Energy-efficient data centres

Data centres house the IT infrastructure required for the smooth running of everyday life, but waste large amounts of energy. The **Center for Energy-Smart Electronic Systems** aims to make data centres more energy efficient by developing cutting-edge systems to monitor and regulate their energy use

HOME TO CRITICAL IT infrastructure including servers and data storage equipment, data centres power a significant proportion of everyday online activity, from using a search engine to shopping online. As the use of the internet has grown substantially, and continues to grow, demand for data centres will too.

Ever increasing in scale, data centres today have significant energy requirements. In the US, such facilities are rapidly becoming one of the biggest energy consumers, owing to the need to both run and cool the equipment they contain. In fact, the energy spent running them comprises around 2.5 per cent of the national energy expenditure – enough to power two medium-sized cities for a year.

The growing demand for online services, and the concomitant increase in data centres that accompanies this demand, brings with it a need for greater energy efficiency. Recognising this, the Center for Energy-Smart Electronic Systems (ES2), a National Science Foundation (NSF) Industry-University Cooperative Research Center (I/URC), is searching for new ways to reduce data centre energy use.

This innovative consortium is made up of four universities from the US: Binghamton University, Georgia Institute of Technology, University of Texas at Arlington and Villanova University, together with over 20 member companies. These companies represent the entire data centre supply chain, from hardware manufacturing and software development to end users, and include industry leaders such as Intel, IBM and Facebook.

INTELLIGENT TECHNIQUES

ES2 brings together computer scientists and mechanical engineers, sharing expertise from the disciplines of IT, dynamic systems control, electronic systems and electronic cooling. A holistic effort to develop energy-smart electronic systems, ES2 also includes industry and government.

The central aim of this unique collaborative research centre is to develop new and efficient methods to operate electronic systems by controlling resource use and managing workloads in order to achieve optimal energy consumption. Intelligent techniques are at the crux of this: ways to enable electronic systems to self-regulate the amount of energy they use. In real terms, this means keeping as many machines on as are needed to deliver current demand – no more and no less – using predictive techniques to estimate future demand and provision the exact power needed, much like the way in which power stations function. Ultimately, the Center aims to reduce the overall power consumed by data centres. To achieve this, each member university is conducting research in a different area – contributing one piece of the energy-efficiency puzzle.

Dr Bahgat Sammakia and PhD student Husam Alissa discussing their research project in the ES2 Data Center at Binghamton University.

BINGHAMTON UNIVERSITY

New York's Binghamton University, represented by Center Director Professor Bahgat Sammakia, and Binghamton ES2 Site Director and Chair of the Computer Science Department Professor Kanad Ghose, conducts research in two key areas: energy proportionality and thermal management. One of their projects aims to tackle the overprovisioning of IT capacity – a common practice in industry which causes significant energy losses. "Overprovisioning is done to deal with workload fluctuations that may lead to performance losses. It involves operating many servers at low utilisation levels, where they have poor energy efficiency," Ghose explains. Many operators of online services keep all of their servers running, regardless of demand. This is extremely inefficient, as even lightly loaded or idle servers use energy and require significant amounts of energy to be cooled.

There is a clear need for more agile practices from the operators of online services, and greater energy proportionality – in other words, only running the servers that are needed. While energy-aware scheduling has been investigated previously, most past research has failed to consider job characteristics or environmental conditions, such as the temperatures near servers.

Without considering these conditions, it is impossible to predictively and synergistically manage servers. To create a scheduler that does consider these factors, Ghose's team applies short-term workload prediction to provision the precise resources needed to handle the immediate workload, without causing any loss of performance. This also means any active servers will run at high utilisation levels – acting on many requests at the same time, where they are most energy efficient.

The novel scheduler, called the Smart Load Balancer (SLB), is an energy-aware data centre load balancing and management system, providing power in an energy-proportional manner. To do so, it concentrates work on the smallest subset of servers possible, such that performance is unaffected while energy is saved. The remaining servers enter idle or sleep states, or are even powered off altogether in order to minimise energy consumption. "Smart allocation is used to allocate workload to the active servers in a way that avoids performance loss while delivering significant energy savings," Ghose summarises. When applied to a data centre containing 150 servers, SLB reduced overall energy consumption by almost 30 per cent.

Sammakia himself is leading another of the projects at Binghamton focused on

the thermal management of data centres. Containment configurations, which prevent the mixing of hot and cold air streams, are commonly used to reduce the amount of energy needed for cooling. However, they can also change the way the IT equipment behaves. Moreover, equipment must compete for the cold airflow supply coming from the tiles beneath, leading to under-provisioning for some servers and poor distribution generally.

To tackle these issues, Sammakia is further investigating containment schemes and characterising the main components of the airflow path. This requires a fundamental understanding of the physics of air and water flow through a data centre, and the impact of temperature on these processes, and feeds into Ghose's efforts: "The first thing to do is allocate workload, then predict the power distribution in the data centre, and finally predict the required thermal management throughout the data centre – these have to work seamlessly together," Sammakia explains. In time, these factors will be coupled through a control system being developed at the University of Texas at Arlington.

UNIVERSITY OF TEXAS AT ARLINGTON

Owing to its longstanding focus on data centre cooling, the University of Texas at Arlington (UTA) is perfectly poised to contribute to ES2 efforts. Cooling is an important issue for the industry, as IT equipment in data centres must be continuously cooled in order for it to operate reliably. Evidence suggests that around half of the total electricity used by such facilities is for cooling.

UTA operates a wide range of projects, covering many different aspects of data centre engineering: "A lot of our research focuses on the areas of IT equipment reliability, oil immersion cooling, cold plate designs and data centre cooling," elaborates Kanan Pujara, research assistant at UTA.

One particular focus area for the University is liquid cooling: an efficient method to remove the heat from data centre electronics that, as its name suggests, involves cooling heated computer components with fluids. Dr Dereje Agonafer, the UTA lead in ES2, is fielding a project to develop cold plates: metal plates through which a liquid coolant is forced. This ongoing research aims to reduce the negative effects of non-uniform power distribution via intelligent sensing.

Since the introduction of Intel's Pentium III microprocessors in 1999, non-uniform power distribution has become common. Varying amounts of power are assigned to different functional units leading to 'hot spots'. These substantial temperature differences are detrimental to performance and stability, and are often observed in high-end devices

such as multi-core chip-scale packages, a type of microchip used in a huge range of electronics.

In order to reduce these temperature differences and therefore the costs of cooling, a dynamic approach is needed whereby resources are allocated based on power distribution. Through this project, the UTA team will develop and test dynamic cold plates for the localised cooling of single-chip modules; electronic packages used by the majority of industry today. The project is making good progress: "Extensive computational analyses have led to a finalised design, samples of which have been fabricated by an industry partner," Agonafer details. "These will be evaluated experimentally in the near future".

Rack testing in the ES2 Data Center at the University of Texas at Arlington.
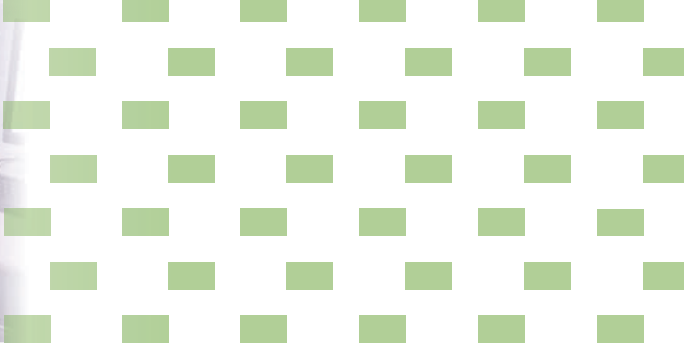
VILLANOVA UNIVERSITY

Efforts to more effectively cool electronics, such as those ongoing at UTA, require expertise in heat transfer, fluid dynamics and thermodynamics – which is where Villanova University in Pennsylvania comes in. This institution, represented by Dr Alfonso Ortega, is contributing expertise in the thermal management of electronic systems, and in modelling and characterisation of energy management systems, in order to aid ES2 in the design of more energy-efficient data facilities.

Like UTA, Villanova researchers are also investigating liquid cooling. In a project led by Professor of Mechanical Engineering Dr Amy Fleischer and Ortega, the Villanova team will look beyond the immediate benefits of commonly used single-phase liquid cooling in order to evaluate the potential of two-phase cooling, which allows a higher coolant operating temperature. Two-phase cooling through microchannel cool plates could increase the efficiency of waste heat recovery. A waste heat recovery testbed is under construction in order to evaluate energy efficiencies gained by coupling the server cooling system to an energy recovery system.

While energy is central to the work of ES2, the lesser known concept of exergy is also an important consideration. Exergy is the available energy in a system, and wasteful practices are said to 'destroy exergy'. Estimating the amount of exergy destruction in different parts of the system is therefore a useful means by which to identify wasteful practices.

Aaron Wemhoff, Associate Professor in Mechanical Engineering, is leading a project investigating new approaches to designing data centres based on the exergy concept. Towards this aim, Wemhoff and the Villanova team created the Villanova Thermodynamic Analysis of Systems (VTAS) tool – a flow network tool for modelling the thermodynamics, fluid mechanics and heat transfer within a data centre. Having used VTAS to show that exergy destruction can be significantly reduced by introducing on-server or near-server air-liquid heat exchange, Wemhoff's team now aims to enhance VTAS by adding capacitance effects and developing a library of validated, platform-independent component models. These changes will enable VTAS to explore new scenarios, and permit the integration of component models being developed in other ES2 projects.

Drs Alfonso Ortega and Kamran Fouladi taking measurements in their lab at Villanova University.

GEORGIA INSTITUTE OF TECHNOLOGY

The final university partner of ES2, Georgia Institute of Technology, contributes capabilities for measuring and testing airflow in data centres. Using their facility to reproduce the cooling capabilities of modern data facilities, and instruments to characterise air temperature flow distribution, the Georgia Tech team is conducting a number of projects to optimise airflow.

In the first, led by Professor Yogendra Joshi – also the Institute's representative in ES2 – the researchers will investigate how air flows through the perforated floor tiles commonly found in data centres. Past studies conducted by Joshi suggest the geometry of these tiles can significantly affect downstream flow and cool air delivery. Joshi's team will now directly resolve the geometrical structure of commercially available tiles and devise new strategies and modelling guidelines to cope with their complex features. In addition, the project team will investigate how air is delivered through the tiles, which will validate computational fluid dynamics models being developed by Agonafer at UTA.

In another thermal management-focused project, Joshi is investigating the varying thermal responses of different data centres.

As a result of the continuously changing computational, and thus thermal, loads on servers, data centres have dynamic thermal environments. A sophisticated management scheme is needed to assess the local and global cooling requirements of these facilities, and to supply the requisite optimal cooling resources. At present, most data centre thermal control is limited to pre-set air temperatures maintained by computer room air conditioning units. To improve on this static model, Joshi will incorporate other parameters such as server inlet air and – more importantly – chip temperatures.

Instrumentation in the research lab at the G W Woodruff School of Mechanical Engineering, Georgia Institute of Technology.

DATA CENTRE ENERGY CONSUMPTION: THE FACTS

- Data centers are one of the largest and fastest growing consumers of electricity. In 2013, US data centers consumed an estimated 91 billion kWh of electricity – enough to power all the households in New York City twice over – and are on-track to reach 140 billion kWh by 2020

- The largest issues and opportunities for energy savings include the underutilisation of data centre equipment and the misalignment of incentives, including in the fast growing multi-tenant data centre market segment

- ES2 aims to improve the energy efficiency and productivity of US data centres by 20-35%, equivalent to removing 1-2 power plants every 3 years

- If successful, this would translate to eliminating greenhouse gas emissions from 1.5-2.7 million cars or 1-1.7 million homes in the US each year

A JOINT EFFORT

While most ES2 projects are led by individual universities, 'Models and Metrics for Dynamic Air and Hybrid Liquid Cooled Data Centers' involves input from Binghamton and Villanova universities, as well as UTA.

This collaborative effort aims to develop new cooling models for data centres using a combination of computational and experimental approaches. The end goal is to provide cooling on demand, as part of a holistically designed energy efficient data centre in which IT load allocation and temperature regulation systems are synergistically controlled. Such control can only be realised with accurate models of cooling system components, and rapid control strategies that allow cooling systems to respond to IT load, which can vary in place and time within a data centre.

Together, the groups will conduct experiments into the steady and dynamic performance of air and hybrid (air/liquid) cooling systems in an effort to validate computational models, allowing system-level studies. By assessing the energy efficiency of hybrid systems as compared to conventional systems, the researchers hope to develop new strategies to optimally control the temperature in data centres using less wasteful, local, on-demand cooling schemes.

A PLATFORM FOR COLLABORATION

ES2 is a truly unique endeavour in energy efficiency, providing benefits for companies and research institutions alike. It provides member companies a platform to conduct research that may not be possible within their own organisations, and particular support to small companies with limited research and development infrastructure. By bringing together universities with different areas of expertise and companies with different activities, the centre is identifying intelligent new ways to monitor electronic systems and regulate the amount of energy they use.

The collaborative environment of ES2 allows its members to address energy efficiency in a holistic way, looking at problems from various angles and incorporating the viewpoints of many different disciplines. Combining their strengths, the ES2 team will find answers to some of the industry's biggest problems. Its energy-efficient solutions could save the data centre industry millions of dollars, but the lessons learnt could also lead to practical solutions for all areas of the electronics industry.

DR BAHGAT SAMMAKIA, distinguished State University of New York Professor of Mechanical Engineering, is the Vice President for Research at Binghamton University and the Center Director of the NSF Industry/University Cooperative Research Center for Energy-Smart Electronic Systems.

DR KANAD GHOSE is Professor and Chair in the Department of Computer Science at Binghamton University and Binghamton University ES2 Site Director.

DR ALFONSO ORTEGA is the James R Birle Professor of Energy Technology and Associate Dean for Graduate Studies and Research at Villanova University. He is the Villanova ES2 Site Director and Director of the Laboratory for Advanced Thermal and Fluid Systems which he founded in 2005.

DR DEREJE AGONAFER is Professor and founding Director of the Electronics, MEMS & Nanoelectronics Systems Packaging Center at the University of Texas, Arlington and the UTA ES2 Site Director.

DR YOGENDRA JOSHI is Professor and John M McKenney and Warren D Shiver Distinguished Chair in the GW Woodruff School of Mechanical Engineering at the Georgia Institute of Technology, and ES2 Site Director at the same institution.