A parameterized test bed for carbon aware job scheduling

Ein parametrisierbares Testbed für kohlenstoffbewusste Jobplanung

Vincent Opitz

Arbeit zur Erlangung des Grades “Master of Science” der
Digital-Engineering-Fakultät der Universität Potsdam

A parameterized test bed for carbon aware job scheduling

Ein parametrisierbares Testbed für kohlenstoffbewusste Jobplanung

Vincent Opitz

Arbeit zur Erlangung des Grades “Master of Science” der Digital-Engineering-Fakultät der Universität Potsdam

A parameterized test bed for carbon aware job scheduling
(Ein parametrisierbares Testbed für kohlenstoffbewusste Jobplanung)
von Vincent Opitz

Arbeit zur Erlangung des Grades "Master of Science" der Digital-Engineering-Fakultät der Universität Potsdam

Betreuer:in: Prof. Dr. rer. nat. habil. Andreas Polze
Universität Potsdam,
Digital Engineering-Fakultät,
Fachgebiet für Betriebssysteme und Middleware

Gutachter:in: Prof. Dr. Jack Alsohere
University of San Serife
Facultiy of Computer Doings

Amelia van der Beenherelong
ACME Cooperation

Datum der Einreichung: 27. August 2024

Zusammenfassung

ummarize thesis

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Contents

1 Introduction **3**

2 Background **5**

3 Related Work **7**

4 Methodology **9**

4.1 Improving the current Job model 9

4.1.1 Power Measurements on Machine Learning Jobs 9

4.1.2 Defining a new model 9

4.2 Choosing an implementation approach 10

4.2.1 Carbon-aware scheduling via a slurm plugin 10

4.2.2 Using a Simulation approach 10

4.3 building ontop of the existing gaia sim 12

4.4 Evaluating carbon-aware scheduling with the new job model 12

5 Results **13**

6 Discussion **15**

7 Future Work **17**

References **19**

Zusammenfassung

S

ummarize thesis

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

1 Introduction

- Warum sollte man sich mit Scheduling von Jobs in Data Centers beschäftigen?
- why is carbon aware scheduling useful? we want to emit less carbon per unit of work, reason being climate change
- welche contribution mache ich konkret?
- which hypothesis do I have?
- for example, we could think that having more information about how our workload's power changes over time, could lead to higher carbon efficiency, as we could for example match high-power to low-carbon periods.

warum carbon aware scheduling? ? ? ? welche forschungsfragen

2 Background

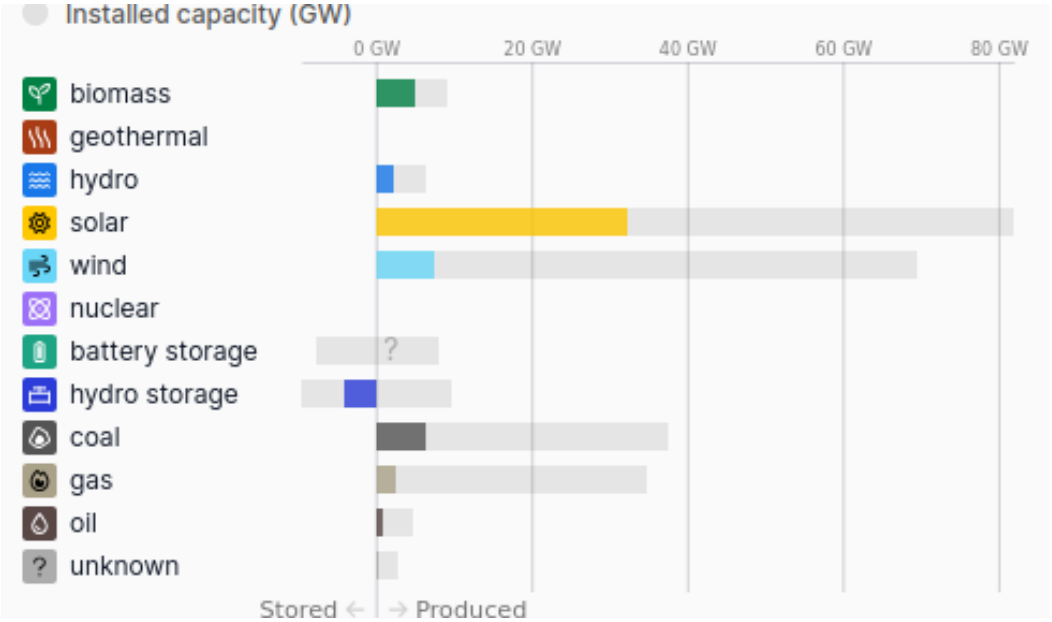
introductory text, easing into this chapter

Computers need electricity to compute. This electricity is usually sourced from the public and local grid, which in turn is made up of many different energy producers. This energy mix is composed of different sources: low-carbon technologies like solar, wind, or hydroelectricity and carbon-intensive sources like coal, gas, or oil.

Over the day, the demand and supply for power changes. For example, during the midday, solar is produced as the sun shines. In the morning or evening hours however, as people may enjoy a cup of coffee or watch television, additional

Thus, at each point in time the amount of CO_2 per unit of electric power can be determined by averaging the amount of power each source supplies to the grid and how much carbon it emits.

- power grid is made up of different sources
- these sources have a varying amount of carbon per unit of power (solar is efficient, while burning oil is not)
- this mix changes throughout the day, solar power can only be produced as the sun shines, burning fossile fuel can be done whenever and is needed to supply the base load.
- carbon-aware scheduling thus assumes that, if we schedule our jobs during low carbon periods, we emit less CO_2 for the same amount of work
- this only works for deferrable or batch workloads [1]
- secondary benefits such as potentially a lower cost using dynamic energy prices



Types of Carbon Aware Scheduling

- There are different ways to exploit the dynamic carbon emissions: temportal shifting, spatial, ressource scaling. - what do we focus on in this thesis and why?

Different Signals of the grid

- marginal and relative carbon
- what does each signal mean for making decisions?
- why do we use the relative signal?

there was some discussion about that in my bib; if we point to that it'll surely be fine

hintergrund energie netz, stormproduktion welche arten davon gibt es welche benefits hat das? hintergrund signals nach denen entschieden wird (marinal / relative carbon) welche metriken gibt es POI-A, etc etc.

- energy is produced by different sources, leading to a different amount of carbon/unit of energy over time.
- Welche Methoden gibt es carbon einzusparen (vllt. auch in der related work) (temporal, spatial, ressource scaling)
- Begründen warum sich das lohnen kann (marinal / relative carbon), kein wasting von erneuerbaren energien (losses among saving energy into batteries / "curtailment" !!)
- welche metriken gibt es POI-A, etc etc.
- Jobs <-> dynamic ernergie verbrauch, wie wird sowas gemessen?

3 Related Work

A systematic approach I used the following system for finding related work to my topic; first, my supervisors and I would brainstorm for search engine keywords. The specific words are in the attachments, but can be grouped into two groups: one for all things carbon-aware and one for keywords about servers / computing / HPC and such.

Literaturrecherche aufarbeiten und hier verlinken

Using these two groups, I could then create a google scholar queries via the cross product between them. Using the double-quotation feature would further limit the results. For each query, I would then read the abstracts of roughly the first 5 results, depending on if their titles sounded subjectively fit. Additionally, I would also further explored papers by finding new ones through *connected papers*[1] These would then be entered into a spreadsheet: for each query, 5 paper titles. I then "rated" them into three categories:

- green, meaning that they seem very connected and are good first entries into the topic
- orange, which would indicate that are are somehow connected to the paper and might be read at a later date
- red, the paper is either irrelevant or had some other flaw. These would not be touched again in the course of my work

With this approach, X abstracts were read. Figure 3.1 shows that X abstracts were deemed "good" for the purpose of this work. The complete list of papers analysed can be found in

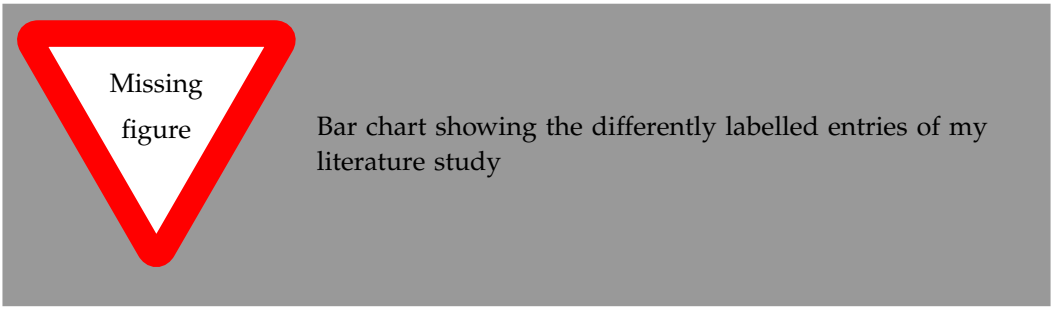
figure out results of the literature stufy

THE ATTACHMENTS

I would like to highlight some papers:

GreenSlot: Scheduling energy consumption in green datacenters[0] seems to be the first paper that deals with carbon aware scheduling by implementing it as a slurm plugin. In

[1] https://www.connectedpapers.com/



Bar chart showing the differently labelled entries of my literature study

Figure 3.1: Results of the literature study

contrast to our scenario, where we try to optimize carbon emissions via the public electricity grid, GreenSlot is about datacenters having their own renewable energy production (solar panels on the roof). Using weather data, *GreenSlot* would then predict when solar energy production is high, scheduling jobs to those time frames.

The War of the Efficiencies: the Tension between Carbon and Energy Optimization [o] outline the different ways of carbon aware computing. Among those is *temporal shifting*, the idea that jobs can be executed later when energy is more carbon efficient, is also the main idea for my work. They also use *spatial shifting*, moving jobs across the globe to areas where higher carbon efficiency is possible. *Ressource scaling* uses dynamic amounts of hardware according to carbon emissions. In the end, *rate shifting* is the idea to also scale hardware frequencies. During carbon-efficient times, CPU speeds would be increased, leading to faster processing speeds and more energy usage.

All of these techniques are then tested under various parameters. My work will only make use of the temporal shifting, and abstract away the other methods of further saving carbon.

Let's wait awhile: how temporal workload shifting can reduce carbon emissions in the cloud is to be one the paper I'll be building on the most. [**wiesner_lets_2021**] uses a simulation approach, to simulate temporal shifting. Their workload model consists of known length jobs that can use *checkpoint & restore* to be executed at different time slices. They would further use different traces and test these traces under the assumption of different job-deadlines, meaning that each job would have to be completed by a certain timeframe, and also different regions of the world as described in the background part. Their main take-aways are that increased deadlines lead to reduced carbon emissions, but that this effect also has diminishing returns. They also deduced that regions such as california, with high amounts of solar power, have higher potential for carbon-savings in comparison to nuclear-heavy regions such as france.

think about linking this to the earlier part

4 Methodology

Based on our related work, describe (using the goals outlined in the beginning) the approach I take: creating a better model to be used in carbon-aware scheduling using the model and evaluating it.

4.1 Improving the current Job model

We should probably argue why the current job model used in literature is not sufficient (WaitAWhile assumed constant power usage and no overhead from stoppig and resuming)

4.1.1 Power Measurements on Machine Learning Jobs

Für die Durchführung von Messungen ist es zwingend erforderlich, die verwendete Testumgebung (Hardware wie auch Software) zu dokumentieren sowie Messparameter gewissenhaft zu wählen. Zu den wichtigsten Parametern gehren die Anzahl der Messwiederholungen, die Anwendung von Warmup-Läufen sowie die Identifikation und Vermeidung möglicher störender Einflüsse.

Describing the experiment setup

- I used an MCP, why is that preferrable over using another power measurement tool like rapl and such?
- what hardware did I use?
- describe the specifics of my measurement setup (multiple runs, system at rest, measure before and after)
- also be sure to mention that we added logging to the script execution (to be used for the phases later)

Measurement results

- Warum ist das interessant
- which jobs did I measure and why?
- Wie habe ich die Messungen ausgeführt, MCP beschreiben, sowie pinpoint als Schnittstelle
- Experimentier-parameter, also wodurch versuche ich sicherzustellen, dass meine Messungen auch sinn machen / reproduzierbar sind usw usw.
- Wie habe ich das ausgewertet, schöne graphs zeigen

4.1.2 Defining a new model

Now that we know what a high-level job looks like, we can pick it apart and reduce the real-world measurements of one program to a more generic model.

- we can deduce phases
- each phase has a constant power draw
- give an example of how to represent the real-world measurements into a model
- now we should proof that the model actually represents the reality to a certain degree (error analysis)
- have a cute graph showing the measurements and the model-"measurements" next to each other
- also show that the stop-resuming functionality can be represented with our model

4.2 Choosing an implementation approach

We first need to explain why we chose our approach (building upon exisiting work inside GAIA). The other option that is not using a simulation would be to schedule real jobs, for example by creating a slurm plugin.

We can then evaluate how well a slurm plugin would work for our given Forschungsragen. End that section by deeming the plugin idea as unfit, we can then shift to arguing for the simulation approach as that is also something that just came out in related work (perhaps we should see wether we list GAIA as related work or introduce it just then)

4.2.1 Carbon-aware scheduling via a slurm plugin

thank god I made notes

- why do we choose slurm specifically, and not other software like kubernetes etc. => because its also being used in scorelab at the same location as I am in
- was ist slurm? => "open source, fault-tolerant, and highly scalable cluster management and job scheduling system for large and small Linux clusters"
- how does the slurm plugin system work? what do we need to do to get there?
- what problems occured?
- ...thus I ultimatly choose not to pursue implementing a plugin

4.2.2 Using a Simulation approach

Thankfully, just at that time a new paper [o] was released. They made a prototype testbed for simulating job scheduling on cloud providers. These Jobs could be executed on spot instances (cheap VMs that seek to increase cloud utilization), on-demand instances (short-notices VMs that are thus more expensive) or pre bought VMs (medium cost, but may be wasted), the paper then discussed balancing carbon- and dollar costs.

improve this

The good part is that within that testbed, many scheduling approaches outlined in the related work section were implemented - a WaitAWhile Implementation for example . I could now extend this testbed and would ontop get something to compare against!

The way to do this section would be to a) describe what was aleady there, and b) what I changed

This would show a class diagram similar to the one in the GAIA paper, with markings which files I edited / added / deleted

Describe the figure, explain why I am for example removing the part about the dollar-costs and the slurm-scheduler adapter. I should also describe which parts of the program I am modifiying to tackle my Forschungsfragen

Stuff I should describe about the simulation before I added anything:

Assumptions of the simulation

- Joblängen sind bekannt
- Jobs können zeitlich verschoben werden (begründet daraus, dass sie als Batch Jobs submitted werden, andere Jobs werden hierbei nicht betrachtet)
- User geben dabei an, wie lange der Job verschoben werden darf
- Die carbon curve auf dem electrical grid ist für kurze Zeiträume in der Zukunft bekannt
- Die Hardware ist zZ nicht begrenzt. Das war in der related work auch nicht so. Eigentlich wäre es spannend sich das anzuschauen, allerdings sind die bisherigen Scheduler halt darauf garnicht gemünzt, da werden alle Jobs unabh. voneinander gescheduled. Man könnte das via publicCloud argumentieren, allerdings wäre das questionable, in wie fern der scorelab trace benutzt werden kann (da das ja auf in einem lokalem datacenter läuft)
- TODO: Joblängen sollten dem Scheduler nicht bekannt sein. Die Workloads aus GAIA werden allerdings so gescheduled als ob man perfekte Knowledge hat. Das reicht zwar für ein upper bound an carbon savings, ist aber nicht sehr realistisch.

Data being used here i could describe which data is already being used (the traces, aswell as the historical carbon data)

- Welche Traces gibt es, wodurch werden die characterisiert? (Länge, Anzahl, etc, etc) Vllt. kann man hier nen coolen vergleich erstellen, Auch könnte man ein paar Sätze darüber schreiben, wie die bisher in GAIA aufgenommen wurden.
- Wie den scorelab trace benutzen und übersetzen? Gerne auf ner halben Seite aufschlüsseln, was die einzelnen Attribute aus sacct bedeuten.
- Ansonsten kann man noch die dynamic ernergy sachen als Datenquelle auflisten, bzw. das mini experiment mit fmnist und roberta

4.3 building ontop of the existing gaia sim

Which parts of GAIA do I add on? => this should just be the schedulers and the part where the carbon is calculated, this ensures that

4.4 Evaluating carbon-aware scheduling with the new job model

Hi!

5 Results

- here we would try to show off the difference between power-and-phase-oblivious scheduling and my new implementation which can make use of that
-

6 Discussion

welche schlüsse können wir aus den ergebnissen ziehen?

7 Future Work

Bibliography

[0] Walid A. Hanafy, Roozbeh Bostandoost, Noman Bashir, David Irwin, Mohammad Hajiesmaili, and Prashant Shenoy. "The War of the Efficiencies: Understanding the Tension between Carbon and Energy Optimization". en. In: *Proceedings of the 2nd Workshop on Sustainable Computer Systems*. Boston MA USA: ACM, July 2023, pages 1–7. ISBN: 9798400702426. DOI: 10.1145/3604930.3605709 (cited on page 8).

[0] Walid A. Hanafy, Qianlin Liang, Noman Bashir, Abel Souza, David Irwin, and Prashant Shenoy. "Going Green for Less Green: Optimizing the Cost of Reducing Cloud Carbon Emissions". In: *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 3*. Volume 3. ASPLOS '24. New York, NY, USA: Association for Computing Machinery, Apr. 2024, pages 479–496. ISBN: 9798400703867. DOI: 10.1145/3620666.3651374 (cited on page 10).

[1] Andrew S. Tanenbaum and Albert Woodhull. *Operating systems: design and implementation: [the MINIX book]*. en. 3. ed. The MINIX book. Upper Saddle River, NJ: Pearson Prentice Hall, 2006. ISBN: 978-0-13-142938-3 978-0-13-505376-8 978-0-13-142987-1 (cited on page 5).

[0] Íñigo Goiri, Kien Le, Md. E. Haque, Ryan Beauchea, Thu D. Nguyen, Jordi Guitart, Jordi Torres, and Ricardo Bianchini. "GreenSlot: scheduling energy consumption in green datacenters". en. In: *Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis*. Seattle Washington: ACM, Nov. 2011, pages 1–11. ISBN: 978-1-4503-0771-0. DOI: 10.1145/2063384.2063411 (cited on page 7).

Eidesstattliche Erklärung

Hiermit versichere ich, dass meine Arbeit zur Erlangung des Grades “Master of Science” der Digital-Engineering-Fakultät der Universität Potsdam mit dem Titel “A parameterized test bed for carbon aware job scheduling” (“Ein parametrisierbares Testbed für kohlenstoffbewusste Jobplanung”) selbständig verfasst wurde und dass keine anderen Quellen und Hilfsmittel als die angegebenen benutzt wurden. Diese Aussage trifft auch für alle Implementierungen und Dokumentationen im Rahmen dieses Projektes zu.

Potsdam, den 27. August 2024

_____

(Vincent Opitz)