

New method to predict the workload for online services

September 30 2015

How can overloads on the Internet be prevented? Cloud computing means that more server space can be rented from large computing resources. Ahmed Hassan has developed algorithms for automatic addition and removal of server resources for a web service based on demand. The study was performed at Umeå University in Sweden.

When Michael Jackson passed in 2009, he almost took the Internet with him. CNN, Google News, TMZ, Twitter, Wikipedia, and most other news websites went down due to the huge increase in the number of people trying to read news about his death. While Jackson's passing is a single event, similar events are quite common, mostly affecting some parts of the Internet, e.g. the servers of the ObamaCare website were overloaded already after a few hours after going live in the United States, and studera.nu, the website for applying to Swedish universities used to be overloaded every year during the application rounds. In short, the Internet infrastructure that our lives rely on is fragile.

Cloud computing is a new IT model that allows for better management of the server resources that the websites depend on. For example, studera.nu now runs in the cloud. A cloud enables ubiquitous network access to a shared pool of configurable [computing resources](#). These computing resources can be easily 'rented' and 'released', similar to a utility that one can use at home, e.g. if you plug in a fridge, you get electricity and pay for the amount of electricity you use. In the Cloud model, a customer can rent computing resources when required, and release them when they are no longer needed. For a website owner, this

means that resources can be rented on demand to handle the overload.

"My dissertation introduces methods and algorithms to automatically add and remove resources to a [web service](#) based on the actual demand. Renting more cloud capacity than needed to run a web service is a costly choice. Renting too little capacity will result in server overloads and service disruptions. To better choose the right amount of resources to rent, the dissertation introduces prediction algorithms for future demand on a web service," says Ahmed Hassan.

To understand how to design good prediction algorithms, the dissertation also contains analyses of some server workloads from major web services.

"First, I analysed the usage of Wikipedia for a period of five and a half years looking at what happens when major events occur such as Michael Jackson's death. We also analysed how people use the premium services of TV4 Video-on-Demand. Some interesting findings is how impatient users are when streaming a video, with 50% of the users abandoning streaming a video after watching less than 12 minutes," says Ahmed Hassan.

The final part of the dissertation demonstrates a tool that is capable of predicting the capacity requirements of various services running in the cloud by using more than one prediction algorithm, improving the overall performance of multiple services at the same time.

The dissertation work was performed in collaboration with multiple people from the Departments of Mathematics and Mathematical Statistics at Umeå University, the Department of Control and the Department of Electrical and Information Technology at Lund University. Many of the ideas in the dissertation are being developed and used in Elastisys, a start-up that was founded based on results from our

research group. Parts of the work is patented by Elastisys.

More information: "Workload characterization, controller design and performance evaluation for cloud capacity autoscaling." umu.diva-portal.org/smash/record.jsf?pid=diva2%3A852794&dswid=-7376

Provided by Umea University

Citation: New method to predict the workload for online services (2015, September 30) retrieved 25 April 2024 from <https://phys.org/news/2015-09-method-workload-online.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.