Two Meetings on Green Data Storage at DIMACS

Organizers:
Stark Draper (University of Toronto)
Emina Soljanin (Bell Labs – Murray Hill, US)

Two short meetings on "Green Data Storage" were held at the Center for Discrete Mathematics and Theoretical Computer Science (DIMACS), Rutgers University, New Jersey, on Dec. 16–18 2013. The event brought together researchers from the coding, queueing, and systems communities, studying data storage in the cloud, operation of data centers, and content delivery. The goal was to better understand performance measures, modeling issues, evaluation methodologies, and how to bridge theory and practice.

Data centers' energy requirements have increased massively in recent years. A substantial (and growing) percentage of worldwide electricity is dedicated to cloud computing, which motivates the design of energy-efficient storage techniques. In the U.S. alone about 100 billion kWh of energy at a cost of $7.4 billon is consumed annually by these centers, producing as much CO2 as all of Argentina. Yet, much of this power goes to waste. Many servers are busy only 5–30% of the time and up to 30% of internal network traffic is devoted to reliability, dealing with equipment failure and the threat of data loss. One reason for these inefficiencies is that users of cloud systems demand that their services be readily available and their content be reliably stored. Cloud providers today strive to meet both demands through overprovisioning: simply keeping processors ready to go at all times and replicating content throughout the storage network over multiple disks, in both cases increasing energy requirements. A large body of recent literature draws on queuing analyses to model and control both processors and content distribution. Erasure and network coding has been proposed as a more bandwidth and storage efficient way to provide reliability and accessibility of stored content. These smart processor and storage algorithms promise to cut data-center energy use significantly, but have so far been studied and evaluated mostly in simplified theoretical settings, without taking into account all systems issues that may arise by introducing the algorithms into practical storage systems.

The working group and workshop brought together experts in coding and queueing theory for distributed storage together with systems experts in order to better understand the issues involved and to find ways to evaluate how well the practice would match the theoretical predictions of energy savings. The first two days took the form of a DIMACS "Working Group" with eight keynote talks by Tom Bostoen (Bell Lab), Alex Dimakis (UT Austin), Muriel Medard (MIT), Thu Nguyen (Rutgers), Kannan Ramchandran (UC Berkeley), Mor Harchol-Balter (CMU), Minlan Yu (USC), and Anshul Gandhi (IBM & Stony Brook). The first day also featured an icy roof-top tour of Rutgers solar powered PARASOL experimental micro-datacenter (http://parasol.cs.rutgers.edu) operated by Profs. Nguyen and Bianchini. The third day was a DIMAC "Workshop" featuring 16 talks by students, faculty, and industrial researchers from institutions across North America and Europe.

"The DIMACS workshop combined the benefits of a full-size conference with hard-hitting quality talks and the schedule of a small workshop that encourages discussion; I saw DIMACS sow the seeds for many a collaboration. I gained as many insights about the state of the field as from two or three normal conferences combined, and getting to know thought leaders at a more personal level was fantastic." —Ulric Ferner, PhD student, MIT

"This was an amazing workshop! It brought together people from information theory, queuing theory, computer systems with a common aim: energy efficiency. It was exciting to see different research approaches and styles on the same topic." —Yanpei Liu, PhD student, UW-Madison

"This was a well-organized workshop with great speakers, from which I not only learned the recent progress of the applications of Information Theory to data storage, but more importantly, had the chance to interact with lots of world-class researchers." —Zhenhua Liu, PhD student, Caltech

"The DIMACS Working Group on Algorithms for Green Data Storage was a unique opportunity for me, as a queueing theorist, to meet people in the area of network coding, and learn of problems that are quite different from standard queueing models. I was exposed to several new problems in multi-server queueing systems, motivated by error-correcting codes and energy management, wherein jobs are purposely replicated across all servers, and are considered "complete" when a subset of the servers complete their tasks. At this point, I have read all the papers on this topic by Prof. Kannan Ramchandran and by Dr. Emina Soljanin, both of whom I met at the workshop. I am working with my students on researching new solutions to some of the problems that Kannan and Emina in the future." —Mor Harchol-Balter, Professor, CMU.

"I learned a lot about energy management in storage and networking. It was very interesting to learn more about coding, which is typically outside of my research area, and how one might use it for improving energy efficiency. The workshop helped me to think about energy efficiency and management more holistically." — Thu Nguyen, Professor, Rutgers

The program with abstracts can be found at
http://dimacs.rutgers.edu/Workshops/WGGreen/program.html for the Working Group and at
http://dimacs.rutgers.edu/Workshops/Green/program.html for the workshop.
Talk slides are available at http://dimacs.rutgers.edu/Workshops/Green/Slides/slides.html