

Researcher studies power consumption of cloud infrastructures

September 4 2019



Credit: Victorgrigas/Wikideia/ CC BY-SA 3.0

In his doctoral dissertation at Umeå University, Jakub Krzywda has developed models and algorithms to control tradeoffs between the power consumption of cloud infrastructures and performance of hosted applications to enable safe and efficient operation under a limited power budget.

Cloud computing [infrastructure](#), which keeps the majority of internet services such as Google, Facebook and Amazon up and running, consume massive amounts of energy, and thus exacerbate climate change.

Interestingly, under some specific conditions, a [data center](#) is in principle not much different from a house with an old electrical installation. Most people can probably recall a case when turning on too many appliances simultaneously tripped a circuit breaker. That happens because the electrical installation was not meant to sustain such high [power](#) surge.

In modern data centers, the power delivery infrastructure, which supplies all the servers with electricity, is often underprovisioned on purpose. In this case, it is unable to sustain the power surge of all the servers running at their full speed. At first glance, it sounds like bad planning, but in practice, it almost never happens that the computing power of all the servers is needed at once. Since the cost of power delivery infrastructure is proportional to the peak power it can sustain, putting a cap on it helps the data center operators to save money that otherwise would be spent on infrastructure that is almost never needed.

However, "almost never" is not enough in the cloud industry. Many cloud providers promise their customers that the infrastructure will be available 99.99 percent of the time—it allows only 52 minutes of downtime per year.

And this is where the contributions of this thesis come in: What is the best way to handle data center operations when there is not enough power available to run all the applications at their full speed? Should operators shut down less important applications completely, or force all of them to slow down? Are some types of applications better candidates for graceful performance degradation? What actions should be taken to ensure that power consumption is reduced but applications still produce

useful results? Which techniques shall be used to enforce that?

"In order to answer the above-mentioned questions, in this thesis, I have developed models to capture relationships between [power consumption](#) and application performance, and proposed a set of power budgeting controllers that act at the application, server, and whole data center levels to minimize performance degradation while enforcing power limits," says Jakub Krzywda.

The findings included in the thesis have practical applications, for example, a set of recommendations for using software techniques available in modern servers that can allow data center operators to run their infrastructures using less power while still ensuring that their customers are satisfied with the performance of applications.

The results and analysis presented in this thesis can be used by data center operators to improve the power efficiency of [servers](#) and reduce overall operational costs while minimizing performance degradation. All of the software produced during this work, including source codes of models, controllers, and simulators, has been open-sourced and made available online to facilitate its deployment in both research and industrial data centers.

More information: The dissertation is available online:
[urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-161363](https://nbn-resolving.org/urn:nbn:se:umu:diva-161363)

Provided by Umea University

Citation: Researcher studies power consumption of cloud infrastructures (2019, September 4) retrieved 20 March 2024 from <https://techxplore.com/news/2019-09-power-consumption-cloud-infrastructures.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.