Increasing Data Centre Renewable Power Share via Intelligent Smart City Power Control

Florian Niedermeier
University of Passau
Computer Networks and Computer Communications
Innstrasse 43
Passau, Germany
florian.niedermeier@uni-passau.de

Wolfgang Duschl
University of Passau
Computer Networks and Computer Communications
Innstrasse 43
Passau, Germany
wolfgang.duschl@uni-passau.de

Torben Möller
Software Engineering Group
University of Mannheim
B6, 26
Mannheim, Germany
moeller@informatik.uni-mannheim.de

Hermann de Meer
University of Passau
Computer Networks and Computer Communications
Innstrasse 43
Passau, Germany
hermann.demeer@uni-passau.de

ABSTRACT

Urbanization has been an increasing trend that has lead to higher population densities in cities worldwide. Providing these large amounts of citizens with a high quality of living is one of the main goals of future smart cities. A high quality of living requires the smart city to offer several services, many of them supported by a sophisticated, power-intensive IT infrastructure. We present a cooperation scheme between smart city and data centres that allows for an effective use of power flexibilities inherently available in data centres. By adapting the power demand of data centres to the availability of renewable energy, smart city goals like a low carbon emission of its IT infrastructure become achievable. To this end, a demand side management scheme under orchestration of a central control system is proposed. The concrete guidelines on power use for data centres are calculated by a component named "Ideal Power Planner", based on smart city goals and renewable power availability forecasts. The effectiveness of the developed approach has been validated in three testbeds.

1. INTRODUCTION

To provide a more sustainable and eco-friendly energy supply, the integration of renewable energy into the power grid has been a major factor in the recent years. Especially for smart cities, which strive to provide a high quality of living to its citizens, a green power grid is a key goal. However, power generation based on renewable sources may be volatile, especially for wind and solar. In turn, the fraction of renewably generated power inside the electricity grid is varying over time. To gain a maximum benefit from renewable generation, power demand should coincide with times of maximum renewable availability in the power grid. Simultaneously, situations of supply surplus or scarcity should be avoided to ensure grid stability and avoid high costs for peak load coverage. Prognosis of weather conditions can help, but this is not always as precise as necessary, especially if weather conditions fluctuate massively during the day. One further approach is using batteries to store electrical energy to even fluctuations in the electricity grid and provide a need oriented return of energy. However, this technology is mostly not considered beneficial enough from an economic point of view to be used at large scale in deployments yet [4]. Considering the problem that volatile energy resources cannot be controlled directly without wasting resources like traditional fossil fuel based power plants, there is a need for an alternative approach to handle this situation.

We propose a demand side management scheme under orchestration of a central control system which is able to perform a higher level optimization compared to autonomous demand side management approaches. The proposed system is based on a two-stage calculation of desired power values for the participating consumers. In the first stage, the high level goals of the smart city are divided into concrete objectives. In the second stage, these objectives serve as an input to the Ideal Power Planner, which derives a time series of power values to be followed by the demand side participants to the system. In the following, it is assumed that the demand side participants to the system are data centres with a certain amount of flexibility regarding their power demand. Additionally, we assume that forecasts regarding the availability of renewable energy are either readily available or easily derivable using information shared by Energy Related Data Suppliers (ERDS). This work contributes three major points:

1. A system architecture description of a cooperation scheme between smart city and data centres
2. The Ideal Power Planner, a component that translates high level goals of the smart city into concrete power values to be followed by a data centre
3. The communication scheme to share smart city goals and objectives with data centres

The remainder of this paper is structured as follows: Section 2 gives an overview of the proposed system. Section 3 introduces the Ideal Power Planner, a central component in the system. Section 4 briefly discusses the influence of metrics. In Section 5, the communication between smart city and data centres is explained. Section 6 presents first results gathered at three trial sites. After positioning our work in relation to previously conducted research in Section 7, the paper is concluded in Section 8.

2. SMART ARCHITECTURE

Through the high concentration of users on the demand side in cities, the energy consumption fluctuates significantly during a day. On the other hand it is important to consider that photovoltaic (PV) and wind power are not necessarily available over a whole day. At least for PV power, it is evident that it is not available during the night. To deal with this situation, it would be best to use renewable power peaks to satisfy the demand side. To cope with volatilities on both demand and generation side, we propose an intelligent control mechanism for demand side management based on the cooperation of smart city and large power consumers [14]. To this end, suitable grid participants are needed, which can provide a certain level of adaptiveness. Adaptiveness means in this context that the particular power demand can be shifted to another point in time, if necessary. Although the proposed scheme can be applied to different types of consumers, this work focuses on data centres. They seem promising candidates to provide the previously discussed adaptiveness as automation frameworks are usually already in place and several mechanisms exist to gracefully scale power demand at the expense of service performance (e.g., dynamic voltage and frequency scaling (DVFS)).

To integrate DCs in a suitable way, the following factors should be fulfilled to the highest possible extent:

- A sufficient amount of non-time critical jobs, whose execution schedule can be adjusted on demand
- Available power saving mechanisms (e.g., p-states) to exploit differences in machine load patterns
- Localized in (or close to) a city and connected to the same power grid or micro grid

2.1 Suitable DC types

The capability of data centres to adapt to a changing amount of available renewably generated power depends on several factors, however for the proposed system the most important are those impacting flexibility [13]. From the perspective of business models two criteria are taken as an example: First, the so-called "Outsourcing Factor". It depends on the business model of a data centre and is defined as the amount of operation control the DC may exert on its hosted services. This also includes control possibilities to increase flexibility. As the proposed system relies on DC flexibilities, DCs with high Outsourcing Factors should be preferred. Second, the suitability depends on the flexibility of service level agreements (SLAs). SLAs determine constraints of a data centre regarding, e.g., response time of services or generation rate of documents. If a SLA allows for a more flexible approach, it also improves possibilities for adaptation. Figure 1 shows the relationship between SLA flexibility, Outsourcing Factor and DC suitability (color of the circles). Data centres offering software as a service are very flexible, for example, their SLAs guarantee high flexibility and their high Outsourcing Factor allows for high control through the data centre administration. Best suitable data centres offer SaaS and are bound to a flexible SLA. From

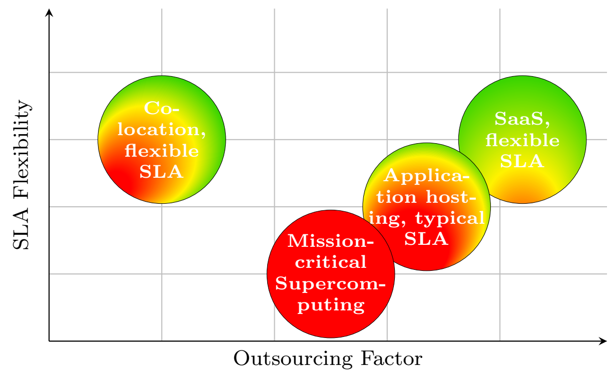

Figure 1: SLA Flexibility (based on [13])

a technical perspective, there are different levels at which data centers may offer adaptation. Depending on where the adaptation is performed, adaptation time, amount of power adaption, and other characteristics of the created flexibility may change drastically. This also has a significant impact on the response time that can be provided and in turn on the adaptation requests which can be fulfilled. On a qualitative level, adaptations which are based on "silicon only", i.e. no mechanical operations are involved, and are performable at runtime qualify for very rapid execution (< 1s). Examples include DVFS, application internal changes or virtual machine resource capping. On the other hand, there are flexibilities which cannot be provided instantaneously. Amongst others, these are virtual machine live migration, switching machines on/off or any other change that is usually only performed after manual review.

2.2 System Interactions

Many different stakeholders are interacting in a smart city environment. These include bulk generation like nuclear or coal power and so called "Distributed Energy Resources" [3] like solar and wind power generation in a power range of 100kW to 10MW. Further, there are grid operators which take care of transporting and distributing electricity to the consumers, an electrical energy market and end users. The central role in a future smart city environment is a management authority which deals with different grid situations. It is called "Energy Management Authority of the Smart City" (EMA-SC). Tasks of an EMA-SC include, amongst others, monitoring of all available producers and consumers as well as their regulation and balancing by creating power plans for the demand side, especially for DCs. Additionally, an EMA-SC has a management role if problems arise during the regulation process. Besides the provision of the required energy to DCs, it is also important for an EMA-SC to have information on directly or locally installed renewable generation, which can be used for own consumption. Additionally, the expected energy mix in the grid or micro grid and the guidelines of the EMA-SC are taken into account. In the course of its task of calculating power plans (a kind of an energy budget) for DCs in the smart city, an EMA-SC makes use of a component named Ideal Power Planner (cf. Section 3), which bases its calculations on metered and forecasted values of energy generation and constraints imposed by an EMA-SC. All this information is used to generate a suitable power plan, which should be followed as closely as possible by the respective DC by splitting and shifting its tasks accordingly. The result is a maximum usage of renewable energy under the constraints of the EMA-SC. The goals for creating power plans are:

- Maximum usage of renewable (especially volatile) energy
- Provide enough energy to DCs to maintain an acceptable quality of service
- Support the implementation of smart city goals by taking them into account during power plan calculation

It is still important to mention that the generation of these power plans is improved by incorporating a city-wide view, which includes all large generation and demand participants. This allows for the creation of a more sustainable power infrastructure under the conditions of optimal use of renewables and provision of enough energy for the demand side. In the next section, the calculation of suitable power budgets by the Ideal Power Planner will be explained in more detail.

3. THE IDEAL POWER PLANNER

For a data centre to be able to comply with smart city regulations, the high-level goals that drive a smart city have to be translated into concrete technical terms that can be implemented. To this end, a software component called Ideal Power Planner (IPP) translates the constraints from smart city side and forecasts on renewable energy availability into bounds on power consumption that should be followed by a data centre.

3.1 Tasks and attributes of the IPP

The IPP is a stateless component that provides to a DC an upper bound regarding its power demand. The upper bound in this case means the maximum power amount that may be used while still complying with the constraints set by an EMA-SC.

- Input values: Basically, the IPP takes two kinds of input data: On the one hand, it needs forecasts on future renewable power availability. This may be provided directly by the distribution system operator (DSO) or computed independently by monitoring the respective environmental parameters. On the other hand, the constraints set by an EMA-SC on power, energy, and energy properties have to be communicated.
- Output values: The output produced by the IPP is an upper bound on power use for a certain time into the future (e.g., 24 hours) in a resolution of e.g., 15 minutes. The concrete values may vary and depend mostly on the prediction accuracy and computing resources available. A DC participating in the system should try to not exceed this limits to be in line with the constraints set by the respective EMA-SC. Of course, a DC has to fulfill its SLAs at the same time. Rewards and penalties for keeping/breaking the power plan may be defined in a contract between an EMA-SC and a DC.
- Internal working process: To calculate power bounds, the IPP has to translate constraints set by an EMA-SC to concrete power values. This can be quite easy (e.g., for a simple constraint on max. power use) or complex (e.g., a constraint on min. % of renewable energy and multiple energy sources). Internally, the IPP can solve these tasks by employing, e.g., a constraint solving engine. For a more detailed description on how objectives are set, please refer to Section 5.1.

3.2 IPP location

The IPP may be located at two different points in the envisioned system: Either centrally on smart city side or locally at each participating DC. The choice of location has several implications on the system.

3.2.1 EMA centrally on smart city side

In a future smart city, calculating a power plan may be performed centrally by an EMA-SC. As an EMA-SC has more detailed information on the power grid state, it may be able to perform a higher level optimization. All data related to power generation and demand are gathered, processed and orchestrated at EMA-SC side. It collects all forecasts and after consolidation it may calculate power budgets that can be distributed among data centres in order to make use of their flexibility. In this case, DCs are relieved completely from having to calculate power values. DCs directly receive power budgets per time slot from an EMA-SC. Figure 2 shows a scenario with central power planning.

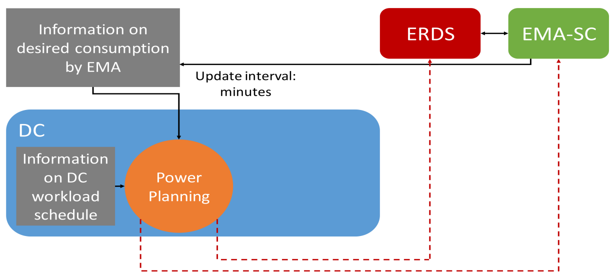

Figure 2: The envisioned future integration with smart cities

3.2.2 *EMA locally on DC side*

However, with current smart cities, it is highly likely that there is no way to directly give restrictions on power use to a DC. The future scenario described above is not currently in place. To be applicable even in absence of a strong EMA-SC and to support legacy systems in the future, the system also offers the possibility to run without a central power planning service. In this case, the system gathers forecasts of renewable energy availability and/or environmental data that is used to create needed forecasts locally. Power is then adapted using power availability information and goals set by the respective EMA-SC. The minimum required functionality from EMA side is therefore limited to setting at least one objective for each DC. Figure 3 shows the architecture with locally calculated power plans.

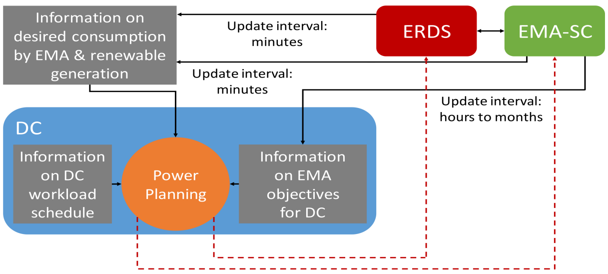

Figure 3: An architecture not depending on power planning capabilities on smart city side

4. METRICS

Metrics are a very important utility to measure and evaluate the results of an adaption process. As an essential requirement, a single metric should be as generically applicable as possible. In case of DCs, it is obvious that the measurement of utility or adaptiveness is specific to certain classes of applications offering different services. The quality of service (QoS) of a web server, for example, may be measured by the number of served websites or response time. In contrast, a backup system has to perform a lot of I/O operations and handle network traffic. If a specific environment is considered, it is possible to measure, e.g., power demand vs. received utility of a service. To find the correct constraints in CPU time, I/O etc. without creating a single-resource bottleneck for the service is however a complex task which required further investigation. Another useful application of metrics would be to quantify the overall flexibility of a DC while still fulfilling its SLAs.

5. COMMUNICATION

An EMA-SC and a data centre communicate with each other in two ways. First, an EMA-SC sets objectives a data centre should accomplish. A data centre tries to shift its workload to accomplish these objectives. If this is not possible, it will escalate the process by sending a message to the EMA-SC so the organization can try to find further solutions.

5.1 Setting Objectives

The objectives communicated to a DC are derived from high level goals of the respective smart city. A goal may be very abstract, like a certain percentage of renewable energy used.

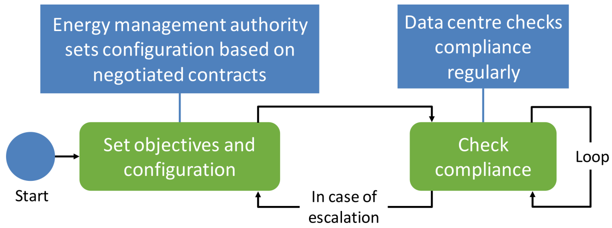

Figure 4: State diagram of EMA-SC and DC interaction (based on [6])

This goal is subdivided into objectives, which are more concrete and easier to implement. These objectives are further specified into targets. The relationship of goals, objectives and targets is shown in Figure 5.

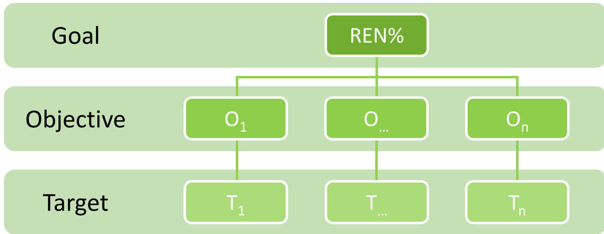

Figure 5: Goals, objectives and targets (based on [6])

In order to set objectives, an EMA-SC negotiates with the owner of a data centre and notes agreed on objectives in a contract. Objectives are set with respect to the local context (e.g., local renewable supply). To enable technical enforcement of objectives, they are modeled by three different types of constraints:

- Energy property constraints, like the renewable energy ratio and CO_2 emissions can be used for long term goals. E.g., to reach at least a renewable energy percentage of 80%.
- Energy constraints are used to limit the energy used in a certain time frame while retaining a certain flexibility regarding the exact time of power use.
- Power constraints work in a similar way to energy constraints. They are however limited to power, i.e. instantaneous demand.

These objectives are fixed for a given time frame. It is also possible to define recurrence of objectives, in case they are expected to happen again in certain intervals. E.g., weekends can be handled similarly every week by using an interval of one week.

5.2 Deriving Objectives

Objectives are based on metrics that are expressive and can be applied to every data centre. Energy-related information like terms fixed in already existing energy supply contracts or energy bill details can be taken into account. This information can be used to generate a prediction of future energy needs. From data centre side the workload based on size or customer amount can be used for predictions.

It should be noted that all this data has to be used in the local energy context and time-dependent consumption in mind. For example, workload can depend on day-night or season cycles.

5.3 Escalation

If one or more of the objectives agreed on are not possible to accomplish, the respective data centre starts an escalation process. A determined person is informed via a negotiated communication channel on important information the data centre can provide to determine the affected objective, time frame and possible solutions.

An EMA-SC may react in multiple ways: Usually one of the solutions applied to the data centre is, e.g., migrating services to other data centres participating in the system. The objectives, SLAs or the contract could be changed to avoid violations. If none of these steps can be applied the EMA-SC is left with two possibilities: Either incur a penalty according to the contract or allow the data centre to break the contract once.

6. TRIALS

The proposed system was evaluated in the course of the EU project "DC4Cities" in three different data centers. The trial sites have different core workloads and therefore exhibit different load patterns. In the following, more details on the trials will be given. Afterwards, first results are discussed.

6.1 Test setup

Trial site 1 was built upon two independent data centres. In these data centres several services like web crawling or backup are hosted. The focus was put on a compute and I/O intensive video transcoding batch job, to which privileged users can upload videos, which are converted to save storage space. Since the SLA requires a certain amount of videos to be transcoded every 24h, it is possible to shift the jobs intra-day to times of high renewable energy availability.
The trials at trial site 2 are executed in collaboration with a local health agency. The local health agency needs a report generation application, which generates high load on the systems. This report generation can be shifted to time frames with high renewable energy availability. Shifting is possible, since there is no need for on-demand generation. The generation can be scheduled and executed ahead.
Trial site 3 hosts a web application e-learning platform in its trial. This platform is available 24/7. During the trial the number of backend servers available to the web application is adapted to the availability of renewable energy while keeping SLA requirements.

6.2 Current results

The key performance indicator of success of the trials is the percentage of renewable energy of the total amount of consumed energy.

$$RenPercent = \frac{\sum_{i=1}^{N}(E_{DCgrid_i} \cdot \frac{E_{ren_i}}{E_{sys_i}} + E_{DCself-cons_i})}{\sum_{i=1}^{N} E_{DC_i}} \quad (1)$$

The calculation of RenPercent is explained in equation 1. For each time period i the DC energy consumption from grid is multiplied with the percentage of renewable energy in the whole energy system plus the energy produced and consumed locally. The result is divided by the total DC energy consumption. A more detailed description of the trials, formula and results can be found in [9].
Trial site 1 shows an increase from 40.09% to 41.78%. The biggest increase on one day of the trial was from 38.75% to 47.76%, which is an improvement of 23.25%. The second data centre of trial site 1 has shown an improvement of 10.11% (from 41.74% to 45.96%). This was reached by tuning the start and stop times of services and batch jobs.
In trial site 2 the data centre was tuned by using parallelism and adjusting start and stop times of the jobs. The difference is that in this trial the SLAs define a stringent timing. In trial site 2 two configurations were tested. The first one did not show any improvement due to the high idle power of the legacy servers used in this trial. The second configuration was set to shut down idle servers and increased the percentage of renewable energy from 43.13% up to 57.88%. That is an improvement of 34.20%.
Trial site 3 has shown an average improvement of 4.63% in the first run (from 57.45% to 60.11%) and 4.47% in the second run (from 62.40% to 65.19%). The improvements of trial sites 1 and 3 are significantly lower compared to those of trial site 2 as their infrastructure features more up-to-date hardware and was already tuned for energy efficiency before the trials.

7. RELATED WORK

This work has connections to many research fields, amongst others energy efficiency, power flexibility and smart city control. In the following, previous work in the most significant areas is discussed.

Extensive research has been performed in the area of data centre energy efficiency. In this field, both software and hardware approaches are discussed. The work [2] aims to improve DC energy efficiency by measures like improving the cooling system efficiency, usage of variable frequency drivers, or using direct current power systems. Also, management techniques like virtualization and consolidation of servers are considered. [7] researches the possibility of using power management policies (consolidating, P-states, etc.) to design a power management plan that matches the supply of power with the demand for power in DCs. In [5] the SESAMES architecture is presented, aiming to reduce energy demand while reducing both cost and environmental impact in future supercomputing (HPC) environments.

The work of [12] introduces a green wide-area testbed. The focus is on reducing greenhouse gas emissions. In [1] the authors propose an approach to negotiate flexibility between energy suppliers and energy consumers. By an increased flexibility the supplier can react easier to changes in renewable energy supply. DC4Cities uses the gained knowledge of All4Green for contracts between smart city and data centre owners.

Regarding the adaptation of workload, [8] propose a combination of electrical storage and workload adaptation to power DCs by renewable energy. The work also considers the cost implications of peak power management, storing energy on the grid, and the ability to delay the MapReduce jobs. In [10] the GreenCassandra system is introduced. It uses a combination of prediction (workload and energy availability) and modeling (performance and power) to improve the use of self-generated solar power while keeping SLAs in distributed structured storage systems.

Other work also takes SLAs between DC and its customers into account. In [11] the authors state that the environmental impact of a service offered to a user should have an influence on its cost. Therefore, a collaboration between DSO, DC and its customers is suggested, called Green SLAs. The work of [15] investigates how to dynamically distribute service requests among data centres in different geographical locations, based on the local weather conditions, to maximize the use of renewable energy. The authors introduce the middleware system "GreenWare", that conducts dynamic request dispatching to maximize the percentage of renewable energy used to power a network of distributed DCs.

Our work differs from all stated approaches in the respect that it specifically aims at integrating DCs better into future smart cities. Additionally, in contrast to other approaches, we consider an optimization that builds on a cooperation of smart city and DCs, which we believe will increase the effectiveness of our work and enable a higher level of optimization.

8. CONCLUSION AND FUTURE WORK

In this work we presented a novel cooperation scheme between smart city and data centres aimed at achieving a better integration of data centers into future urban environments. At its core, the proposed system relies on predictions of future renewable energy availability and guidance by a smart city energy management authority. The system may operate both in a tight coupling with a smart city or, if a strong EMA-SC is not available, in a more autonomous mode relying only on objectives given by an EMA-SC. The effectiveness of the approach has been validated in a first set of trials, which show promising results. Future work will include further improvements to the power planner algorithm and research on including data centre federation in the system.

9. ACKNOWLEDGMENTS

This work was carried out within the European Project DC4Cities (FP7-ICT-2013.6.2).

10. REFERENCES

[1] R. Basmadjian, G. Lovász, M. Beck, H. de Meer, X. Hesselbach-Serra, J. Botero, S. Klingert, M. Perez Ortega, J. Lopez, A. Stam, R. van Krevelen, and M. di Girolamo. A generic architecture for demand response: The all4green approach. In *Cloud and Green Computing (CGC), 2013 Third International Conference on*, pages 464–471, Sept. 2013.

[2] M. Bramfitt, A. Bard, R. Huang, and M. McNamara. Understanding and designing energy-efficiency programs for data centers, Nov. 2012.

[3] J. Bruinenberg, L. Colton, E. Darmois, J. Dorn, J. Doyle, O. Elloumi, H. Englert, R. Forbes, J. Heiles, P. Hermans, et al. Smart grid coordination group technical report reference architecture for the smart grid version 1.0 (draft) 2012-03-02. *CEN, CENELEC, ETSI, Tech. Rep*, 2012.

[4] P. Denholm, J. Jorgenson, M. Hummon, T. Jenkin, D. Palchak, B. Kirby, O. Ma, and M. OŠMalley. The value of energy storage for grid applications. *Contract*, 303:275–3000, 2013.

[5] M.-M. Diouri, O. Gluck, and L. Lefevre. Sesames: a smart-grid based framework for consuming less and better in extreme-scale infrastructures. In *Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing*, pages 187–194. IEEE, 2013.

[6] Florian Niedermeier et al. DC4Cities - D4.1: First results on renewable energy generation and smart city authority coordination of single DCs. `http://www.dc4cities.eu/en/wp-content/uploads/2014/12/D4.1-First-results-on-renewable-energy-generation-and-smart-city-authority-coordination-of-single-DCs.pdf`, 2014. Acccessed: 2015-03-25.

[7] D. Gmach, J. Rolia, C. Bash, Y. Chen, T. Christian, A. Shah, R. Sharma, and Z. Wang. Capacity planning and power management to exploit sustainable energy. In *Network and Service Management (CNSM), 2010 International Conference on*, pages 96–103. IEEE, 2010.

[8] I. Goiri, W. Katsak, K. Ley, T. D. Nguyen, and R. Bianchini. Parasol and greenswitch: managing datacenters powered by renewable energy. *ASPLOS '13 Proceedings of the eighteenth international conference on Architectural support for programming languages and operating systems*, pages 51–64, 2013.

[9] Jordi Guijarro et al. DC4Cities - D6.2: Description on the experimentation phase1, 2015.

[10] W. Katsak, Í. Goiri, R. Bianchini, and T. D. Nguyen. Greencassandra: Using renewable energy in distributed structured storage systems. Technical report, Rutgers University, 2015.

[11] S. Klingert, A. Berl, M. Beck, R. Serban, M. di Girolamo, G. Giuliani, H. de Meer, and A. Salden. Sustainable energy management in data centers through collaboration. In *Energy Efficient Data Centers*, volume 7396 of *Lecture Notes in Computer Science*, pages 13–24. Springer Berlin Heidelberg, 2012.

[12] K.-K. Nguyen, M. Cheriet, M. Lemay, M. Savoie, and B. Ho. Powering a data center network via renewable energy: A green testbed. *IEEE Internet Computing*, 17(1):40–49, 2013.

[13] Sonja Klingert et al. DC4Cities - D2.2: First market analysis. `http://www.dc4cities.eu/en/wp-content/uploads/2014/12/D2.2-First-market-analysis.pdf`, 2014. Acccessed: 2015-03-25.

[14] U.S. Department of Energy. Energy Efficiency & Renewable Energy. The importance of flexible electric supply. Technical Report 1 of 3, NREL National Renewable Energy Laboratory, May 2011.

[15] Y. Zhang, Y. Wang, and X. Wang. Greenware: Greening cloud-scale data centers to maximize the use of renewable energy. *Lecture Notes in Computer Science*, 7049:143–164, Dec. 2011.