



GreenLight Technical Report

2009

Virtualization

Amin Vahdat: Conserving Resources through Virtualization

Research Activities

We have been active in investigating the network requirements for emerging data center architectures as well as virtualization techniques to increase the level of multiplexing and energy efficiency available from emerging container-based compute infrastructures. On the first front, we are considering GreenLight as an exemplar of emerging compute and storage containers supporting 1000-4000 individual processors and multiple petabytes of storage. This computation will be striped across multiple thousands of discrete machines each capable of delivering 10 Gb/s or more of network bandwidth across emerging 10 Gb/s switches. The ability to leverage all available computing and storage in these containers will hinge upon the ability to deliver low-latency and high bandwidth communication between individual elements. Our goal is to build a modular multi-stage network switch capable of delivering full bisection bandwidth (e.g., 20+ Tb/s in aggregate across 2,000 individual cores). We are in the process of designing a switch built around a fat-tree topology and emerging high-port count 10 Gb/s Ethernet switch silicon as the basis for GreenLight networking.

On the virtualization front, we recently completed our implementation and evaluation of Difference Engine, to mitigate perhaps the largest barrier to higher degrees of multiplexing in data center deployment environments. In current deployments, CPU and network utilization on a per-virtual machine basis is highly bursty. This allows data center engineers to leverage statistical multiplexing to more efficiently utilize available energy, compute, and network resources in the data center. Where previously machines that might have average utilization of 10% but bursty utilization to 80-100% had to have their own dedicated physical machines, virtualization allows for higher degrees of consolidation by assuming that periods of high activity will typically not be correlated and that when they are, VM migration can be leveraged to ease hot spots. However, the total amount of memory available on a physical machine limits the maximum degree of multiplexing. With current virtualization technology, each VM must be allotted its full amount of physical memory regardless of its dynamically changing behavior. Our work on Difference Engine leverages the observation that many physical memory pages are shared across VM's running similar operating systems and applications. Difference Engine dynamically and efficiently locates these pages and collapses them to a smaller physical memory footprint on individual physical machines. Difference Engine takes this process one step further by efficiently finding similar but not quite identical pages and collapsing them to base pages plus a compact representation of the difference. We perform all of this work independent of the operating system or the application, enabling us to transparently increase the available degree of multiplexing by a factor of 2-3 for a range of workloads.

For further information please contact Amin Vahdat, mvahdat@ucsd.edu