# The Greening of Datacenters

**Mazin Yousif, PhD**
**Avirtec Corporation**
**mazin@avirtec.net**

Green has become the buzzword of the industry for few years; you hear about green datacenters, green enterprise computing, green grids, green servers, green initiatives and green everything. It has become so contagious that everyone tries to out-green their competition and label their approach greener than the others. Good, bad, overly done, need more concerted efforts or wide mobilization…? All these efforts are blissfully commended as our computing infrastructures have insatiable desires to consume as much power as offered – a model that is not sustainable as the rate at which we add computing resources far exceeds the available and planned power capacities. In fact, one survey by the Uptime Institute (http://uptimeinstitute.org/) showed that 42% of respondents claim they would run out of power capacity in 2-5 years. Of course, not to mention the associated real environmental ramifications such as greenhouse gas emission. The positive side of this is that we realized that we have a power consumption problem (many datacenters consume more power than the cities they are hosted in) and the computing community is working feverishly to funnel efforts to resolve it. Imagine a total electric bill to operate datacenters world-wide in the 10's of Billions of US Dollars.

So how should we go about and tackle such a monster? I believe what is needed is a combination of conservation and efficiency around operations, cooling, ventilation, and autonomic control. Such efforts will need to be coupled with Industry-wide collaboration through standard bodies to establish ways to develop energy-efficient technologies and the means to benchmark and audit them. Of course, easy said than done! But, we should all take part of this process and aim at optimizing performance-power consumption as best as technologically possible. Specifically, just to name few, I am talking about computers and chips manufacturers; software and tools developers; IT administrators; researchers and theoreticians; as well as cooling and power supplies manufacturers. For example, computers and chips manufacturers have recently started paying considerable attention to optimizing performance-per-watt – the achieved performance relative to the amount of power consumed.

One difficulty to greening datacenters is that their resources are not homogeneous. In other words, if we enumerate resources in datacenters, we see very diverse set of resources including, but not limited to, servers, Ethernet network switches and routers, firewalls, load balancers, storage devices such as disk drives, Network Attached Storage (NAS) filers and Storage Area Networks (SAN). Few of these resources have

seen a great deal of transformation to make them more energy-efficient. For example, servers incorporate very advanced power management techniques, especially those deployed in Microprocessors, which constitutes the major share of power consumption in servers. Power supplies, usually rank second in terms of power consumption, are notoriously inefficient, especially when running at low utilization. Memory has received much less attention, but new low-level power management techniques and DRAM ACPI (Advanced Configuration and Power Interface) states support as well as the use of Non-Volatile Memories (NVRAM) promise more optimized memory power management.

Datacenter communications resources constitute less to the overall datacenter power consumption than servers; simply due to their smaller number (relative to servers). For example, network resources such as switches and routers can consume a great deal of power, but they make up smaller percentage of the total datacenter power. That does not mean that we should pay less attention to them. All the techniques adopted to optimize performance-per-watt in servers could also be deployed in network resources. However, one approach that has seen profound discussions is the use of one unified fabric to carry all traffics types (networking, storage and inter-process communication), which could best optimize network power consumption, if deployed along with the appropriate power management techniques.

One major contributor to datacenters high-electric bill is air conditioning, which grows proportionally to the amount of heat generated by the datacenter resources. The rule of thumb that we use here is that one Watt spent on compute and communication requires one watt in cooling. Clearly, this amounts to considerable overhead, which is currently being minimized by architecturally placing equipment racks in ways that would lower the impact of hot air blowing among racks and improve the efficiency of cooling equipments. One last source of power consumption in datacenters is the loss due to electric power transmission, which could also be considerable depending on electric wiring and the layout of the electric circuits. We also quantify that to one Watt for every Watt spent on compute and communication.

Since it is imperative to benchmark the efficiency of power management in datacenters, two metrics have been proposed by the Green Grid (www.thegreengrid.org): *Power Usage Effectiveness (PUE)* and *Data Center Infrastructure Efficiency (DCiE)*. The purpose of the metrics is to compare the amount of electricity the data center consumes for power and cooling with the amount of power used by the data center's compute and communication equipments through a simple-to-understand metric like the Kilometer-Per-Gallon metric used for automobiles. Of course, your goal is to bring PUE as close as possible to 1 – meaning the datacenter’s power is used mainly on compute and communication platforms. The farther away from 1 PUE is, the less efficient the

datacenter energy policies are. For the majority of datacenters, PUE is on average around 2.5, which means there remains a great deal of work for datacenters operations to be optimized. DCiE, on the other hand, is a percentage reciprocal of PUE. It refers to the percentage of power consumed by compute and communication resources relative to the total power consumed by the datacenter. Of course, the higher the percentage, the better! The difficulty with these two metrics is they are not easy to quantify and could generate a great deal of controversial discussions. That said, proposing such metrics is an admirable step in the right direction as it will give a chance to datacenter owners to compare the efficiency of their datacenters with others.

In conclusion, the community has realized that datacenters have a power consumption problem that is impacting the bottom line – Total Cost of Ownership (TCO) – of the datacenter. Looking at the bright size, we have promising technologies such as autonomic power management, virtualization, manageability and effective monitoring that when adopted intelligently could effectively optimize performance-per-watt in datacenters. Here are few words on couple such technologies. Autonomic management borrows from techniques adopted by biological organisms to optimize their operation and life activities based on environmental and runtime conditions through policies-enforcements such as performance-power-specific policies. Virtualization promises to improve the utilization of resources, provide better runtime management and optimize power consumption through dynamic resource allocation based.