average load per server of 5.29×10^{13} MI.

The sizing with reduced QoS is performed with 50 workloads, in order to see the distribution of the number of servers. For these experiments, 50 workloads were generated, following distributions law of Google trace workloads [13]. Figure 4 shows the number of servers distribution needed to satisfy the workloads, with four different QoS level. For a QoS of 100%, we observe a variation that can be up to double for some workloads. This dispersion is drastically reduced as soon as the QoS reaches 99%. It is explained by the fact that the rejections "smooth" the workload, by discarding the epiphenomena. The reduction is also remarkable because the 1% reduction is enough to almost halve the number of servers.

5.2.2 Electrical oversizing

This simulation allows to observe the behavior of the electrical sizing when the battery capacity is limited. This analysis consists in setting BC to a set values, its maximum value being the capacity given by the initial sizing, and then in running a simulation to find a new electrical sizing as describe before and then giving metrics to help the decision maker to choose the appropriate configuration.

To do so, different power profiles required by a initial IT sizing for different studied workloads are considered. An initial electrical sizing configuration that respects the rules of the game is performed of each of them. An alternative battery capacity is then computed in the second step to combat its oversizing if so. Wind speed and irradiation from the city of Lille, France, over the year 2019 are taken from NASA data.

The parameters used for this analysis are given in Table 3, and the profile of the configuration reference is shown in Figure 5.

Location/year	Lille (France) / 2019
Configurations #1	(0, 22921, 6045, 939019509),
Configurations #2	(1, 9687, 3948, 268069986)
Configurations #3	(2, 1596, 2599, 269806939)
Configurations #4	(3, 0, 3230, 826084706)

Table 3: Parameters for electrical sizing analysis. An electrical *Configuration* is defined by (q, Apv, BC, LOH), where q is the number of wind turbine [u], Apv is the solar panel area [m^2], BC the battery capacity [Wh] and LOH the hydrogen capacity [kg].

Table 4 shows the evolution of the difference in solar panel surface and hydrogen storage capacity, for the configuration #3 in Table 3. For a reduction of -16% of the battery capacity, the solar panel surface does not increase, and the needed hydrogen resources stay almost unchanged too, according to the reference configuration shown in Table 3. The

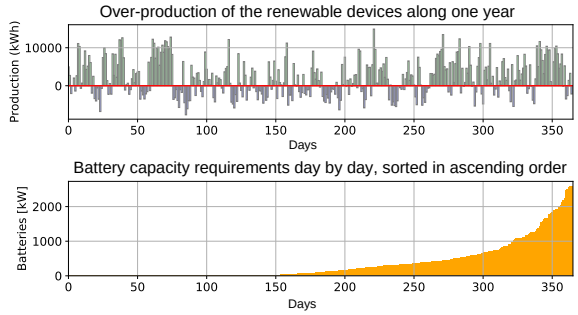


Figure 5: Over-production with a configuration of 2 wind turbines and 1596 m^2 of solar panels in Lille (France) in 2019 and the corresponding battery capacity requirements day by day, sorted in ascending order.

Batteries capacity reduction	-16%	-20%	-30%
PV increasing [%]	0.0%	0.21%	1.11%
LOH increasing [%]	0.13%	0.29%	0.77%

Table 4: PV and H_2 evolution, depending on the battery capacity reduction. The reduction is computed on the basis of the reference configuration, in Table 3

lack of battery is indeed compensated by an addition of photovoltaic panels, however, it is interesting to note that this increase is not significant considering the drastic limitation of the battery. This result is interesting, given the price of solar panels compared to the price of batteries. Indeed, the NREL laboratory had predicted a net increase in the batteries cost, in particular due to its increasing use for electric vehicles or for the manufacture of digital equipment [14]. On the other hand, the price of solar panels seems to be decreasing, thanks to technological advances and the global will to limit carbon emissions [15], [16]. In the case of hydrogen, the production and storage facility remains a significant investment, but one that can be largely amortized if the excess is resold [17], [18].

Overall the reduction of the battery capacity leads to a small increase on the other elements of the electrical infrastructure. In our example, a decrease of 30% of the battery capacity leads to an increase of 1.11% of the PV area, and of .77% of the produced hydrogen.

6 Conclusion

Obtaining the sizing of the infrastructure is divided into two parts: the IT sizing gives the number of servers needed to satisfy a given workload; the electrical sizing deducts from the IT power demand the number of wind turbines, solar panels, short and long term power/capacity of the storage system (resp. batteries and hydrogen system). In a context of reducing the carbon footprint of the total infrastructure, several analysis have been performed to observe the consequences of decreasing the number of servers and battery capacity. An important and interesting reduction has been noted on the IT and electrical sizing. For all workloads studied, a reduction in the number of servers by up to half had little impact on QoS. A resilient reduction would then be possible: Indeed, for a QoS of 95% for example, all the workloads seem to be able to cope with less than half the servers needed for a QoS of 100%. However, it is important to note that the workloads studied follow the same jobs distribution, and that the total load remains modest. An analysis of more workloads, with different profiles, may lead to different results.

Similarly, for electrical sizing, a clear decrease in battery capacity had little impact on the rest of the configurations, both for solar panels and hydrogen storage. In the case of an unexpected event leading to a higher electrical demand, an overconsumption of H_2 can be envisaged, and compensated afterwards on favorable weather conditions, as H_2 is a long-term storage that smooths out epiphenomena. Buying green H_2 remains an open option, where this purchase can be offset by future sales of H_2, again under favorable weather conditions. Overall a reduction of 30% of the battery capacity finally leads to a negligible impact on the overall electrical infrastructure.

However, it is important to estimate the total cost of the infrastructure, both in terms of budget and ecological impact, to measure a real positive impact. For this, a life cycle (LCA) and a budget analysis of the infrastructure are needed. Further experiments need to be conducted, always with the aim of reducing the cost of the infrastructure, in terms of economic and environmental impacts.

Acknowledgement

This work was supported in part by the ANR DATAZERO2 (contract “ANR-19-CE25-0016”) project and by the EIPHI Graduate school (contract "ANR-17-EURE-0002"). We also thank Mr. Igor Fontana de Nardin for his help and valuable discussions on the presentation of green computing challenges.

References

[1] U. Cisco, Cisco annual internet report (2018–2023) wp (2020).

[2] E. e. a. Masanet, Recalibrating global data center energy-use estimates, Science 367 (6481) (2020) 984–986.

[3] W. Xia, P. Zhao, Y. Wen, H. Xie, A survey on data center networking (dcn): Infrastructure and operations, IEEE communications surveys & tutorials 19 (1) (2016) 640–656.

[4] S. Project, Lean ict: Towards digital sobriety (2019).

[5] V. H. et al., Trends in worldwide ICT electricity consumption from 2007 to 2012, Computer Communications 50 (2014) 64–76.

[6] Y. Yao, J.-H. Xu, D.-Q. Sun, Untangling global levelised cost of electricity based on multi-factor learning curve for renewable energy: Wind, solar, geothermal, hydropower and bioenergy, Journal of Cleaner Production 285 (2021) 124827.

[7] J.-M. e. a. Pierson, DATAZERO: DATAcenter with Zero Emission and RObust management using renewable energy, IEEE Access 7 (2019) 103209 – 103230.

[8] L. A. Barroso, U. Hölzle, The case for energy-proportional computing, Computer 40 (12) (2007) 33–37.

[9] L. et al., Optimal sizing of energy station in the multienergy system integrated with data center, IEEE Transactions on Industry Applications 57 (2) (2021) 1222–1234.

[10] L. Cupelli, N. Barve, A. Monti, Optimal sizing of data center battery energy storage system for provision of frequency containment reserve, in: IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society, 2017, pp. 7185–7190. doi:10.1109/IECON.2017.8217257.

[11] K. e. a. Anoune, Sizing methods and optimization techniques for pv-wind based hybrid renewable energy system: A review, Renewable and Sustainable Energy Reviews 93 (2018) 652–673.

[12] M. Haddad, G. Da Costa, J.-M. Nicod, M.-C. Péra, J.-M. Pierson, V. Rehn-Sonigo, P. Stolf, C. Varnier, Combined it and power supply infrastructure sizing for standalone green data centers, Sustainable Computing: Informatics and Systems (2021) 100505.

[13] G. Da Costa, L. Grange, I. De Courchelle, Modeling, classifying and generating large-scale google-like workload, Sustainable Computing: Informatics and Systems 19 (2018) 305–314.

[14] W. Cole, A. W. Frazier, C. Augustine, Cost projections for utility-scale battery storage: 2021 update, Tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States) (2021).

[15] M. Taylor, P. Ralon, A. Ilas, The power to change: solar and wind cost reduction potential to 2025, International renewable energy agency.

[16] P. Perez-Lopez, R. Jolivet, I. Blanc, R. Besseau, M. Douziech, B. Gschwind, S. Tannous, J. Schlesinger, R. Briere, A. Prieur-Vernat, et al., Incer-acv project. uncertainties in life-cycle environmental impact assessment methods of energy production technologies. final report, Tech. rep., International Atomic Energy Agency (2021).

[17] G. Parks, R. Boyd, J. Cornish, R. Remick, Hydrogen station compression, storage, and dispensing technical status and costs: Systems integration, Tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States) (2014).

[18] B. James, C. Houchins, J. Huya-Kouadio, D. Desantis, Final report: Hydrogen storage system cost analysis 2016.