SEVENTH FRAMEWORK PROGRAMME

GreenDataNet

D2.2 – Analytical Models for DC

Status: Final

Prof. David Atienza
Embedded Systems Laboratory (ESL) – EPFL

Rev 4.3

Contributors:

- Ali Pahlevan, EPFL
- Pablo Garcia, EPFL
- Davide Brunelli, UNITN
- Maurizio Rossi, UNITN

TABLE OF CONTENTS

REVISION SHEET

Revision Number	Date	Brief summary of changes
Rev 1.0	17/02/2015	Baseline document.
Rev 2.0	05/03/2015	In-detail description of the model.
Rev 3.0	19/03/2015	Added new model from UNITRN.
Rev 3.1	22/03/2015	Added explanation of the model. Formatted text.
Rev 4.0	29/03/2015	Added experimental results and review feedback.
Rev 4.1	30/03/2015	Added review feedback (from EPFL). Updated index.
Rev 4.2	30/03/2015	Added review feedback (from Trento).
Rev 4.3	31/03/2015	Layout update

KEY REFERENCES AND SUPPORTING DOCUMENTATIONS

[1] CEN/CLC/ETSI CG GDC : *Standardization landscape for energy management and environmental viability of green Data Centres* (Jan 2014) ftp://ftp.cencenelec.eu/EN/EuropeanStandardization/HotTopics/ICT/GreenDataCentres/GDC_report.pdf

[2] Whitney, Josh, Delforge, Pierre et al.: *Data Center Efficiency Assessment Scaling Up Energy Efficiency Across the Data Center Industry: Evaluating Key Drivers and Barriers.* Natural Resources Defense Council, Issue Paper, (August 2014) http://www.nrdc.org/energy/files/data-center-efficiency-assessment-IP.pdf

[3] *System Modelling and energy management for grid connected PV systems with storage.* 23rd European Photovoltaic Solar Energy Conference and Exhibition, 1-5 September 2008, Valencia, Spain.

[4] *An in depth analysis of the maths behind Peukert's Equation (Peukert's Law)* http://www.smartgauge.co.uk/peukert_depth.html

[5] T. Benson and et al. *Understanding data center traffic characteristics*. ACM SIGCOMM Computer Communication Review, 40(1):92{99, 2010}.

[6] M. Pedram and et al. *Power and performance model ing in a virtualized server system*. In Parallel Processing Workshops (ICPPW), 2010 39th International Conference on, pages 520-526

[7] GreenDataNet Deliverable: D2.3 - Server Multi-level Software Management Specification and Implementation

[8] GreenDataNet Deliverable: D2.4 - Racks Multi-objective Energy Management Specification and Implementation

[9] GreenDataNet Deliverable: D3.2 - Electricity Consumption Forecasting Tool Design and Implementation

[10] GreenDataNet Task: T4.1 - Specification of Demonstration Sites and Validation Methodology

[11] GreenDataNet FP7 Collaborative Project Green and Smart Data Centres Network Design

[12] *Eaton 93PM 30-50kW UPS Technical Specification*.

[13] PUE definition in Wikipedia: http://en.wikipedia.org/wiki/Power_usage_effectiveness

[14] KMEANS clustering algorithm in Wikipedia: http://en.wikipedia.org/wiki/K-means_clustering

[15] J. Kim, et al., “Free cooling-aware dynamic power management for green datacenters,” in Proc. HPCS, 2012.

1. INTRODUCTION

The main goal of the GreenDataNet project is to apply new technologies and tools to enhance the power efficiency of the urban data centres. In the pursuit of this goal, it is crucial to minimize the energy consumption of both information technology (IT) and cooling by using renewable energies and optimizing the allocation of tasks to servers. Therefore, one of the outcomes of GreenDataNet is to provide tools for analysis and optimization of green datacentres.

In this deliverable, we develop analytical models that describe the physical structure of a datacentre; from aspects like the properties of the workload of the datacentre (CPU usage, communication patterns…) and the dynamic behaviour of the tasks (short living, long living), to the relation between generated power, cooling infrastructure, control knobs, and temperature. The analytical model combines it all to enable realistic simulations of online thermal management and multi-objective energy optimization of datacentres.

1.1 DOCUMENT PURPOSE

The roadmap of the GreenDataNet project started with the WP1, where a study of the state-of-the-art in DCs was made. The trends in this area allowed us not only to define the HW and SW characteristics of future urban green datacentres, but also the specification of the infrastructure surrounding the IT equipment: cooling system, IT room, DC facility, and inter-DC communication network. Then, with WP2 and WP3 started the implementation of a methodology to design and study future DCs: WP2 creating the management system to optimize the power consumed, and WP3 adding the smart grid integration and the forecasting algorithms that will ensure maximum renewable energy utilization.

Deliverable D3.2 introduced the electricity consumption forecasting tool, with the key component being an engine to estimate the power consumption in the datacentre. This consumption was related to the configuration of the datacentre, of course, as well as to external parameters, like the cost of the electricity, or the availability of renewable energy. The tool, then, will use all these information to minimize the energy consumption by choosing an adequate allocation of tasks (Virtual Machines, or VMs) to servers.

In this deliverable, we revisit this tool, but describing the problem from the analytical point of view: First of all, we define in an accurate way, through equations, the behaviour and constraints of the different components (not only the IT infrastructure, but also the renewable energy sources, batteries…). Then, in the same way, we describe the metrics that allow us to assess the efficiency of the datacentre and set the optimization constraints. Finally, the algorithm that solves the problem is presented, along with a real-life example that demonstrates the internals of the model.

The complete specification of this model is the first step towards the creation of higher level tools that will further optimize the operation of green datacentres, like the ones that will be investigated in deliverables D2.3 and D2.4:

- D2.3 aims at implementing multi-level SW management algorithms that work, first, using distributed local controllers (with different heuristics) and, then, at a global level, using an advanced hypervisor (hierarchical controller) to coordinate the decisions coming from the multiple local controllers.
- D2.4 explores the existing trade-offs between system-level performance, power consumption and temperature at the rack level. The rack controller will interact with the server controller to define how to jointly adjust rack cooling, workload allocation and server power state at runtime.

For D2.3 and D2.4 to provide effective management algorithms, we need a model of the DCs, accurate enough, to provide realistic simulations and allow complete studies. Therefore, the analytical model described here will be used as the tool to validate the proposed techniques hereafter.

1.2 DOCUMENT OVERVIEW

The rest of the document explains the details of the analytical model used in GreenDataNet to characterize green datacentres. Since, typically, they are made by one or more DC locations, we start in Section 2 by introducing the model for the network latency (both inside a DC and among datacentres). Then, section 3 describes the rest of the components (system and IT power and battery models), so that the whole optimization problem can be defined (Section 4). Next, Section 5 details the challenges of this problem, and proposes the algorithm to resolve it. Finally, Section 6 contains a small experiment presented as an example to demonstrate how the model and the algorithm work.

2. NETWORK MODEL

From the communications point of view, a GreenDataCenter is composed of machines that interchange information. The information is stored locally in network-attached storage devices, and it can be transferred from one DC to another one. Therefore, for the network model, we have considered a full duplex peer-to-peer global optical fibre link between each two datacentres and a local link inside each datacentre to access to its network-attached storage. For global and local connections, we have considered the backbone (B_{bb}) and Local bandwidth (B_L) respectively. Additionally, we take into account the bit error rates and their probabilities associated to the transmission. In the equations, the speed of light and distance between two datacentres are also present to model global link.

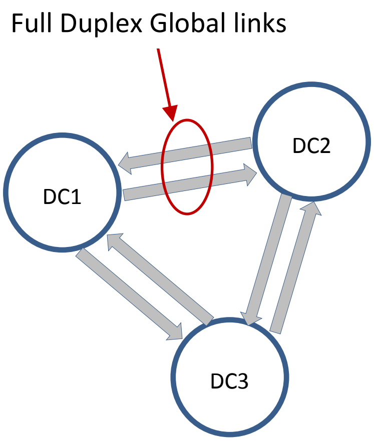

Figure 2.1 – Network Model: Global connections of the GreenDataCentre

2.1 LATENCY MODEL

To compute the total latency (L_t^j) for migrating a set of VMs from all datacenters to a certain datacenter, we take into account two parts:

1) Local and global latency for the ith source datacentre, i.e. L_l^i and $L_g^{i,j}$ respectively, to transmit selected VMs through the local and global networks to the jth destination datacentre
2) Local latency for the jth destination datacentre (L_l^j) to transmit VMs collected from all datacenters to its network-attached storage.

Equation 1 represents that the total latency for the jth destination datacentre to receive the collected VMs from the sources is equal to the summation of the maximum latency for transmitting the corresponding VMs through local and global links to the destination datacentre among all source datacentres and the local latency inside the destination datacentre.

Equation 1

$$L_t^j = max_i \left(L_l^i + L_g^{i,j}\right) + L_l^j \qquad i = 1\ to\ N_{DC}\ \ and\ \ i \neq j$$

Equation 2 states that the local latency of the ith source datacentre is dependent on the total size of the VMs ready to be transferred to jth destination datacentre ($Size_t^{i,j}$) and its local bandwidth.

Equation 2

$$L_l^i = \frac{Size_t^{i,j}}{B_L}$$

The local latency of the jth destination datacentre is related to the total size of VMs received from the source datacentres and its local bandwidth. This constraint is specified in Equation 3.

Equation 3

$$L_l^j = \frac{\sum_{i=1,i\neq j}^{N_{DC}} Size_t^{i,j}}{B_L}$$

Equation 4 is used to compute the global latency between two datacentres through the global link. The global latency includes propagation latency as a primary source and data latency with regard for the amount of data being transmitted. Propagation latency is a function of how long it takes information to travel at the speed of light (S_l) in the communications media from source to destination distance ($Dist_{i,j}$). Data latency is a function of effective bandwidth ($B_e(t)$) in presence of bit error rate ($BER(t)$) to resend the lost data until all data (VMs) are transmitted to destination properly.

Equation 4

$$L_g^{i,j} = \frac{Dist_{i,j}}{S_l} + L_e$$

To compute the data latency in presence of bit error rate (L_e), first we calculate the effective bandwidth at each time (every second), then we send the data through the channel. In this case, if the size of data is more than the effective bandwidth we send a part of data in a second and then, at the next time, we compute the current effective bandwidth to send the remaining data until all data is sent (i.e. the size of data becomes less than the effective bandwidth). Algorithm 1 describes this process analytically.

Algorithm 1:

$while(1)\{$

$$B_e(t) = B_{bb} - B_{bb} \times BER(t) = (1 - BER(t)) \times B_{bb}, \qquad BER(t) \propto P_{BER}(i)$$

$$\begin{cases} L_e = L_e + \frac{Size_t^{i,j}}{B_e(t)} \;\; and \;\; Break; & if \;\; Size_t^{i,j} \leq B_e(t) \\ Size_t^{i,j} = Size_t^{i,j} - B_e(t) \;\; and \;\; L_e = L_e + 1; & if \;\; Size_t^{i,j} > B_e(t) \end{cases}$$

}

Table 2.1 and **Table** 2.2 show all of the parameters and variables used for network model throughout this document.

Table 2.1 – Network notation and definitions [Parameters]

Symbol		Definition
B_{bb}		Backbone Bandwidth (100 Gb/s)
B_L		Local Bandwidth in Each DC (10 Gb/s)
S_l		Speed of Light (300,000 Km/s)
N_{BER}		Number of Bit Error Rate (5)
$P_{BER}(i)$	$1 \leq i \leq N_{BER}$	Probability of Each BER (0.54, 0.2, 0.15, 0.1, 0.01)
$Dist_{i,j}$		Distance Between DCi and DCj (2000, 4000, 2500 Km)
$Size_{VM}^{k}$		k^{th} VM Size (2, 4, 8 GB)
N_{DC}		Number of DCs

Table 2.2 - Network notation and definitions [Variables]

Symbol	Definition
$B_e(t)$	Effective Bandwidth at Time t
$BER(t)$	Bit Error Rate at Time t ($10^{-6}, 10^{-5}, \dots, 10^{-2}$)
$Size_t^{i,j}$	Total Size of VMs to be Transferred from DC_i to DC_j
$DataS_t^{i,j}$	Total Amount of Data to be Transferred from DC_i to DC_j Due to Data Correlation
L_t^j	Total Data (VMs) Latency from Other DCs to DC_j Until Running VMs on DC_j
L_d^j	Total Amount of Data Latency from Other DCs to DC_j in Presence of Data Correlation (i.e. Amount of data should be transmitted online between two VMs in different DCs within time t and t+1. Note that this latency is completely different from VM migration)
L_l^j	Local Link Latency Inside of DC_j
$L_g^{i,j}$	Global Link Latency from DC_i to DC_j
L_e	Global Error Latency Due to BER

3. DATACENTER SYSTEM MODEL

As introduced in the previous section. a GreenDataCentre is modelled as a network of interconnected DCs. This section presents the datacentre system framework and its power management model. Two components account for the total power consumption: IT equipment and cooling unit.

3.1 SYSTEM AND POWER MANAGEMENT MODEL

Figure 3.1 depicts the schematic view of the system model used to analyse the power flows in a datacentre, according to the GDN specifications (Document [11] and Deliverable D3.2 [9]). For the initial scenario, we assume two simplifications with respect to the general model (more details in D3.2 [9]): one single renewable source, and no injection of energy back to the Grid.

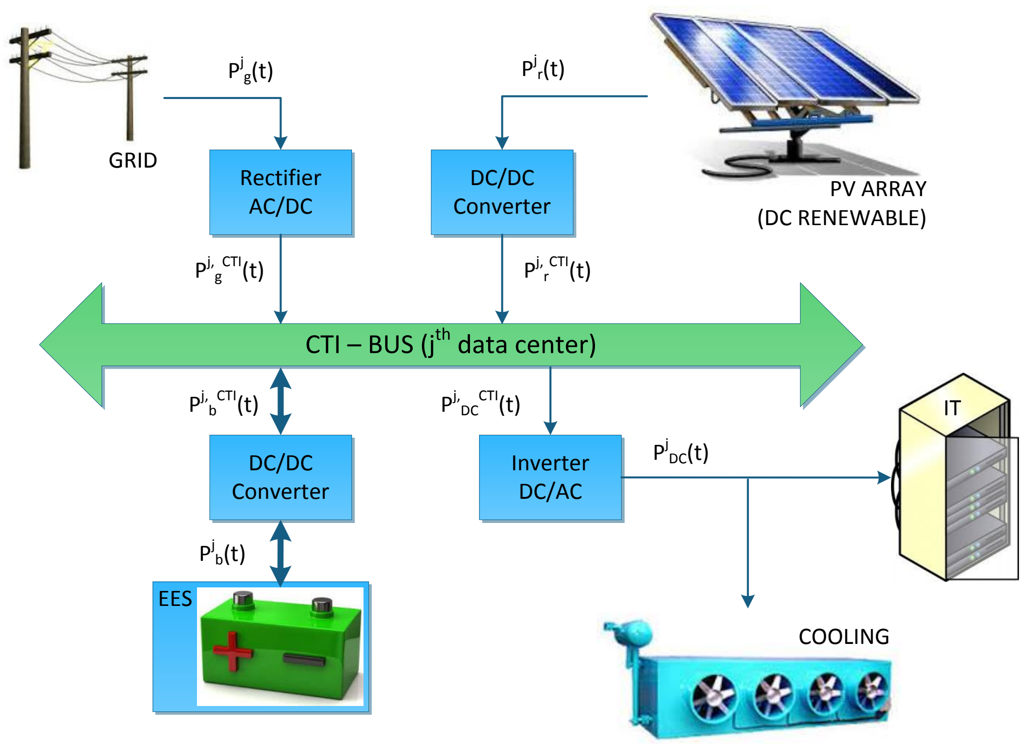

Figure 3.1 – Overview of the system and management model

The power management problem is analysed and solved at the Charge Transfer Interconnect bus (CTI) level which is a Direct Current (DC) path. Conversely, the system comprises both AC and DC sources/loads thus, for the former ones, it is required to consider the power factor component in the conversion. For example, considering the power intake from the Grid, if we measure the total apparent power that inputs the rectifier ($\mathrm{P_g^j}$ [VA] $= \mathrm{V_{RMS} \cdot I_{RMS}}$), this can be converted into active power (the useful power available on the DC side) according to the $\mathrm{P_g^{j,CTI}}$ [W] $= \mathrm{P_g^{j,CTI}}$ [VA] $\cdot$ $\cos(\varphi)$ where φ is the angle between Voltage and Current waves. In addition, the converter's efficiency must be added to any transformation, since it depends on the actual power flowing with respect to the nominal one. The characteristic curve (or table) of the efficiency must be provided by the manufacturer or empirically evaluated.

The following set of equations represents the analytical model of the schema depicted in Figure 3.1. The first one (Equation 5) is the power balance equation that states that the sum of the input from the Grid, PV and battery arrays must be equal to the datacentre requirements. For the battery array term there is a directional parameter, b, which can be -/+1 depending on the charging/discharging status (source or load of the system).

Equation 5

$$P_{DC}^{j,CTI}(t) = P_g^{j,CTI}(t) + \sum_{m=1}^{M} P_{r_m}^{j,CTI}(t) + b \cdot P_b^{j,CTI}(t)$$

The following equations describe the AC-to-DC and DC-to-DC conversion functions used for each system component.

Equation 6

$$P_g^{j,CTI}(t) = P_g^j(t) \cdot \cos(\varphi) \cdot Eff_{Rect}^j\big(\rho(t)\big) \Longrightarrow V_g^{j,CTI}(t) \cdot I_g^{j,CTI}(t) = P_g^j(t) \cdot \cos(\varphi) \cdot Eff_{Rect}^j\big(\rho(t)\big)$$

Equation 7

$$P_{r_m}^{j,CTI}(t) = P_{r_m}^j(t) \cdot \cos(\varphi)^d \cdot Eff_{Rect}^j\big(\rho(t)\big) \Rightarrow V_{r_m}^{j,CTI}(t) \cdot I_{r_m}^{j,CTI}(t) = P_{r_m}^j(t) \cdot \cos(\varphi)^d \cdot Eff_{Rect}^j\big(\rho(t)\big)$$

Equation 8

$$P_b^{j,CTI}(t) = P_b^j(t) \cdot \left(Eff_{Conv}^j(\rho(t))\right)^b \Longrightarrow V_b^{j,CTI}(t) \cdot I_b^{j,CTI}(t) = P_b^j(t) \cdot \left(Eff_{Conv}^j(\rho(t))\right)^b$$

Equation 9

$$P_{DC}^{j,CTI}(t) = P_{DC}^j(t) \cdot \cos(\varphi) \cdot Eff_{Trnf}^j\big(\rho(t)\big) \Longrightarrow V_{DC}^{j,CTI}(t) \cdot I_{DC}^{j,CTI}(t) = P_{DC}^j(t) \cdot \cos(\varphi) \cdot Eff_{Trnf}^j\big(\rho(t)\big)$$

According to EATON's GDN UPS specifications [12], the CTI voltage level can be considered constant (720VDC); thus, the optimization problem results simplified since it's a design parameter that doesn't depend on time:

Equation 10

$$V_g^{j,CTI}(t) = V_{r_m}^{j,CTI}(t) = V_b^{j,CTI}(t) = V_{DC}^{j,CTI}(t) = V^{j,CTI} = 720\ [V]$$

3.2 IT EQUIPMENT POWER MODEL

The power consumption of the IT equipment: server ith in datacentre jth, is composed of static ($P_{i,Static}^{j}$) and dynamic ($P_{i,Dynamic}^{j}$) power when a server is in active mode. Then, $U_i^j(t)$ represents CPU utilization of the ith server in the jth datacentre.

Equation 11

$$P_i^j(t) = P_{i,Static}^{j} + P_{i,Dynamic}^{j} \times U_i^j(t)$$

Therefore, the power consumed by server clusters in the jth datacentre can be calculated as the sum of power consumption of its corresponding servers, as follows.

Equation 12

$$\boldsymbol{P_s^j(t)} = \sum_{i=1}^{N_s^j} \boldsymbol{P_i^j(t)}$$

According to the definition of PUE [13], the power consumed by the cooling system in the jth datacenter, $P_c^j(t)$, can be calculated as follows:

Equation 13

$$\boldsymbol{P_c^j(t) = (PUE^j(t) - 1).\, P_s^j(t)}$$

In the general case, the PUE of a DC varies along time. More exactly, it depends on the temperature of the room, on the ambient temperature, and on the power consumed by the server clusters: $PUE^j(t) = f\ (T_{room}, T_{amb}, P_S^j(t))$. For clarification purposes, however, in the experiments section, a constant value for the PUE, specific to each DC, will be assumed.

As stated in the introduction of this section, the total power consumed in the datacentre is the addition of the IT requirements power and the cooling power, defined in Equation 12 and Equation 13, respectively, and can be written as:

Equation 14

$$P_{DC}^j(t) = P_S^j(t) + P_c^j(t)$$

3.3 BATTERY MODEL

The battery model is based on the work proposed by CEA's group in [3]. The goal is to model Hybrid Electrical Systems (HES), that combine the advantages of the different battery technologies (lead-acid and lithium-ion). For this first version of the model, only one battery array is considered. The module, as all the modules in the model, has been conceived as a plug-and-play component; therefore, upon modifications and future upgrades of the model, it can be easily replaced by the latest version.

Equation 15 defines the State of Health of the battery (SoH) as a ratio between currently available charge capacity and the nominal one. Equation 16 defines the charge capacity as a linear combination of the previous Charge and a term that depends on the charge drained. The following two equations (Equation 16 and Equation 17) allow to determine the State of Charge and the equivalent battery current with respect to the nominal battery parameters. The role of these two equations is explained in detail in [4].

The SoH of the battery decreases only during discharge, so it is calculated only during discharge, whereas the SoC is computed during both charge and discharge cycles.

Equation 15

$$SoH_b^j(t+1) = \frac{C_{b,ref}^j(t+1)}{C_{b,Nom}^j}$$

Equation 16

$$C_{b,ref}^j(t+1) = C_{b,ref}^j(t) - C_{b,Nom}^j \cdot Z_b \cdot \left(SoC_b^j(t) - SoC_b^j(t+1)\right)$$

Equation 17

$$SoC_b^j(t+1) = \frac{C_{b,ref}^j(t) \cdot SoC_b^j(t) - \left(I_{b,eq}^j(t) \cdot t_{SL}\right)}{C_{b,ref}^j(t)}$$

Equation 18

$$I_{b,eq}^j(t) = \left(\frac{\left|I_b^j(t)\right|}{I_{b,ref,b}^j}\right)^{(k_b-1)} \cdot I_b^j(t)$$

3.4 SUMMARY OF THE PROBLEM NOTATION AND DEFINITIONS

As a summary, **Table** 3.1 and **Table** 3.2 contain all of the parameters and variables used to define the problem statement and the datacentre system model throughout this document.

Table 3.1 – Problem notation and definitions [Parameters]

Symbol	Definition
$P_{i,Static}^{j}$	Static Power of i^{th} Server in j^{th} DC
$P_{i,Dynamic}^{j}$	Dynamic Power of i^{th} Server in j^{th} DC
N_{s}^{j}	Number of Servers in j^{th} DC
$th_{b}^{ch,min}, th_{b}^{ch,max}$	Min/Max thresholds for battery charge phase
$th_{b}^{ds,min}, th_{b}^{ds,max}$	Min/Max thresholds for battery discharge phase
$\cos(\varphi)$	Power Factor, to convert Apparent power [VA] in Active power [W]
d	Renewable selector: DC source (0) AC source (1), binary value
p	Penalty parameter Unit to compute Cost of Battery-Usage
$C_{b,Nom}^{j}$	Nominal Charge Capacity of Batteries in j^{th} DC (new device, from datasheet)
Z_{b}	Correction factor = $3\text{x}10^{-4}$
$V^{j,CTI} = V_{DC}^{j,CTI} = V_{g}^{j,CTI} = V_{r}^{j,CTI} = V_{b}^{j,CTI}$	Voltage in the CTI-BUS in j^{th} DC = 720 [V]
$I_{b,ref,b}^{j}$	Reference charge/discharge Current for Batteries in j^{th} DC (from datasheet)
k_{b}	Peukert's coefficient depending on battery technology/model, can be estimated using reference charge/discharge currents
$PR_{g}^{j}(t)$	Electricity Price for j^{th} DC at Time t (function of time-slot length)
N_{VM}	Total Number of VMs Running on DCs
$\widehat{MIPS}_{VM}^{i}(t)$	Maximum VM MIPS Required in Period of [t-1, t)
$E_{r,max}^{j}(t)$	Maximum Available Renewable Energy in j^{th} DC at Time t
DoD	Depth of Discharge of Battery
E_{BatMax}^{j}	Maximum Energy Capacity of Battery in j^{th} DC
$f_{i,max}^{j}$	Maximum Frequency of i^{th} Server in j^{th} DC
$N_{i,Core}^{j}$	Number of Cores of i^{th} Server in j^{th} DC
$N_{i,f}^{j}$	Number of Frequency Levels of i^{th} Server in j^{th} DC
$f_{i,k}^{j} \quad 1 \le k \le N_{i,f}^{j}$	k Frequency Values of i^{th} Server in j^{th} DC
$\widehat{MIPS}_{VM}^{i,j}(t)$	Worst-case Peak MIPS when the Peaks of Two VMs (i and j) Coincide in Period of [t-1, t)
$DataCorr_{i,j}^{VM}(t)$	Data Correlation Between i^{th} and j^{th} VMs (i.e. amount of data should be transmitted between two VMs within time t and t+1)
TH_{QoS}	QoS Threshold as a Network Latency to Migrate the VMs

Table 3.2 - Problem notation and definitions [Variables]

Symbol	Definition
t_{SL}	Integration time (or time-slot length in discrete time analysis: 1 hr, 10 min, …)
$U_i^j(t)$	Utilization of ith Server in jth DC at Time t
$P_i^j(t)$	Power Consumption of ith Server in jth DC at Time t
$P_s^j(t)$	Total Computing Power Consumption of jth DC at Time t
$P_c^j(t)$	Cooling Power Consumption of jth DC at Time t
$P_{DC}^j(t)$	Total Computing Power Consumption of jth DC at Time t
$P_g^j(t)$	Power taken from the Grid in jth DC at Time t
$P_{r_m}^j(t)$	Power taken from the mth renewable source in jth DC at Time t
$P_b^j(t), V_b^j(t), I_b^j(t)$	Power, Voltage and Current taken (discharge case) from (provided to in recharge case) the Battery in jth DC at Time t
$\rho(t)$	Ratio of Output [$P_{out}^{Conv}(t)$] / Nominal [$P_{Nom}^{Conv}(t)$] (rated output in the datasheet) Power to compute converters' efficiency at Time t: $P_{out}^{Conv}(t) \big/ P_{Nom}^{Conv}(t)$
$P_{DC}^{j,CTI}(t), I_{DC}^{j,CTI}(t)$	Total Computing Power and Current used for computation of jth DC at Time t
$P_g^{j,CTI}(t), I_g^{j,CTI}(t)$	Power and Current of the Grid component in the CTI-BUS in jth DC at Time t
$P_{r_m}^{j,CTI}(t), I_{r_m}^{j,CTI}(t)$	Power and Current of the renewable source in the CTI-BUS in jth DC at Time t
$P_b^{j,CTI}(t), I_b^{j,CTI}(t)$	Power and Current taken (discharge case) from (provided to in recharge case) the Battery Banks in jth DC at Time t
$V^{j,CTI}(t)$	Voltage in the CTI-BUS in jth DC at Time t
$SoH_b^j(t)$	State of Health of batteries in jth DC
$SoC_b^j(t)$	State of Charge of batteries in jth DC
$C_{b,ref}^j(t)$	Current Charge Capacity of Batteries in jth DC at time t (Coulomb, where 3600=1Ah)
$I_b^j(t)$	Current to/from the Batteries in jth DC at time t
$I_{b,eq}^j(t)$	Equivalent Current to/from the Batteries in jth DC at time t (according to Peukert's model)
$Eff_{Trnf}^j(\rho)$	Efficiency of Transformers (DC-AC) in jth DC
$Eff_{Rect}^j(\rho)$	Efficiency of Rectifiers (AC-DC) in jth DC
$Eff_{Conv}^j(\rho)$	Efficiency of Converters (DC-DC) in jth DC
$E_g^j(t)$	Energy Taken from the Grid in jth DC at Time t

$E_r^j(t)$	Energy Taken from the Renewable Sources in j^{th} DC at Time t
$E_b^j(t)$	Energy Taken from the Battery Banks in j^{th} DC at Time t
b	Battery charge (-1) / discharge (1)
$E_{DC}^j(t)$	Total Energy of j^{th} DC at Time t
$Cycle_b^j(t)$	Number of Battery Charge/Discharge Cycles in j^{th} DC at Time t
$C_{cc,dd}^j(t)$	Cost of Battery for Charge to Charge and Discharge to Discharge States in j^{th} DC in Period of [t-1, t]
$C_{cd,dc}^j(t)$	Cost of Battery for Charge to Discharge and Discharge to Charge States in j^{th} DC in Period of [t-1, t]
$N_{VM}^{n,j}(t)$	Number of Selected VMs to be Transferred from DC_n to DC_j at Time t
$N_{VM}^j(t)$	Number of VMs Running on j^{th} DC
$N_{i,VM}^j(t)$	Number of VMs Allocated to i^{th} Server in j^{th} DC
$E_{BatAvailabe}^j(t)$	Available Energy in Battery in j^{th} DC at Time t
$f_i^j(t)$	The Frequency of i^{th} Server in j^{th} DC at Time t (discrete)
$f_{i,w}^j(t)$	Determined Frequency of i^{th} Server in j^{th} DC at Time Considering the CPU Correlation (continuous)
$r_{i,k}^j(t)$	i^{th} Server in j^{th} DC Runs at k^{th} Frequency Level at Time t
$Place_{i,k}^j(t)$	The Placement of k^{th} VM in i^{th} Server in j^{th} DC at Time t
$Cost_{i,j}^{VM}(t)$	CPU Correlation Cost Between i^{th} and j^{th} VMs
$Cost_{i,Server}^j(t)$	CPU Correlation of i^{th} Server in j^{th} DC at Time t
$w_{i,k}^j(t)$	Weight of k^{th} VM Allocated to i^{th} Server in j^{th} DC at Time t

4. PROBLEM STATEMENT

The complete optimization problem consists in dispatching VMs among several datacentres, and allocating them to the servers using v/f scaling, while respecting the constraints described in the previous chapter. More exactly, the problem requires the allocation of VMs at a global level, fully exploiting the CPU and data correlation among VMs, in order to minimize the total money spent to purchase energy from the Grid (as the sum of the expenses in each DC location), and finally to obtain the best battery pack utilization, meaning to use as much as possible of it, with continuous c/d cycles. Note that, in our current setup, only one battery pack is managed by the Dispatcher at the global level. In this section, the proposed method is formulated using equations.

4.1 COST FUNCTION

Equation 19 is an objective function to minimize the energy cost of geo-distributed datacentres, by maximizing the usage of free energies, along with maximizing battery lifetime.

Equation 19 – Cost function

$$\min \sum_{j=1}^{N_{DC}} \left[\left(\frac{E_g^j(t)}{\left(E_r^j(t)+E_b^j(t)\right).1} \right) . PR_g^j(t) + Cycle_b^j(t) \right]; \qquad Cycle_b^j(t) = C_{cc,dd}^j(t) + C_{cd,dc}^j(t)$$

4.2 CONSTRAINTS

4.2.1 ENERGY CONSTRAINTS

On-site renewable energy, such as solar panels or wind turbines, is produced by each datacentre to reduce the carbon footprint and the electricity cost. In this particular case, we assume that datacentres are powered by solar energy only. Then, the total amount of energy used by the jth datacentre ($E_{DC}^j(t)$ in Equation 20) is the sum of energies taken from the grid ($E_g^j(t)$), renewable ($E_r^j(t)$) and battery ($E_b^j(t)$) sources when binary variable b is '1'. If battery is in charging mode, b is '-1' and b is '1' for discharging mode. This amount of energy is equal to the energy consumption of the jth datacentre used by the IT equipment and cooling system (Equation 21).

There are some constraints on utilizing renewable and battery energies. Equation 22 shows that the available battery energy in time slot t ($E_{BatAvailable}^j(t)$) is equal to the amount of available battery energy from the previous time slot and given/taken battery energy according to charge/discharge state. The amount of energy taken from the renewable source and available energy stored in the battery is lower and upper bounded by Equation 22 and Equation 23.

Equation 20

$$E_{DC}^{j}(t) = E_{g}^{j}(t) + E_{r}^{j}(t) + b.E_{b}^{j}(t)$$

Equation 21

$$E_{DC}^{j}(t) = P_{DC}^{j}(t) \cdot t_{SL}$$

Equation 22

$$E_{BatAvailable}^{j}(t) = E_{BatAvailable}^{j}(t-1) - b.E_{b}^{j}(t)$$

Equation 23

$$0 \leq E_{r}^{j}(t) \leq E_{r,max}^{j}(t)$$

Equation 24

$$DoD \times E_{BatMax}^{j} \leq E_{BatAvailabe}^{j}(t) \leq E_{BatMax}^{j}$$

4.2.2 SERVERS (IN ALL DCS) UTILIZATION AND FREQUENCY CONSTRAINTS

The following constraints determine the utilization of active servers with respect to their selected frequencies in each datacentre guaranteeing the number of active servers does not exceed the total number of available servers in each datacentre.

Equation 25

$$U_{i}^{j}(t) = \frac{f_{i}^{j}(t)}{f_{i,max}^{j}}$$

Equation 26

$$0 \leq U_{i}^{j}(t) \leq 1$$

Equation 27

$$\sum_{k=1}^{N_{i,f}^{j}} r_{i,k}^{j}(t) \leq 1$$

Equation 28

$$r_{i,k}^{j}(t) = \{0,1\}$$

Equation 29

$$f_{i}^{j}(t) = \sum_{k=1}^{N_{i,f}^{j}} f_{i,k}^{j}.r_{i,k}^{j}(t)$$

Equation 30

$$\sum_{i=1}^{N_s^j}\sum_{k=1}^{N_{i,f}^j} r_{i,k}^j(t) \leq N_s^j$$

4.2.3 VM CONSTRAINTS

Regarding the constraints for the VMs, we guarantee that each VM is allocated only to one server and emphasize that VMs placed in a server should not exceed the server capacity. Then, we map the optimal selected frequency, in continuous range, of each server to the closest available discrete frequency (from the frequency levels set) for each server. After allocating all the VMs, the number of migrated VMs from other datacentre to the j^{th} datacentre can be also calculated as follows.

Equation 31

$$N_{VM} = \sum_{j=1}^{N_{DC}} N_{VM}^j(t)$$

Equation 32

$$N_{VM}^j(t) = N_{VM}^j(t-1) + \sum_{i=1, i\neq j}^{N_{DC}} N_{VM}^{i,j}(t) - \sum_{i=1, i\neq j}^{N_{DC}} N_{VM}^{j,i}(t)$$

Equation 33

$$\sum_{j=1}^{N_{DC}}\sum_{i=1}^{N_s^j} Place_{i,k}^j(t) = 1$$

Equation 34

$$Place_{i,k}^j(t) = \{0,1\}$$

Equation 35

$$N_{i,VM}^j(t) = \sum_{k=1}^{N_{VM}} Place_{i,k}^j(t)$$

Equation 36

$$\sum_{i=1}^{N_s^j}\sum_{k=1}^{N_{VM}} Place_{i,k}^j(t) = N_{VM}^j(t)$$

Equation 37

$$f_{i,w}^{j}(t) \leq f_{i}^{j}(t)$$

Equation 38

$$f_{i}^{j}(t) \leq f_{i,max}^{j}$$

4.2.4 VM DEPENDENCIES (CORRELATION) CONSTRAINTS

Variability and fast-changing characteristics of scale-out applications affect the energy consumption of servers due to the dependency to external factors, e.g., number of clients/queries in the system. To this end, the impact of the energy consumption of the servers on the usage of green energy becomes more substantial and the management of consumed energy will play a major role in lifetime and operation of energy storage systems.

Due to the correlation of CPU utilization among virtual machines within a cluster of applications in virtualized datacentres (CPU correlation), we have considered a correlation-aware VM allocation scheme as the datacentre power management solution. The CPU correlation-aware VM allocation method has been proposed to efficiently compact more VMs (in terms of CPU Million Instructions per Second (MIPS)) to the lowest number of servers across a certain time horizon. Finally, an optimal voltage/frequency (V/f) level is provided to achieve power savings without any QoS degradation. The VMs are allocated such that the CPU correlation among the allocated VMs in the server is minimized, and the number of the active servers is minimized while satisfying performance requirements.

Equation 39

$$Cost_{i,j}^{VM}(t) = \frac{\widehat{MIPS}_{VM}^{i}(t) + \widehat{MIPS}_{VM}^{j}(t)}{\widehat{MIPS}_{VM}^{i,j}(t)}$$

Equation 40

$$Cost_{i,Server}^{j}(t) = \sum_{k=1}^{N_{VM}} Place_{i,k}^{j}(t).\left[w_{i,k}^{j}(t).\left(\sum_{l=1\ \&\ l\neq k}^{N_{VM}} \frac{Cost_{k,l}^{VM}(t)}{N_{i,VM}^{j}(t)-1}\right)\right]$$

Equation 41

$$w_{i,k}^{j}(t) = \frac{\widehat{MIPS}_{VM}^{k}(t)}{\sum_{l=1}^{N_{VM}} Place_{i,l}^{j}(t).\widehat{MIPS}_{VM}^{l}(t)}$$

Equation 42

$$f_{i,w}^{j}(t) = \left(\frac{1}{Cost_{i,Server}^{j}(t)}\right).\left(\frac{\sum_{k=1}^{N_{VM}} Place_{i,k}^{j}(t).\widehat{MIPS}_{VM}^{k}(t)}{N_{i,Core}^{j}}\right)$$

Additionally, another type of correlation exists among VMs: Data Correlation. This type of correlation indicates the amount of data that will be interchanged amongst two VMs at runtime. Therefore, highly data-correlated VMs will be clustered together by the optimization algorithm. Due to its nature, the Data Correlation is taken into account in the QoS and migration section.

4.2.5 QOS AND MIGRATION CONSTRAINTS

The following constraints state that a set of VMs should be selected and migrated to datacentre j[th] from the other datacentres so that their transmission latency does not exceed a certain value as a threshold to guarantee the QoS constraint.

Equation 43

$$N_{VM}^{n,j}(t) = \sum_{k=1}^{N_{VM}} \left[\sum_{i=1}^{N_S^n} Place_{i,k}^{n}(t-1) \, . \sum_{i=1}^{N_S^j} Place_{i,k}^{j}(t) \right]; \quad n \neq j$$

Equation 44

$$N_{VM}^{j,n}(t) = \sum_{k=1}^{N_{VM}} \left[\sum_{i=1}^{N_S^j} Place_{i,k}^{j}(t-1) \, . \sum_{i=1}^{N_S^n} Place_{i,k}^{n}(t) \right]; \quad n \neq j$$

Equation 45

$$Size_t^{n,j}(t) = \sum_{k=1}^{N_{VM}} \left[\sum_{i=1}^{N_S^n} Place_{i,k}^{n}(t-1) \, . \sum_{i=1}^{N_S^j} Place_{i,k}^{j}(t) \, . Size_{VM}^{k} \right]; \quad n \neq j$$

Equation 46

$$L_t^j(t) \leq TH_{QoS}$$

Equation 47

$$DataS_t^{n,j}(t) = \sum_{l=1}^{N_{VM}} \sum_{k=1 \,\&\, k \neq l}^{N_{VM}} \left[\sum_{i=1}^{N_S^n} Place_{i,l}^{n}(t) \, . \sum_{i=1}^{N_S^j} Place_{i,k}^{j}(t) \, . DataCorr_{l,k}^{VM}(t) \right]; \quad n \neq j$$

Equation 48

$$L_d^j(t) \leq TH_{QoS} \qquad \rightarrow \qquad Data\ Correlation\ Latency\ According\ to\ Amount\ of\ Data$$

4.2.6 BATTERY CONSTRAINTS

To maximize the battery lifetime and ensure its correct functioning, each battery manufacturer defines limits on the charge/discharge currents that a specific accumulator can sustain. The following constraints permit to specify the limits in terms of power, which is more appropriate when dealing with very large battery arrays and customizable serial/parallel/mixed configurations.

The cost constraints define a cost, p, for the continuous charge/discharge utilization of the battery array and a 10x higher cost for a discontinuous utilization cycle. Generally speaking, Li-ion battery lifetime can be maximized if they are used with continuous cycles. By considering the sign of the current we can understand if the power flow is the same in two consecutive time slots or not, and the optimization algorithm can give more priority (lower price) to the continuous cycles management strategy under evaluation.

Additionally, in the case of certain constraints, or an empty battery, for example, only a change in the current flow is possible. This situation is covered by Equation 23 and Equation 24, from Section "Energy Constraints".

Equation 49 - Charge/discharge battery rates

a. $P_b^j(t) = 0 \quad or \quad th_b^{ch,min} \le P_b^j(t) \le th_b^{ch,max} < 0 \quad ; \quad if\ b = -1 : charge\ case$

b. $P_b^j(t) = 0 \quad or \quad 0 < th_b^{ds,min} \le P_b^j(t) \le th_b^{ds,min} \quad ; \quad if\ b = 1 \ : discharge\ case$

Equation 50 - Charge/Discharge cost definition

c. $C_{cc,dd}^j(t) = p \quad , if : sgn\left(I_b^j(t)\right) \cdot sgn\left(I_b^j(t-1)\right) > 0\, , \ else\ C_{cc,dd}^j(t) = 0$

d. $C_{cd,dc}^j(t) = 10p\, , \ if : sgn\left(I_b^j(t)\right) \cdot sgn\left(I_b^j(t-1)\right) < 0\, , \ else\ C_{cd,dc}^j(t) = 0$

5. ALGORITHM FOR PROBLEM RESOLUTION

The complete optimization problem for VMs dispatching among several DCs, according to the model described, can be summarized in the following list of tasks:

a. Gathering of VMs statistics for the time-slot just expired;
b. Gathering of free energy resources (predictions) and battery state (state of charge) available for the next time-slot;
c. Computation of VMs correlation, based on MIPS, during the previous time-slot.
d. Minimization of the Cost Function (minimization of the energy purchased from the Grid plus the cost of battery usage), by varying the free parameters:
    i. threshold on the cost for each cluster of VMs allocated in a server;
    ii. battery usage (c/d/off);
    iii. voltage on the CTI;
    iv. maximum migration latency;

   which affect:
    i. the corresponding frequency and power consumption of each server;

ii. the overall DC power consumption;
iii. the VMs migration latency for each DC (depends on the total size of the set);
iv. the efficiency of the converters involved;
v. the money spent for purchasing energy from the Grid.

This optimal problem allows to manage the VMs allocation at a global level, fully exploiting the correlation among VMs, to minimize the total money spent to purchase energy from the Grid, as the sum of the expenses in each DC location, and finally to obtain the best battery pack utilization, meaning to use as much as possible of it with continuous c/d cycles.

The global approach removes the need to run the VM allocation algorithm inside the DCs; we only need to run the on-line energy manager to actively adapt the real power consumption to the real free power from renewable sources and batteries.

5.1 PROPOSED ALGORITHM

The initial algorithm proposed provides the perfect solution to the problem of VM allocation. However, its high complexity prevents its practical application in real life scenarios. Next, we reformulate the problem of "energy- and cost-saving of geo-distributed data centres with correlation-aware VM placement and lifetime-aware battery banks" as an optimization problem, and show that it is NP-complete.

Theorem 1

The problem of optimizing price and power consumption in data centres with correlation-aware VM placement and lifetime-aware battery banks is NP-complete.

Proof

The placement of VMs onto the hosting DCs and then servers has direct impacts on the operating frequency and number of turned on servers optimization and the final energy and price conservation and batteries states. To achieve the final objective in the Cost Function, we could firstly enumerate all possible mapping combinations between *VMs* and *DCs* on a *n*-to-1 basis considering correlation, network latency, battery states, available free energy (renewable energy) and price. Then, for each VM cluster mapping trial (to DCs), we will have a corresponding matrix representing correlations between VM pairs, which can consequently be used for flow assignment and allocation optimization to minimize the number of servers available in each DC. At last, we can compute the power consumption, price, batteries lifetime, QoS for each trial, and find the optimal solution from the results. It is obvious that energy- and cost-aware optimization is inevitable in the above process, no matter how large or small the problem size is. Therefore, the whole optimization problem can be proved as NP-complete by restriction.

Although the enumeration-based method given in the proof is simple and direct, it cannot scale to the size of current data centres and, thus, is impossible to be used in practice. It is necessary to develop the solution and algorithms with acceptable time complexity for our purpose. We decompose the optimization problem and find that, in fact, consists of two classic NP-complete problems, namely: (1) VM grouping for each DC, (2) VM-group to server-rack mapping.

A new approach:

Next, the problem is defined formally, and a practical approach to solving it is described.

A task (Virtual Machine) is defined with the following parameters:

- VM size (2, 4, 8 GB)
- CPU correlation, of one machine with respect to another: high correlation between i and j means that the machine j will have high utilization of the CPU whenever the machine i does.
- DATA correlation, of one machine with respect to others: high correlation indicates that both machines interchange a lot of data at runtime.
- Time a, arrival time, given in time slots
- Time f, finalization time, given in time slots

Based on this properties, two **cost functions** are defined to calculate the "forces of attraction and repulsion":

- Repulsion, based on the CPU correlation -> (0, 1]
- Attraction, based on the data correlation -> [-1, 0]

Algorithm:

All the virtual machines are represented as dots in a 2D plane. Between any pair of points, there are forces of attraction and repulsion (given by the previous equations). Initially, all the dots are in the coordinate (0,0). Then, one by one, the forces are calculated and the points remapped in the 2d-plane, increasing or reducing their distances to balance the forces. This process is iterated until the forces are balanced (not completely neutralized, but smaller than an epsilon).

Each datacentre has a capacity (in Joules) according to the battery and renewable energy available in the current time slot. We cluster the machines in datacentres according to the available energy, and their energy consumption.

We utilize a modified version of the K-means algorithm [14] to cluster VMs with respect to each cluster capacity (the battery and renewable energy available, and the power consumed during the previous time slot), and distance between two VMs obtained from repulsion and attraction phase in the 2D plane. In this step, we do not consider network latency as a QoS criterion to obtain the optimal solution in the 2D plane. Then, we should revise the k-means output due to meet the QoS constraint as follow.

The clustering plan obtained in the previous step is revised taking into the account the latency. According to K-MEANS output, for each Datacentre (cluster), we take into account two queues are prepared: outgoing and incoming. The first one contains the candidates to be migrated outside, to another datacentre, while the second one contains the candidates to be migrated to this datacentre. First, we select one VM from the incoming queue of one cluster. If the latency and capacity allow us to migrate this VM we do it otherwise we select another VM from the queue. If there is no VM to accept we select one VM from incoming queue of another cluster. After accepting the VM, we select another VM from outgoing queue of current cluster has maximum distance (force) to accepted VM. If we could migrate this VM, we go to destination cluster and select from its outgoing queue and repeat this. Otherwise, if we could not migrate the selected VM from outgoing queue, we select the second VM has the maximum correlation with the accepted VM in the cluster. We repeat algorithm until violating the latency or datacentres capacity or there is no action to do. Unallocated VMs will be stayed in its previous position.

Each Virtual machine must be allocated to a server, and the optimal Voltage/Frequency operating for the servers should be computed. In this step, we consider only CPU correlation to allocate VMs to the servers. Therefore, we utilize the best state-of-the-art algorithm [15] in this phase.

NOTES:

For the initial implementation, we consider that the new virtual machines have no latency (i.e.: they are available to be spawned in any datacentre).

6. EXPERIMENTS

In order to see the analytical model at work, and evaluate the effectiveness and applicability of the proposed framework and algorithm to larger scale problems, a real life experiment is included in this section.

6.1 EXPERIMENTAL SETUP

We consider a GreenDataCentre made of 3 Locations in Europe (Lisbon, Zurich and Helsinki), on those depends the distance (for network model), the shift in the time of day, the temperature/irradiance sequences, the size of the Battery, the size of the Photovoltaic Plant and the energy price. The locations considered are different from Barcelona, Zurich and Amsterdam, used for the renewable energy maps in WP1 because it was necessary to have a bigger distance between the locations in order to have a more realistic simulation on the network model.

Data centres:

- 3 Locations (size of the Photovoltaic Plant, size of the Battery, time shift for each time zone):
  - Lisbon: PV 5kWp, Battery 19.2kWh, shift 0,35 servers
  - Zurich: PV 3.7kWp, Battery 14.4KWh, shift 1 hour (wrt Lisbon),25 servers
  - Helsinki: PV 10kWp, Battery 9.6kWh, shift 2 hours (wrt Lisbon),15 servers
- Distances
  - Lisbon -> Zurich: 2000 km
  - Lisbon -> Helsinki: 4000 km
  - Zurich -> Helsinki: 2500 km

Workload (VMs):

- Two weeks simulation horizon.
- Workload traces obtained from a real DC setup and real irradiance and temperature profiles. To simulate the DC workload and energy demand we sampled the CPU utilization of a real DC setup every 5 min. for one day, then we duplicated the samples up to 14 days. Finally, to generate different samples for each day, we synthesized fine-grained samples per 5 sec. with a lognormal random number generator [5], whose mean is the same as the collected value for the corresponding 5-minute sample rate.
- Maximum allowed VM in the network of DC: 400.
- Data correlation between two VMs is randomly generated by uniform distribution and maximum 10 MB.
- Arrival and finalization time of each VM, given in time slots, are randomly generated by uniform distribution.

Power:

- Three Homogeneous Urban Datacentres in Europe, irradiance and temperature profiles and double price scenario (regulated electricity market) have been taken into account.
- The urban Datacentres consist of medium sized facilities with two components: (i) computing power consumption (IT equipment) and (ii) computer room air conditioning (CRAC) power consumption as the cooling unit.
- Machines: Intel Xeon E5410 server configuration which consists of 8 cores and two frequency levels (2.0GHz and 2.3GHz), and used the power model proposed in [6].
- Efficiency of converters constant and equal to 92%.
- Battery: On-line battery management only (- Only Lithium-Ion Battery Bank managed.

Network:

- High-speed network link between them (optical fiber, 100Gb/s, full-duplex, WDM…) with very low tx/rx transmission time considering distance between the DCs (included in the computation).
- 10 Gb/s full-duplex intranet speed inside the DC, between server. VMs size in the range 2, 4, 8 GB randomly generated according to the following distribution: 60%, 30%, 10%.
  Randomly generated network errors that slow down the migration between datacentres and affect the QoS. We modelled the random noise as a stochastic process with a fixed discrete probability distribution function, every second we experience a bit-error-rate that is chosen randomly following this distribution: 54% probability of 10-6, 20% prob. of 10-5, 15% of 10-4, 10% of 10-3 and 1% of 10-2.

Other considerations:

Dispatcher Scheme Invoked every 1 Hour -> VM Allocation Scheme (in each DC) Invoked every 1 Hour.

The Quality of Service (QoS) is defined as 98% of dispatching time period. This is the maximum time consumed to migrate VMs from other DCs to a certain DC through the network; until all the batch of VMs are ready to allocate and run on the servers in all datacentres.

Energy management and optimization at global level (dispatching time) based on forecasted irradiance profiles.

Energy management and optimization at local level (online part between two time slots) based on real irradiance profiles.

6.2 RESULTS

The setups, presented in the previous section, have been formulated following the indications and specifications of the industrial partners of the GreenDataNet project; in particular, Credit Suisse has indicated the ranges of sizes and number of servers, as well as setup of the complete IT setup, while Eaton has indicated the specifications and models for the PDUs and related equipment. For the PV and storage system and their interaction with the smart grid, we replicated the setup published in [3] by CEA, as the characterization of Nissan batteries was not available at the moment of the preparation of the simulation tool . The details for the complete simulation, a total of 336 hours (or 14 days), required 60 minutes approximately to complete in our Intel Xeon X5650 -based server (@2.67GHz).

Figure 6.1 depicts the normalized SoC of the batteries; therefore, a 1 indicates that the battery is fully charged. The degradation of the battery depends on the charge/discharge cycles, as well as the current flows (see the battery equations). The graphical evolution of the SoC is depicted in Figure 6.2.

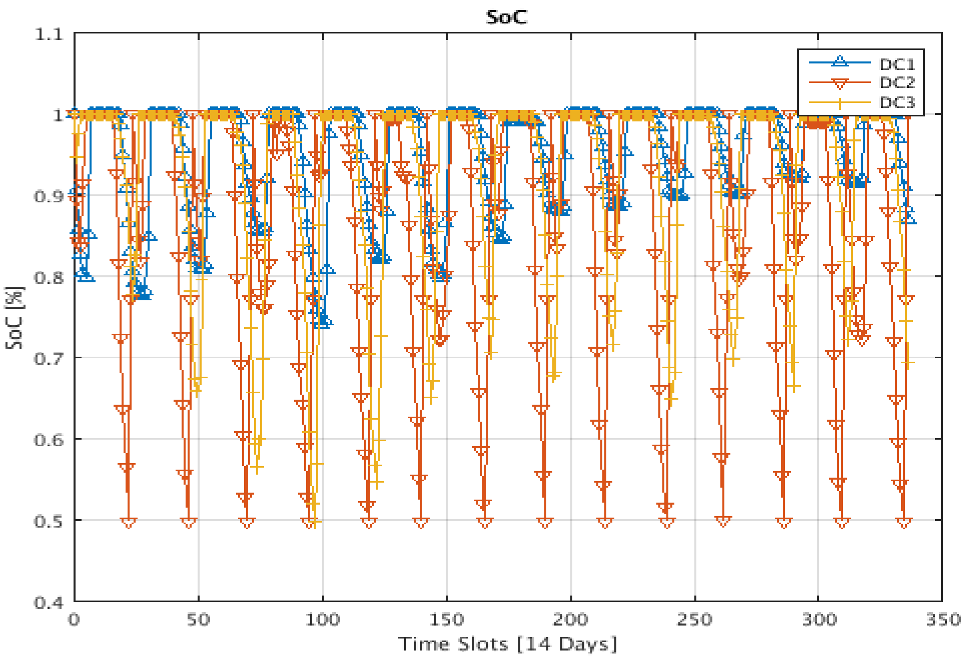

Figure 6.1 – State-of-Charge of the batteries, for the 3 DCs.

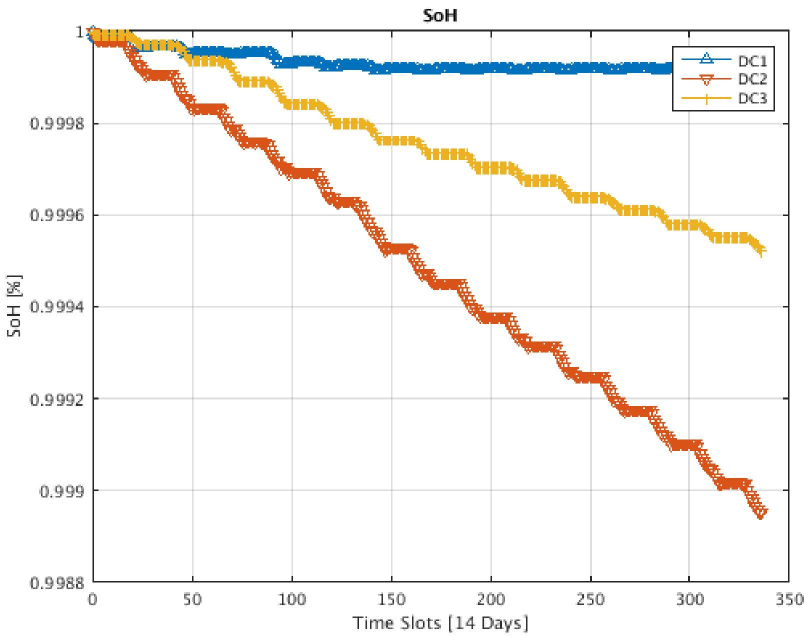

Figure 6.2 – State-of-Health of the batteries, for the 3 DCs.

The following three figures (Figure 6.3, Figure 6.4 and Figure 6.5) study the power profile inside the datacentres (1, 2 and 3, respectively). For each DC, it is represented the power drawn from the battery (note that negative values indicate that the battery is being charged, whereas positive values indicate that the DC is being powered by the battery), the power drawn from the GRID, and the power generated in the PV (divided into used and unused power).

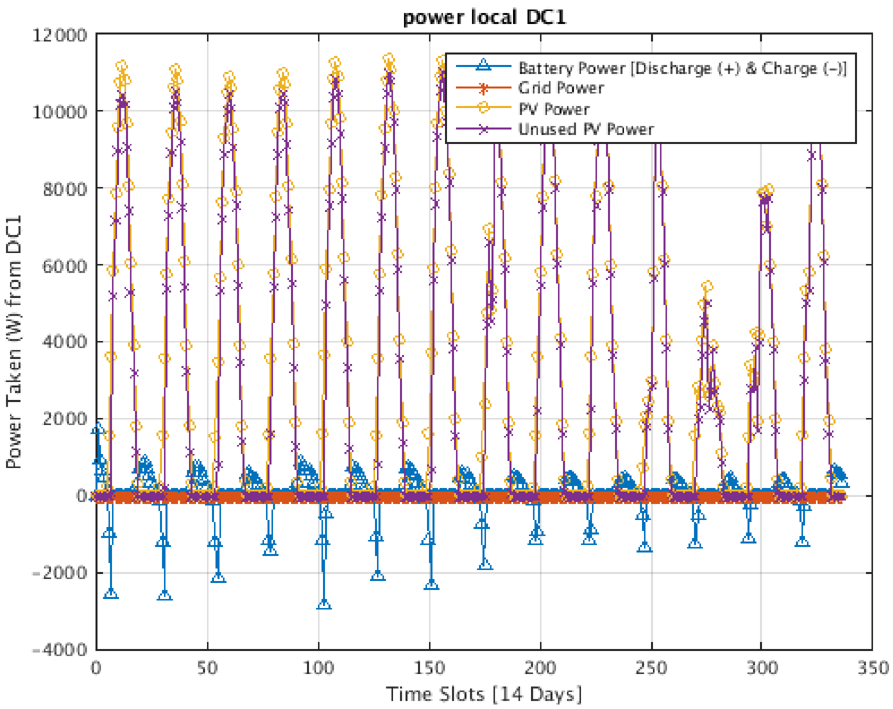

Figure 6.3 – Power consumed by DC1. Detailed for the different sources.

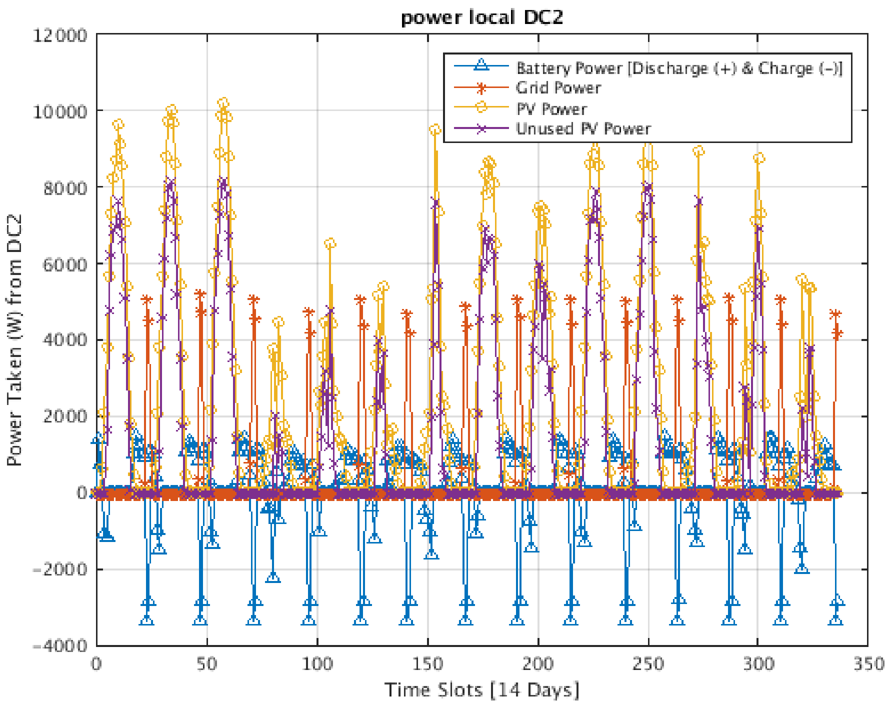

Figure 6.4 - Power consumed by DC2. Detailed for the different sources.

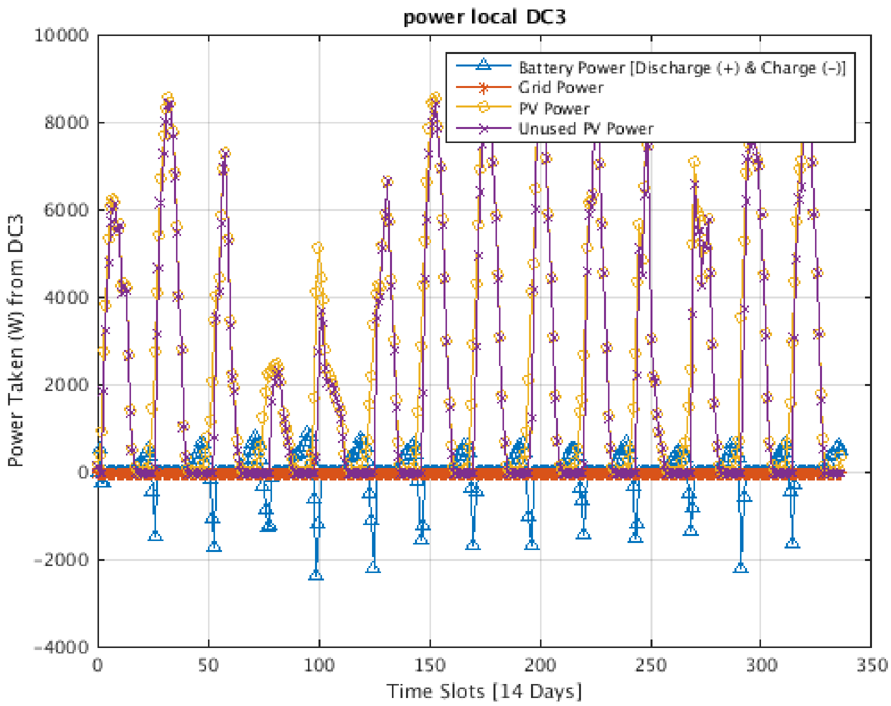

Figure 6.5 - Power consumed by DC3. Detailed for the different sources.

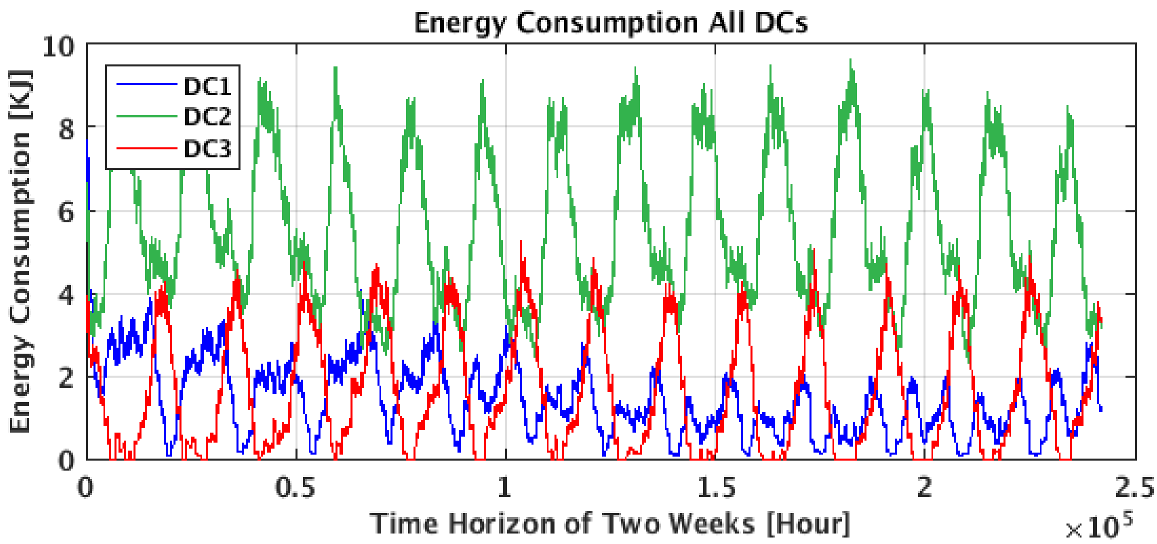

Figure 6.6 –Consumption of energy in the DCs.

The next two figures represent the energy consumption of the whole datacentre: cooling + IT (Figure 6.6), that can be derived from the previous figures, and the cost of the energy consumed (i.e. demanded from the GRID) by each DC .The total energy consumed in the green datacentre is 2.08 GJ. Table 6.1 summarizes the contribution from each DC location.

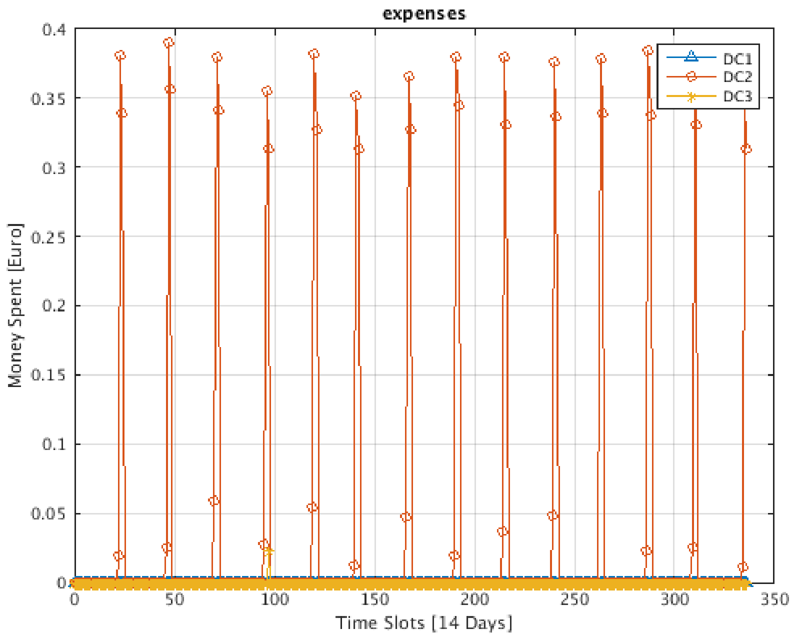

Figure 6.7 – Cost of the energy consumed in each DC.

	Lisbon DC (DC1)	Zurich DC (DC2)	Helsinki DC (DC3)
Total Energy (GJ)	0.35	1.35	0.38

Table 6.1 – Total Energy Consumption per DC.

We have considered two factors to assess the Quality of Service (QoS). The first one is related to the dispatching algorithm. It is defined as the maximum ratio of time consumed to migrate the VMs from other DCs to a certain DC through the network (until the batch of VMs are ready to allocate and run on its servers), to the dispatching time period during the two weeks. The graph is depicted in Figure 6.8, where it can be observed that the QoS never decreases below 98%, the given constraint (the worse-case).

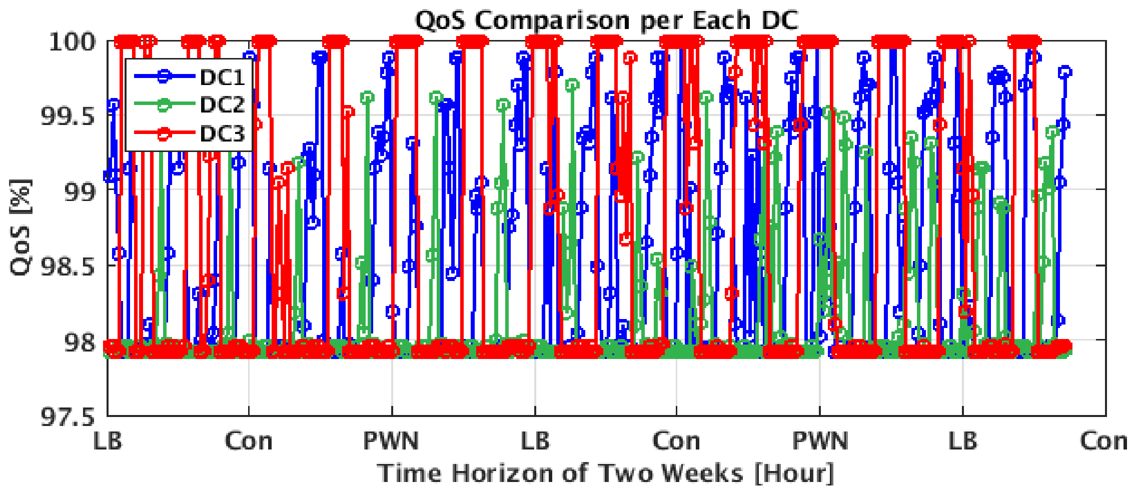

Figure 6.8 – QoS for the DCs

The second factor to measure the QoS (Local part; inside of each DC), is the number of violations, defined as the maximum per-period ratio of the number of over-utilized servers (i.e., when the aggregated utilization among co-located VMs is beyond the CPU capacity of a corresponding server) to the total number of servers in DCs between two VM allocations in each DC. It can be observed in Figure 6.9. The maximum number of violations per DC (for a given slot) is given in Table 6.2. The global average is 2.65%.

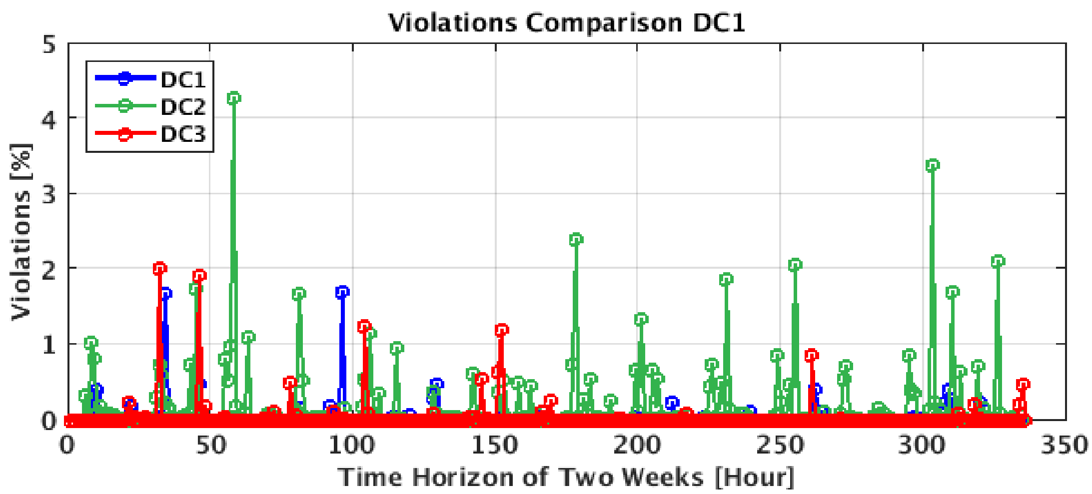

Figure 6.9 – Number of violations per period.

	Lisbon DC (DC1)	Zurich DC (DC2)	Helsinki DC (DC3)
Max #Violations (%)	1.7	4.2	2.05

Table 6.2 – Maximum number of violations.

The total number of migrations in the system is 11600 migrations. Figure 6.10 shows the number of migrations per time slot, and Figure 6.11 depicts the ratio of outgoing vs. incoming migrations.

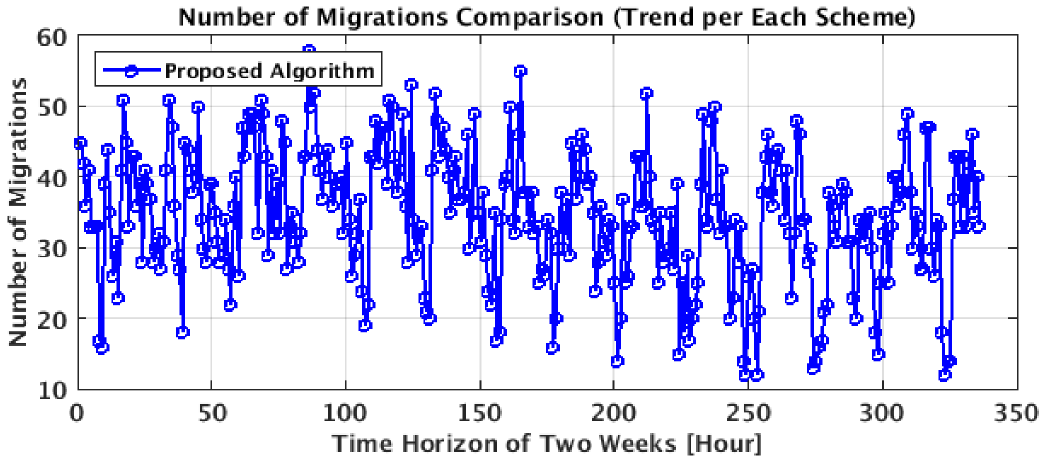

Figure 6.10 – Number of migrations per time slot.

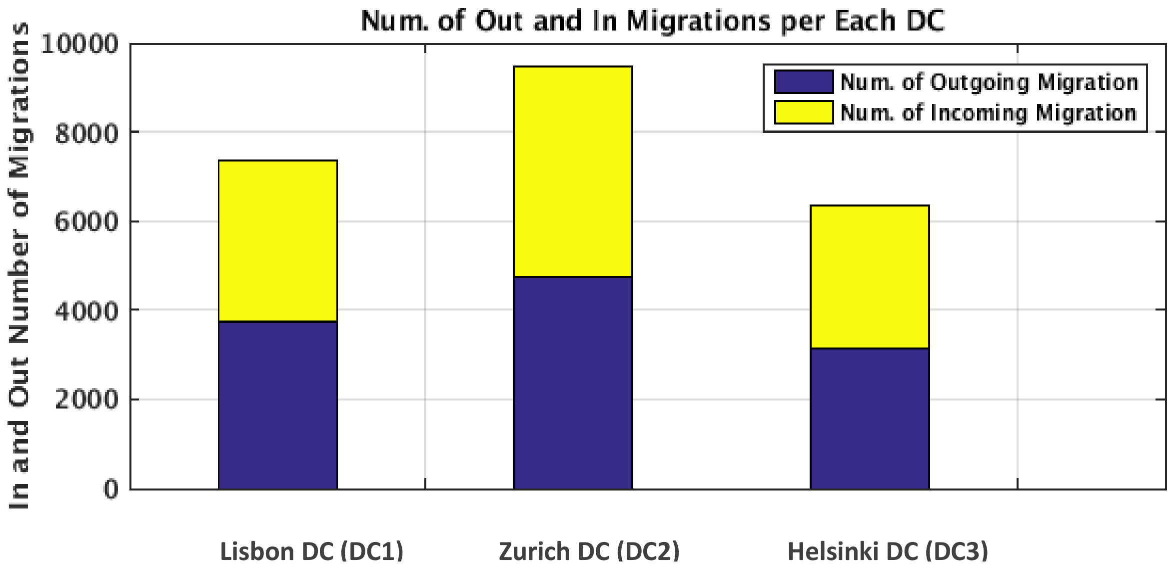

Figure 6.11 – Percentage of outgoing vs incoming migrations, per DC.

7. CONCLUSIONS

In this deliverable, we have presented the analytical model used for green datacentres: detailing the whole DC structure, the system constraints and, finally, an algorithm to optimize the energy consumption altogether. The last section includes also the simulation of a real case scenario using a real-life configuration.

This is the analytical model on top of which the tool presented in deliverable D3.2, the electricity consumption forecasting tool, is based. It optimizes the energy consumption in the green DC by choosing an adequate allocation of tasks (VMs) to servers.

The experimental results, Section 6, are not only an example of how the tool works and the type of studies that we can conduct with it but, at the same time, these results also prove that the proposed green datacentre is not optimized: The overall setup suffers from a serious underutilization of the grid; this is inferred from figures 6.1 to 6.7 presented in the experiments section, where DC2 is the only one that demands important amount of current from the grid (c.f. Figure 6.6 and Figure 6.7). This result will also be taken into account in the dimensioning of the storage of the demonstrator, as the size of the batteries proposed in deliverable 1.6 can be considered as a first iteration in order to get to the optimized storage size.

The other two datacentres run, almost exclusively, off the renewable energy (battery + PV). Apart from this aspect, the scheduler proposes an allocation scheme that meets the given constraints and tries to minimize the overall power consumption.

All the values (c.f. Section Experimental Setup) utilized in our current configuration come from documents like datasheets or previous deliverables that have been reviewed/proposed by our industrial partners and taken from real scenarios. However, when put together, they result in an inefficient architecture, requiring the resizing/substitution of some of the elements. This is one of the goals of GreenDataNet: to demonstrate that an accurate model is extremely convenient and necessary to tune the configuration of a real green datacentre before actually building it. Future deliverables, like 3.6 and 3.11, for instance, will detail the whole process of profiling the different components of a datacentre and selecting the right ones according to the given constraints and simulation results.

In the context of GreenDataNet, the model described here interacts with the different forecasting tools implemented in WP2 and WP3, like the ones to estimate the PV production (D3.4) and the IT energy (D2.6). Additionally, the complete specification of this model is the first step towards the creation of higher level tools that will further optimize the operation of green datacentres, such as those investigated in deliverables D2.3 and D2.4: Server Multi-level SW Management Specification and Implementation and Racks Multi-Objective Energy Management Specification and Implementation, respectively.