

Position Summary: Energy Management for Server Clusters

Jeff Chase and Ron Doyle

Department of Computer Science, Duke University

{chase, doyle}@cs.duke.edu

The Internet service infrastructure is a major energy consumer, and its energy demands are growing rapidly. For example, analysts project that 50 million square feet of data center capacity will come on line for third-party hosting services in the US by 2005. These facilities have typical power densities of 100 watts per square foot for servers, storage, switches, and cooling. These new centers could require 40 TWh *per year* to run 24x7, costing \$4B per year at \$100 per MWh; price peaks of \$500 per MWh are now common on the California spot market. Generating this electricity would release about 25M tons of new CO_2 annually.

The central point of this position paper is that energy should be viewed as an important element of resource management for Web sites, hosting centers, and other Internet server clusters. In particular, we are developing a system to manage server resources so that cluster power demand scales with request throughput. This can yield significant energy savings because server clusters are sized for peak load, while traces show that traffic varies by factors of 3-6 or more through any day or week, with average load often less than 50% of peak. We propose *energy-conscious service provisioning*, in which the system continuously monitors load and adaptively provisions server capacity. This promises both economic and environmental benefits.

Server energy management adds a new dimension to *power-aware resource management* [1], which views power as a first-class OS resource. Previous research on power management (surveyed in [1]) focuses on mobile systems, which are battery-constrained. We apply similar concepts and goals to Internet server clusters. In this context, energy-conscious policies are motivated by cost and the need to tolerate supply disruptions or cooling failures.

Our approach emphasizes energy management in the *network OS*, which configures cluster components and coordinates their interactions. This complements and leverages industry initiatives on power management for servers. Individual nodes export interfaces to monitor status and initiate power transitions; the resource manager uses these mechanisms to estimate global service load and react to observed changes in load, energy supply, or energy cost. For example, under light load it is most efficient to use server

power management (e.g., ACPI) to step some servers to low-power states. The servers may be reactivated from the network using Wake-On-LAN, in which network cards listen for special wake packets in their low-power state.

Our premise is that servers are an appropriate granularity for power management in clusters. Although servers consume less energy under light load, all servers we measured draw 60% or more of their peak power even when idle. Simply “hibernating” idle servers provides adequate control over on-power capacity in large clusters, and it is a simple alternative to techniques (e.g., voltage scaling) that reduce server power demand under light load. Since load shifts occur on the scale of hours, power transitions are not frequent enough to increase long-term hardware failure rates.

Dynamic request redirection provides a mechanism to allow changes to the set of active servers. Our system is based on reconfigurable switches that route request traffic toward the active servers and away from inactive servers. This capability extends the redirecting server switches (L4 or L7 switches) used in large-scale Web sites today. It enables the system to concentrate request traffic on a subset of servers running at higher utilizations.

Like other schemes for dynamic power management, energy-conscious service provisioning may trade off service quality for energy savings. Servers handle more requests per unit of energy at higher utilizations, but latency increases as they approach saturation. This fundamental tradeoff leads to several important research challenges. For example, it motivates load estimation and feedback mechanisms to dynamically assess the impact of resource allotments on service quality, and a richer framework for Service Level Agreements (SLAs) to specify tradeoffs of service quality and cost. This would enable data centers to degrade service intelligently when available energy is constrained.

References

- [1] A. Vahdat, A. R. Lebeck, and C. S. Ellis. Every joule is precious: The case for revisiting operating system design for energy efficiency. In *Proceedings of the 9th ACM SIGOPS European Workshop*, September 2000.