データセンタとサーバルームの動的な電力変動

ジム・スピタエルス

White Paper #43

APC
Legendary Reliability™

要約

データセンタおよびサーバルームに必要な電力条件は、コンピュータの処理量に応じて刻々と変化します。この変化の量はサーバおよび通信機器への電力管理技術の配備により劇的に増加し続けています。このような変化は、可用性と管理に関連する新たな問題を生み出しています。

はじめに

データセンタとサーバルームでは、設置された全てのIT機器で消費される電力の合計量が必要とされます。従来、このような機器の電力消費量は、コンピュータの処理量または動作モードに応じて多少変化するにすぎませんでした。

ラップトップコンピュータの場合、バッテリの駆動時間を延ばすために、プロセッサの電力を管理する必要がありました。このためプロセッサの電力消費量は電力管理技術により、負荷の低い場合で最大90%低減することができるようになりました。そしてこの技術が成熟すると、サーバ設計への適用が始まりました。その結果、新しく開発されたサーバの電力消費量は、作業負荷により大きく変化するようになりました。

電力が時間とともに変化する場合、データセンタおよびサーバルームの設計と管理に各種の新たな問題が発生します。この問題は、数年前には無視できる程度のものでしたが、現在では重要なレベルに到達しており、問題の規模は拡大し続けています。

電力消費量の変動は、データセンタとサーバルーム環境に、サーキットブレーカのトリップ、オーバーヒート、冗長電源システムの冗長性の損失といった望ましくない結果を引き起こすことがあります。このような状況により、データセンタとサーバルームの設計または運用に携わる人々に新たな課題が生じています。

動的な電力変動量

1990年代を通して、ほぼすべてのサーバがほとんど同じ量の電力を消費していました。サーバの電力変動の主な原因は、ディスクドライブのスピンアップと温度制御付きファンの速度変化にあり、プロセッサとメモリサブシステム上の処理量による電力消費量の変動は、全体から見るとわずかでした。通常の小規模ビジネス用、すなわちエンタープライズサーバでは、電力の総変動の大きさは5%であり、計算負荷はほとんど無関係でした。

電力消費量を大幅に削減するには、BIOS、チップセット、プロセッサ、オペレーティングシステム間の連携が欠かせません。このような電力管理下にあるシステムでは、プロセッサの稼動率が100%未満になると、常にオペレーティングシステムでアイドルスレッドが実行され、プロセッサは低電力状態に移行します。低電力状態に費やされる時間は、システム上の計算負荷に反比例します（たとえば、CPU稼動率20%で動作するプロセッサの場合、総時間の80%は低電力状態です）。

低電力状態の達成に用いられる技術は、ベンダーおよびプロセッサファミリにより異なりますが、最も一般的な技術としてクロック数の低減または停止、プロセッサ、チップセット、メモリの異なる部分に印加される電圧の停止または低下させることがあげられます。

近年、プロセッサベンダーでは、CPUが高稼動率で動作する間に電力を節減する技術を導入しています。この方法は、クロックの周波数とプロセッサに流れる電圧の大きさとを変化させ、非アイドル状態のプロセッサに加わる作業負荷に適合するよう調整をはかるものです。

プロセッサの電力を条件付きで低減する技術は、いずれもシステムで消費される平均電力を低減するにすぎないことに注意する必要があります。最大電力量は変化せず、CPUの世代交代ごとに増加する傾向にあります。また、サーバの総電力消費量に占めるプロセッサの電力量の割合が高くなると、処理量によるサーバの総電力消費量の変動は高い率で増加することにも注意する必要があります。したがって、マルチプロセッササーバとディスクドライブ数が非常に少ないサーバ（ブレードサーバなど）は、動的な電力変動率が最大になります。

いくつかのサーバで実際に測定された値を**表1**に示します。ここでは、コンピュータにさまざまな処理量を実施させた場合の、AC電源の測定値の変動を示しています。

表1 - 実際のサーバの動的電力変動

プラットフォーム	プロセッサ	軽負荷時 電力消費	重負荷時 電力消費	変動率
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 フルシャーシ - 14ブレード	Dual Xeon 3.4 GHz	2.16 kW	4.05 kW	88%
HP BladeSystem BL20pG2 フルシャーシ - 8ブレード	Dual Xeon 3.06 GHz	1.55 kW	2.77 kW	79%

動的電力変動にともなう問題

動的な電力変動により、次のような新しい種類の問題が発生します。

分岐回路の過負荷

ほとんどのサーバは、その時間の大半を軽い処理量での動作に費やします。電力管理機能を持つサーバの場合、これはサーバが最大電力消費量よりも少ない電力を消費することを意味します。しかし、データセンタおよびサーバルームの設備または保守を担当する職員のほとんどは、通常測定されるサーバの電力消費が、高い処理量の下での最大電力消費よりも相当低くなることを認識していません。このような状況では、データセンタまたはサーバルームのオペレータやITスタッフは分岐回路に容量以上のサーバを接続してしまいます。

分岐回路上の各サーバの最大電力消費量の合計が分岐回路の定格を超える場合、過負荷が生じる可能性があります。この条件下では、特定のサーバグループは、十分な数のサーバが同時に重負荷の状態になるまで正常に動作します。過負荷に至るような計算条件が発生するのは非常にまれであるため、システムは数週間、あるいは数か月の間、故障を起こさずに正常に動作します。

上記に述べた状況により過負荷の状況にある間は、分岐回路は回路の定格よりも大きな電力で稼動します。データセンタまたはサーバルーム環境では、このような状況が発生した場合、まず分岐回路のブレーカのトリップと、計算機器への給電の停止が起こります。いずれも、きわめて好ましくない現象であることは明らかです。さらに、処理量が大きくなった時点でこのような結果が生じるため、計算機器が大量のトランザクションを処理することになり、特に重要なタイミングで故障が起こる確率が非常に高くなります。

オーバーヒート

データセンタとサーバルームでは、計算機器で消費される総電力消費量は熱として放出されます（電力の大部分をVoIP電話向けのイーサネットケーブル、Wi-Fiアクセスポイント、その他の給電機器に送るPoEスイッチは例外です）。計算機器の電力消費が処理量により変動する場合、機器の発熱量も変化します。データセンタのある区画の機器の電力消費が突然増加した場合、データセンタの局所的なホットスポット状態が生じます。データセンタの空調システムは通常の電力放散機能で熱均衡を実現している可能性があるため、電力の局所的な倍増により、空調システムの設計段階では考慮されていない温度上昇が発生することがあります。温度上昇により機器の停止や異常動作が発生したり、機器が故障したりする可能性があります。

冗長性の損失

多くのサーバには二系統の冗長電力入力が装備され、高可用性データセンタおよびサーバルームのほとんどは、この機能を活用してサーバへの二重の給電経路を確保しています。このようなシステムはどちらかの給電経路が完全に損傷しても、停止に至らず稼動し続けます。通常の動作では、コンピュータの設計上、2つの給電経路は半分ずつ負荷を分担します。

いずれかの給電経路に故障が発生した場合、サーバの負荷はすべてもう一方の給電経路に切換えられます。その結果、切換えられた先の給電負荷は2倍になります。この理由により、二系統給電システムで機器に給電するAC主電源の分岐回路は、常に定格電流容量の50%未満にしか流れないようにし、必要に応じて全負荷を供給するのに十分な容量を残しておく必要があります。

分岐回路への負荷量をその定格の50%未満に確実に抑える作業は、負荷により電力消費が変動を示す場合には困難なものになります。システムの設置後にテストを実施し、分岐回路が定格の50%未満で安全に動作することを確認した後、処理量が増大したある時点で、システムが定格の50%を上回る率で稼動を始める場合があります。

二系統給電システムの分岐回路が、負荷が容量の50%を超える状態に増大した場合、システムの冗長性は損なわれます。給電経路のいずれかに故障が発生すると、もう一方の給電経路はすぐに過負荷状態になり、上記の項で述べたようなブレーカのトリップが起こる可能性が生じます。この場合も、処理量が増大した時点でこのような結果が生じているため、計算機器が大量のトランザクションを処理し、特に重要なタイミングで冗長性が損失する確率が非常に高くなります。

問題の埋没

動的な電力消費を示す機器は、データセンタまたはサーバルームの総電力消費量に占める割合は低いかもしれません。例えばデータセンタの機器の5%が動的電力変動比2:1を示し、残りの機器が一定の電力を消費する場合には、データセンタの主電源または分電盤で測定される総電力量の変動は2.5%にすぎません。このためオペレータは、実際にはブレーカのトリップ、過負荷、冗長性の損失といった重大なリスクが存在する場合でも、動的電力消費変動に特に問題は生じていないと判断してしまうようになります。したがって、未熟なオペレータが問題を認識しないまま放置しておくという可能性が非常に現実味を帯びてきます。

動的な電力変動の管理

前項で述べた問題を軽減するために、データセンタおよびサーバルームの設計者と管理者は現実的な動的電力消費に対応する必要があります。この場合、採用できるいくつかの手段があり、その一部を以下に説明します。

分岐回路のサーバ間の分離

各サーバに個別の分岐回路を配置した場合、分岐回路の過負荷は生じません。どのサーバも、設計された専用の分岐回路を通じて動作することが保証されるためです。これによって、分岐回路の過負荷の問題が解消し、冗長性の損失の問題も解決されます。熱放出の問題は解決されませんが、熱放出は一般にそれほどリスクは高くありません。ただし、このソリューションでは1Uサーバ、2Uサーバなどの小型のサーバが配備され、ラックあたり非常に多くの分岐回路が必要になるため複雑で高価なものになります。極端な場合、二系統給電の1Uサーバを配置したラックには、2基の大型回路ブレーカパネルボードに対応する84の分岐回路が必要になります。このソリューションは、大型のサーバまたはブレードサーバを使用する場合には有効です。

最悪な条件下でのセーフティマージンの設定と設置時の測定

データセンタおよびサーバルームのオペレータの大半は、負荷量の誤差の基準を設定しています。これは通常、全分岐回路定格に対する率で表されます。通常の値は分岐定格の60%から80%の間から選択され、75%が電力容量、コスト、可用性の適切なトレードオフとして認められます。基準への準拠については、実際の分岐回路の負荷を測定して確認します。しかしこの方法には重大な問題があることに留意してください。システムが動的に変動する電力消費を示す場合、測定時に処理量を把握するのが困難なためです。理想的には、保護された機器の処理量を増大にして測定し、最悪な条件下でも動作可能なことを保証します。

最悪な条件下でのセーフティマージンの設定と計算

別のケースとして、各分岐回路に接続された機器に関する正確な明細を記録し、各機器が消費する最大定格負荷または最大測定負荷を維持し、合計して、特定の分岐回路に過負荷が起こらないようにする方法があります。各種機器の最大負荷に関する詳細は、個々の機器のメーカにお問い合わせください（負荷が実際より大きく記載されている場合が多い）。分岐回路の詳細な明細を記録する作業は、大規模な高可用性データセンタでは一般的に行われています。ただし、これにはオペレータが各分岐回路に接続される品目を、常に正確に把握する必要があります。大半のサーバルームおよび小規模なデータセンタでは、ユーザによる機器の移動や交換、あるいは別のコンセントへの差込を確認するのに十分な管理体制が整っていません。したがって、この方法は多くの設備で実用的ではありません。

最悪な条件下でのセーフティマージンの設定と稼動時の監視

セーフティマージンを設定し、自動監視システムにより、稼動時に連続的にすべての分岐回路を監視します。分岐の負荷がセーフティマージンを超えた場合、警告が発せられます。たとえば、分岐負荷の基準を60%とした場合、負荷率が60%を超えた時点でアラートを送信します。セーフティマージンは、問題が発生した場合の警告が事前にオペレータに知らされ、過電流状態に陥る前にオペレータが対処できるように設定します。この方法は、前述のその他の方法と併せて使用できます。この方法が非常に有利な点は、ユーザがデータセンタの管理者に通知せずに機器を設置または移動する、あるいは別のコンセントに差し込む可能性がある状況に適していることです。このような状況は、サーバルーム、コロケーション設備、中規模のデータセンタではごく一般的です。この方法は、冗長性の損失を防止する際も警告を発し、データセンタの管理者が、変化のある環境で動的な電力変動を管理する際に用いる最も強力なツールです。

結論

サーバルームまたはデータセンタのIT機器の負荷率は、負荷の変動により大きく変化する電力消費として表され、時間とともに増加します。このような状況において、データセンタのインフラのオペレータはいくつかの不測の問題に直面します。過負荷のリスクを抑えるために従来から用いられてきた手順は、この新しい問題に適応させる必要があります。新旧を問わず、多くの台数のサーバが設置される施設で可用性を維持するためには、正しい計画と分岐回路の電力の適切な監視を欠かすことができません。

著者について

ジム・スピタエルスはAPCのコンサルティングエンジニアです。ウースター工芸大学で電気工学の学士号および修士号を取得しています。APCでの14年間の在職中に、UPS、通信製品、アーキテクチャおよびプロトコル、機器のエンクロージャ、配電製品を開発し、複数の製品開発チームを管理しています。また、UPSと電源システムに関連して3件の米国特許を保有しています。