End-to-End Energy Management

Yung-Hsiang Lu, *Purdue University*
Qinru Qiu, *Syracuse University*
Ali R. Butt and Kirk W. Cameron, *Virginia Tech*

To improve energy efficiency, we must consider the end-to-end energy use of a task involving multiple computer systems.

The green IT community has been aflutter in recent months, with numerous workshops, summits, and meetings devoted to identifying and meeting the challenges of sustainable computing. Achieving consensus among such a growing, vibrant group of researchers is a challenge in itself. However, the stakes are high, with limited funding opportunities available.

An analysis of the literature in IEEE Xplore highlights a glaring gap in our understanding of energy efficiency in computer systems. Since 2005, researchers around the world have published more than 20,000 papers on energy management. Many if not most of these articles discuss techniques to improve the energy efficiency of individual components or systems: processors, memory, wireless networks, laptops, supercomputers, datacenters, handheld devices, and so on.

In reality, energy is task- not system-centric. Several disparate systems, or systems of systems, collectively use energy to accomplish a given task and satisfy service-level expectations. Consider, for example, someone who takes a photo with a smartphone and sends it to a friend. Taking and transmitting the photo consumes energy from the smartphone, the data transfer consumes energy from the Internet routers, and the recipient's local system consumes energy to display the image.

To improve energy efficiency, we must consider the end-to-end energy use of a task involving multiple computer systems.

REDEFINING ENERGY EFFICIENCY

When systems are interconnected, defining energy efficiency isn't straightforward. Saving energy on one computer can possibly cause another to consume more energy, increasing overall carbon emissions. In the case of file sharing, for example, should a computer compress a file before sending it? Although compression reduces data size and thus saves network energy, the processor's energy consumption could outweigh this savings. The solution depends on many factors including the compression ratio, network data rate, and number of hubs between sender and receiver.

The systems in use impact these energy efficiency calculations. For a smartphone, the wireless network is a significant energy consumer. In contrast, a desktop's processor, memory, display, and hard drive together consume much more energy than the network. Moreover, if the desktop has a wired broadband connection, compression could result in negligible network energy savings.

Cloud computing further complicates energy management. For mobile users, battery life is the leading factor in customer satisfaction, and some studies suggest that offloading computation to cloud servers could extend battery lifetime (M. Satyanarayanan, "Mobile Computing: The Next Decade," *Proc. 1st ACM Workshop Mobile Cloud Computing and Services* [MCS 10], ACM, 2010, article no. 5; E. Cuervo et al., "MAUI: Making Smartphones Last Longer with Code Offload," *Proc. 8th Int'l Conf. Mobile Systems, Applications, and Services* [MobiSys 10], ACM, 2010, pp. 49-62; K. Kumar and Y-H. Lu, "Cloud Computing for Mobile Users: Can Offloading Computation Save Energy?," *Computer*, Apr. 2010, pp. 51-56).

However, this approach gives little or no consideration to network or server energy use. Extending smartphone battery life could come at the cost of more carbon emissions from the network and servers supporting the offloaded computation.

For all these reasons, it's necessary to redefine energy efficiency. In high-performance systems, energy efficiency is often measured as gigaflops per watt—for example, the Green500 supercomputers are ranked on this basis. However, this metric doesn't capture many salient energy-efficiency aspects of connected systems, including mobile devices, access points, network backbones, and servers. The basic definition of energy efficiency is still the ratio between the amount of completed work and the energy used, but the numerator and denominator have become much more complex.

From an environmental impact perspective, any energy dissipation related to computation should be counted. Definitions of energy use should thus include the charging and discharging of the battery of mobile devices, the energy for cooling servers, and the energy lost in power distribution systems. New definitions of the amount of work performed should include meaningful examples such as the number of minutes videos can be watched on a mobile phone.

END-TO-END ENERGY MANAGEMENT

New energy-efficiency techniques must reflect the reality that computer systems no longer exist in isolation. *End-to-end energy management* considers the effects of energy management on the computers that comprise a system of systems.

Many researchers estimate the energy dissipation and carbon footprint of different tasks, but these estimates are based on empirical data, and validation is nearly impossible because no single entity owns the entire system of systems. Even when large companies own many system pieces, energy-efficiency details are often trade secrets. Consequently, existing estimations aren't precise enough to validate energy-efficiency management techniques targeting systems of systems (K.W. Cameron, "The Challenges of Energy-Proportional Computing," *Computer*, May 2010, pp. 82-83).

To achieve end-to-end energy management, academia and industry must recognize that true energy efficiency can be accomplished only by considering the complex interactions within systems of systems. Given the scale of these challenges, researchers and engineers must work together to develop new benchmarks, metrics, models, and measurement techniques.

True energy efficiency can be accomplished only by considering the complex interactions within systems of systems.

Benchmarks

Benchmarks need to incorporate networks and heterogeneous computers as intrinsic elements. In addition to throughput-centric workloads, system optimizations must consider the myriad effects of energy management—for example, whether slight performance degradation that saves significant energy in a server would cause unacceptable disruptions to mobile users.

Metrics

Current metrics such as Gflops/W or energy-delay products are insufficient. In fact, any calculation that summarizes complex interactions in a single number is probably undesirable. Metrics should acknowledge that energy consumption and other factors can affect one another—for example, that there is a tradeoff between energy and reliability (G. Wang, A.R. Butt, and C. Gniady, "On the Impact of Disk Scrubbing on Energy Savings," *Proc. 2008 Workshop Power Aware Computing and Systems* [HotPower 08], Usenix, 2008, pp. article no. 16). Energy-reduction solutions should also adapt to users' preferences: a user wants to maximize battery lifetime on a road trip but performance at the office.

Models

Researchers should investigate new techniques to model how changes in energy use at one location affect the performance and energy dissipation of the rest of the system. Such models could include a map profile of energy distribution over the network. As the application environment, network traffic, and server workload change, the causal relationships and their impact also vary; hence, the models become time-variant. This makes model development particularly challenging. Multi-agent-based distributed techniques should be considered to improve the scalability and reduce the complexity of the models (Y. Ge, Q. Wu, and Q. Qiu, "A Multi-Agent Framework for Thermal Aware Task Migration in Many-Core Systems," *IEEE Trans. VLSI Systems*, Aug. 2011; doi: 10.1109/TVLSI.2011.2162348).

Measurement techniques

New measurement techniques are urgently needed to support scientific evaluation of end-to-end energy management (K. Kant, "Toward a Science of Power Management," *Computer,* Sept. 2009, pp. 99-101). Building such facilities requires sophisticated equipment and expertise in many fields.

In the simplest sense, performance can be measured with a stopwatch. In contrast, energy consumption must be measured using equipment that might significantly alter a system's very nature—a mobile phone connected to a power meter is no longer mobile. Furthermore, the lack of "energy counters" in chips and circuit boards means that measurement must be intrusive.

End-to-end energy management is a new frontier ripe with opportunities for eager researchers. Because we currently lack the infrastructure to study energy efficiency in systems of systems, most end-to-end techniques must rely on disparate empirical data. In addition, we have neither accepted metrics nor accepted models of performance and energy efficiency for interconnected systems.

We've been in this situation before. As we entered the current millennium, there was a dearth of infrastructure and tools to enable energy management in isolated systems. Yet, in the past six years alone, research has led to vast improvements in mobile phone battery life and capability as well as mobile laptops that outperform workstations from just a few years ago. And as before, the necessity of addressing the challenges of end-to-end energy management is likely to drive innovation in the coming years to ensure that we meet our sustainable computing goals. ■

Yung-Hsiang Lu *is an associate professor in the School of Electrical and Computer Engineering at Purdue University. Contact him at yunglu@ecn.purdue.edu.*

Qinru Qiu *is an associate professor in the L.C. Smith College of Engineering and Computer Science at Syracuse University. Contact her at qiqiu@syr.edu.*

Ali R. Butt *is an assistant professor in the Department of Computer Science at Virginia Tech. Contact him at butta@cs.vt.edu.*

Kirk W. Cameron*, Green IT editor, is an associate professor in the Department of Computer Science at Virginia Tech. Contact him at greenit@computer.org.*