Dynamic Power Variations in Data Centers and Network Rooms

By Jim Spitaels

White Paper #43

APC
Legendary Reliability®

Revision 2

Executive Summary

The power requirement required by data centers and network rooms varies on a minute by minute basis depending on the computational load. This magnitude of this variation has grown and continues to grow dramatically with the deployment of power management technologies in servers and communication equipment. This variation gives rise to new problems relating to availability and management.

Introduction

Data centers and network rooms draw a total electrical power, which is the sum of the power consumed by the installed Information Technology equipment. Historically, this equipment consumed power at a value that varied only slightly depending on the computational load or the mode of operation.

The notebook computer created the requirement that processor power be managed to lengthen battery run time. Power Management technology enabled the power consumption of laptop computer processors to be reduced by up to 90% when lightly loaded. As this technology has matured it has begun to migrate into server design. The result is that newly developed servers can have a power consumption that varies dramatically with workload over time.

When power varies with time, a variety of new problems occur for the design and management of data centers and network rooms. A few years ago, this problem was negligible. The problem has now reached a point where it is significant and the magnitude of the problem is growing.

Fluctuations in power consumption can lead to unplanned and undesirable consequences in the data center and network room environment; including tripped circuit breakers, overheating, and loss of redundancy in redundant power systems. This situation creates new challenges for people designing or operating data centers and network rooms.

Magnitude of Dynamic Power Variation

Through the 1990's, almost all servers drew a nearly consistent amount of power. The primary drivers of power variation in servers were related to disk drives' spin-up and speed changes in temperature controlled fans. The computational load placed on processors and memory subsystems caused a negligible variation in overall power consumption. On a typical small business or enterprise servers, the total power variation was on the order of 5% and was almost independent of the computational load.

Significant reductions in power consumption require cooperation between the BIOS, chipset, processor, and the operating system. In such a power managed system, whenever the processors are at less than 100% utilization, the operating system will execute an idle thread which will cause the processors to enter a low power state. The amount of time spent in the low power state is inversely proportional to the computational load on the system (e.g. a processor that is operating at 20% CPU utilization will spend 80% of it's time in a low power state).

The techniques used to achieve low power states vary among vendors and processor families but the most common techniques involve reducing or stopping clocks and removing or reducing voltages applied to various parts of the processor, chipset and memory.

Recently, processor vendors have introduced techniques to conserve power while the CPU is actively performing work. These methods involve changing the frequency of the clocks and the magnitude of the voltages applied to the processors to better match the workload applied to the processor in the non-idle state.

It is important to note that any technique which conditionally reduces processor power only reduces the average power consumed by the system; the maximum power remains unchanged and is trending higher with each new generation of CPUs. It is also important to recognize that when the processor power becomes a larger fraction of the total power consumption of the server, the variations in total server power consumption due to computational loading become correspondingly larger on a percentage basis. Multi-processor servers and those with very few disk drives (e.g. blade servers), will therefore have the highest percentage dynamic power variation.

The actual measurements of some servers are shown in **Table 1**. This shows the variation in AC power measured when different computational loads were placed on the computer.

***Table 1** – Dynamic power variation of actual servers*

Platform	Processor	Light load power draw	Heavy load power draw	Percentage variation
Dell PowerEdge 1150	Dual Pentium III - 1000	110 W	160 W	45%
Intel Whitebox	Pentium 4 - 2000	69 W	142 W	106%
IBM BladeCenter HS20 Full chassis – 14 blades	Dual Xeon 3.4 GHz	2.16 kW	4.05 kW	88%
HP BladeSystem BL20pG2 Full chassis – 8 blades	Dual Xeon 3.06 GHz	1.55 kW	2.77 kW	79%

Problems Associated with Dynamic Power Variation

Dynamic power variation gives rise to the following new types of problems:

Branch circuit overload

Most servers spend much of their time operating at light computational loads. For servers with power management, this means that the server will be drawing less than its potential power draw. Most people installing or maintaining data centers and network rooms, however, are unaware that the typical observed server power consumption may be much lower than the potential power consumption when under a high computational load. This situation can lead a data center or network room operator or IT staff to accidentally put too many servers onto a branch circuit.

When the sum of the maximum power consumptions of the servers on a branch circuit exceeds the branch circuit rating, the potential for overload is present. Under this condition, a group of servers will operate successfully until the condition occurs where enough of the servers are simultaneously subject to heavy loading. The computing conditions that result in such an overload may occur very infrequently, so the system may operate successfully without failure for weeks or even months.

During an overload condition caused by the situation described above, the branch circuit will operate at a higher current than the circuit rating. In the data center or network room environment, the most significant consequence of this situation is the branch circuit breaker may trip and terminate the power fed to the computing equipment. This is obviously an extremely undesirable event. Furthermore, since it is happening at a time of high computational load, it is likely that the computing equipment is handling a large number of transactions so the failure is very likely to be occurring at a particularly undesirable time.

Overheating

In the data center and network room, all of the electrical power consumed by computing equipment is dissipated as heat (one exception to this PoE switches that send a significant fraction of their power down the Ethernet cables VOIP telephones, Wi-Fi access points and other powered devices). When the power consumption of computing equipment varies due to computational load, the heat output also varies. If equipment in one part of the data center suddenly increases its power consumption, this can create a local hot-spot condition in the data center. The data center cooling system may have been balanced using typical power dissipation and a doubling of power in a local area may result in an undesirable temperature rise that the cooling system is not designed for. This may cause equipment to shut down on over temperature, cause the equipment to act abnormally, or it may void equipment warrantees.

Loss of redundancy

Many servers have dual redundant power inputs, and most high availability data centers and network rooms take advantage of this feature to provide dual power path feeds to the server. These systems can survive a complete failure of any point in either power path and continue operation. During normal operation, the computers are designed so that both power paths equally share the load.

When there is a failure in one power path, the full load of the server is transferred to the remaining power feed. This causes the load on that power feed to double. For this reason, the AC mains branch circuits feeding the equipment in a dual path system must always be loaded to less than 50% of rated ampacity, so they have sufficient remaining capacity to take over the complete load if necessary.

Ensuring that a branch circuit is loaded to less than 50% of its rating is a task made more difficult when the loads exhibit dynamic power consumption. A system may be tested upon installation and be found to have branch circuits operating safely below 50% of its rating, and then at some future time of high computational load the system may begin operating at greater than 50% of its rating.

If a branch circuit in a dual path system enters the condition where the load is over 50% of its capacity, then the redundancy of the system is lost. If one feed were to fail, the second feed would then be immediately overloaded and its breaker would be likely to trip as described in the previous section. Again, since this is happening at a time of high computational load, it is likely that the computing equipment is handling a large number of transactions so the loss of redundancy is very likely to be occurring at a particularly undesirable time.

Masking of the problem

The equipment that exhibits dynamic power consumption may represent only a small fraction of the total power consumption of a data center or network room. If 5% of the equipment in a data center has a dynamic power variation of 2-to-1 and the rest of the equipment draws constant power, then bulk power measurements of the data center at the main power feed or at a Power Distribution Unit might only vary 2.5%. This might cause an operator to believe that no significant dynamic power variation problem is occurring when in reality a significant risk of breaker tripping, overheating, or loss of redundancy may be occurring. Therefore, there is a very real possibility that the problem may exist yet be unrecognized by experienced operators.

Managing dynamic power variation

To mitigate against the problems described in the previous sections, designers and managers of data centers and network rooms must adapt to the new realities of dynamic power consumption. There are a number of means that can be used to accomplish this, and some are reviewed as follows:

Separate branch circuit for each server

If a separate branch circuit is provided to each server, then branch circuit overload cannot occur. This is true because every server is assured to operate from a dedicated branch circuit by design. This solves the issue of branch circuit overload and solves the problem of loss of redundancy. It does not solve the thermal problems, but these are typically not the largest risk. However, this is a very complex and expensive solution where small servers are deployed such as 1U or 2U servers since this could require an extremely high number of branch circuits per rack. In the extreme case, a rack filled with dual corded 1U servers could require 84 branch circuits, which corresponds to two large circuit breaker panel boards. This solution is more practical when larger servers or blade servers are used.

Establish safety margin standards for worst case and measure compliance at install

Most data center and network room operators have standards for loading margins, which are typically expressed as a fraction of the full load branch circuit rating. Typical values chosen are between 60% and 80% of the branch rating, with values of 75% being considered a reasonable tradeoff between power capacity, cost, and availability. To verify compliance with the standard actual branch circuit loads are measured to ensure compliance with the standard. Note that there is a serious problem with this approach when the systems exhibit dynamically varying power consumption because it may be difficult to know the computational load at the time of measurement. Ideally, a heavy computational load would be placed on the protected equipment during the measurement to ensure compliance at worst case.

Establish safety margin standards for worst case and calculate compliance

In another case, detailed inventories of exactly what equipment is connected to each branch circuit are kept, and the maximum published or measured load drawn by the equipment is kept and summed to ensure that a particular branch circuit is not overloaded. Information regarding maximum load for various equipment is available from the individual equipment manufacturer (where the load is often considerably overstated) or from UPS selector applications such as those found on www.apcc.com. Keeping detailed branch circuit inventories is a common practice in large high availability data centers. However, this requires that the operator know exactly what is plugged into every branch circuit at all times. For most network rooms and smaller data centers there is insufficient control of users to ensure that equipment is not moved, or exchanged, or simply plugged into a different outlet. Therefore this approach is not practical in many installations.

These margins can be further reduced to provide for dynamic power increases. For example, the safety margin specification can be that the measured branch load cannot be beyond 35% of branch circuit rating when the equipment is operating in an idle condition.

Establish safety margin standards for worst case and monitor compliance ongoing

In this case safety margins are established and all branch circuits are continuously monitored on an ongoing basis by an automatic monitoring system. Warnings are sent out when branch loading begins to enter the safety margin area. For example, when using a 60% branch-loading standard, send alerts when the loading passes 60%. The safety margin is established such that the operators would have significant advance warning of a problem area and could take corrective action before an over current condition occurs. This method can be used in conjunction with the other methods described previously. The great advantage of this method is that it works in situations where users are likely to install or move equipment or plug it into a different outlet without the knowledge of the data center manager; a situation is very common in network rooms, collocation facilities, and medium security data centers. This approach can also warn on impending loss of redundancy. This is the most powerful tool that the data center manager can use to manage dynamic power variations in an ever-changing environment.

Conclusion

The percentage of Information Technology loads in the network room or data center, which exhibit a power consumption that varies significantly with load, is increasing over time. This situation gives rise to a number of unanticipated problems for operators of data center infrastructure. The procedures historically used to minimize the risk of overload must adapt to this new reality. Proper planning and branch circuit power monitoring are critical for ensuring availability in both new and existing facilities where large numbers of servers will be installed.

About the Author:

Jim Spitaels is a Consulting Engineer for APC. He has Bachelors and Masters Degrees in Electrical Engineering from Worcester Polytechnic Institute. During his 14 years with APC he has developed UPSs, communications products, architectures and protocols, equipment enclosures, power distribution products and he has managed multiple product development teams. Jim also holds 3 US Patents related to UPSs and power systems.