THÈSE

Présentée et soutenue le *31/01/2023* par :
Igor FONTANA DE NARDIN

On-line scheduling for IT tasks and power source commitment in datacenters only operated with renewable energy

JURY

Premier MEMBRE	Professeur d'Université	Rapporteur
Second MEMBRE	Professeur d'Université	Rapporteur
Troisième MEMBRE	Professeur d'Université	Examinateur
Quatrième MEMBRE	Professeur d'Université	Examinateur
Cinquième MEMBRE	Professeur d'Université	Examinateur

Abstract

Résumé

Acknowledgments

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Context

Global warming is one of the biggest challenges humanity is facing. A recent rapport shows that we are walking toward a global mean temperature increase by 2100 of 2.7°C, well above the 1.5°C defined by the Paris Agreement [1]. The same rapport predicts the rise in mean global temperature will be around 1.8°C even after implementing all announced Paris Agreement goals. Achieving 1.5°C demands an engagement of all sectors to reduce greenhouse gas (GHG) emissions. One significant greenhouse gas (GHG) emitter is the Information and Communications Technology (ICT) sector. It produces around 1.8-2.8% of the world's total GHG [2]. Inside ICT, Data centers and transmission networks are responsible for nearly 1% of energy-related GHG emissions [3]. Due to its uninterrupted operation, the data centers sector is one of the most electricity-expensive ICT actors. A report revealed that Google data centers consumed the same amount of energy as the entire city of San Francisco in 2015 [4]. In addition, the situation tends to get even worse due to the improvements reduction in processor technologies and the predicted expansion of internet usage [5, 2].

Big cloud providers such as Google and Amazon are trying to reduce energy consumption and increase the power coming from renewable sources (RES) [6]. Renewable sources generate energy coming from clean sources such as wind and solar. A significant drawback of RES is they are highly dependent on weather conditions, creating power intermittence. These providers smooth this intermittence by not migrating entirely to RES, maintaining a connection to the grid. Therefore, they are not 100% clean. A renewable-only data center must consider this intermittence in its decision-making.

1.2 Problem Statement

1.3 Research Goals

1.4 Summary of Contributions

1.5 Publications and Communication

Submitted Peer Reviewed Conferences:

- I. F. de Nardin, P. Stolf and S. Caux, "Adding Battery Awareness in EASY Backfilling" 2023 IEEE 35th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD), Porto Alegre, Brazil, 2023.

Accepted Peer Reviewed Conferences:

- I. F. de Nardin, P. Stolf and S. Caux, "Analyzing Power Decisions in Data Center Powered by Renewable Sources", 2022 IEEE 34th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD), Bordeaux, France, 2022, pp. 305-314;

- I. F. de Nardin, P. Stolf and S. Caux, "Evaluation of Heuristics to Manage a Data Center Under Power Constraints", 2022 IEEE 13th International Green and Sustainable Computing Conference (IGSC), Pittsburgh, PA, USA, 2022, pp. 1-8;

- I. F. de Nardin, P. Stolf and S. Caux, "Mixing Offline and Online Electrical Decisions in Data Centers Powered by Renewable Sources", IECON 2022 – 48th Annual Conference of the IEEE Industrial Electronics Society, Brussels, Belgium, 2022, pp. 1-6;

- I. F. de Nardin, P. Stolf and S. Caux, "Smart Heuristics for Power Constraints in Data Centers Powered by Renewable Sources", Conférence francophone d'informatique en Parallélisme, Architecture et Système (COMPAS 2022), Jul 2022, Amiens, France. paper 7.

Others Disseminations:

- Talk: Analyzing Power Decisions in Data Center Powered by Renewable Sources, GreenDays@Lyon, March 2023.

1.6 Dissertation Outline

The remaining dissertation has the following organization:

Chapter 2 - Context and Related Work: This chapter presents the fundamentals to understand this dissertation. Considering the scope of the topic, the context consists of four parts. First, we introduce the context of global and ICT GHG emissions. Then, we describe renewable energy as an alternative to replace brown energy. After, we explain the usage of renewable to power a data center. Then, we define the uncertainties of weather and workload in a renewable-only data center. This last part also clarifies the importance of using predictions but with an online adaptation. After presenting the context, we introduce a list of works that solve part of our problem, highlighting the existing gaps in the state-of-the-art;

Chapter 3 - Modelling, Data, and Simulation: In this chapter, we describe the model to deal with the several elements that compose a renewable-only data center. Datazero2 creates a division between Offline and Online decisions. We present the model to deal with offline decisions using predicted power demand and production. Then, we demonstrate the output of Offline used by the Online. Finally, we define the Online model, which englobes the job scheduling and modifications in the Offline plan. After describing the model, we explain the source of the different data (e.g., workload, weather, servers) applied in the simulations. We present an explanation of the work done in the traces of the literature. Finally, we present the simulation tools used in this work;

Chapter 4 - Introducing Power Compensations: This chapter describes the proposed optimization to react to power uncertainties. We created four heuristics to find the best place to compensate for battery changes, which aim to reduce the number of killed jobs and the distance between the battery level and the target level. The results presented are related to the publications [7] and [8];

Chapter 5 - Learning Power Compensations: This chapter presents the idea and the results of the introduction of Reinforcement Learning (RL) in the power compensation problem. We propose two RL algorithms (Q-Learning and Contextual Multi-Armed Bandit) to learn the best moment to compensate;

Chapter 6 - Adding Battery Awareness in EASY Backfilling: This chapter explains a heuristic to mix scheduling and power compensation decisions. This heuristic is based on the EASY Backfilling scheduling algorithm but considers the battery's State of Charge to make better decisions;

Chapter 7 - Conclusion and Perspectives: Finally, in this chapter, we summarize the contributions of this work, providing a discussion about future works.

Chapter 2

Context and Related Work

2.1 Global Warming and ICT Role

2.2 Renewable Energy

2.3 Renewable-only Data center

2.3.1 Electrical elements

2.3.2 IT elements

2.4 Sources of Uncertainty

2.4.1 Weather Uncertainties

2.4.2 Workload Uncertainties

2.4.3 Optimization Strategies for Dealing with Uncertainties

2.5 Literature Review

2.5.1 Discussion and Classification of the Literature

Chapter 3

Modelling, Data, and Simulation

3.1 Model

3.1.1 Offline Decision Modules

Power Decision Module

IT Decision Module

3.1.2 Offline Plan

3.1.3 Online Decision Modules

Job scheduling

Modifying Power Plan

Modifying IT Plan

3.2 Data

3.2.1 Workload Trace

3.2.2 Weather Trace

3.2.3 Platform Configuration

3.3 Simulation

3.3.1 Simulator

3.3.2 Metrics

3.3.3 Datazero2 Middleware

3.4 Conclusion

Chapter 4

Introducing Power Compensations

4.1   Introduction

4.2   Model

4.3   Heuristics

4.4   Results Evaluation

4.5   Conclusion

Chapter 5

Learning Power Compensations

5.1   Introduction

5.2   Algorithms

5.2.1   Random

5.2.2   Q-Learning approach

5.2.3   Contextual Multi-Armed Bandit approach

5.3   States

5.4   Actions

5.5   Rewards

5.6   Results Evaluation

5.7   Conclusion

Chapter 6

Adding Battery Awareness in EASY Backfilling

6.1 Introduction

6.2 Model

6.3 Heuristic

6.3.1 Predictions

6.3.2 Job Scheduling

6.3.3 Power compensation

6.4 Conclusion

Chapter 7

Conclusion and Perspectives

7.1   Conclusion

7.2   Perspectives

Bibliography

[1] C. A. Tracker, "Massive gas expansion risks overtaking positive climate policies," *Warming Projections Global Update, November*, 2022.

[2] C. Freitag, M. Berners-Lee, K. Widdicks, B. Knowles, G. Blair, and A. Friday, "The climate impact of ict: A review of estimates, trends and regulations," 2021.

[3] G. Kamiya, "Data transmission networks," International Energy Agency, Paris, Tech. Rep., 2022.

[4] M. A. Khan, A. P. Paplinski, A. M. Khan, M. Murshed, and R. Buyya, "Exploiting user provided information in dynamic consolidation of virtual machines to minimize energy consumption of cloud data centers," in *2018 Third International Conference on Fog and Mobile Edge Computing (FMEC)*. IEEE, 2018, pp. 105–114.

[5] U. Cisco, "Cisco annual internet report (2018–2023) white paper," 2020.

[6] E. Masanet, A. Shehabi, N. Lei, S. Smith, and J. Koomey, "Recalibrating global data center energy-use estimates," *Science*, vol. 367, no. 6481, pp. 984–986, 2020. [Online]. Available: https://science.sciencemag.org/content/367/6481/984

[7] I. F. de Nardin, P. Stolf, and S. Caux, "Mixing offline and online electrical decisions in data centers powered by renewable sources," in *IECON 2022–48th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 2022, pp. 1–6.

[8] ——, "Analyzing power decisions in data center powered by renewable sources," in *2022 IEEE 34th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD)*. IEEE, 2022, pp. 305–314.