Understanding UPS Overload Capabilities in Data Centers

Many UPSs in data centers and other large, mission-critical facilities are oversized, relative to the size of the load they are protecting. While lightly loaded UPSs are less efficient than more fully loaded systems, system designers often opt for oversizing for increased reliability. The logic behind oversizing is that any UPS is likely to encounter temporary overload conditions due to load energization (inrush), downstream static switch operation, accidental faults, or unplanned addition of new equipment to the UPS output. In these conditions it is crucial that the UPS have enough capacity to protect the load. In addition, it is common for a 2N, or “dual bus,” system to be loaded less than 50% on each bus to ensure the load will be protected even in overload conditions.

Overload limits and duration of support
UPSs have specifications that define the amount of overload that can be supported and the time delay before transfer to bypass (or shutdown if bypass is not available). The overload levels are standardized at <103%, 125% and 150% of nominal output power. There is also a current limit level, often 200%, which if exceeded during an output fault or short circuit, will result in an immediate transfer to bypass. If the bypass source is out of limits, this current limit may cause a reduction in output voltage and trip the UPS on output AC under voltage.

For UPSs in today's market, the levels and ranges of support time are:

<table>
<thead>
<tr>
<th>Overload Level</th>
<th>Support Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>103% overload</td>
<td>10 minutes to continuous</td>
</tr>
<tr>
<td>125% overload</td>
<td>Between 30 sec and 10 minutes</td>
</tr>
<tr>
<td>150% overload</td>
<td>Between 10 sec and 60 sec</td>
</tr>
<tr>
<td>200% overload</td>
<td>10 to 20 cycles (current limit)</td>
</tr>
</tbody>
</table>

Since the above specifications contain a wide range of possible overload support times, engineers, users and designers must determine if a longer overload time limit provides useful protection for the load and if a longer overload support time is a measure of UPS robustness or quality of design.

In order to determine if a longer overload time limit is needed for adequate protection, a system designer should weigh options such as how 10 minutes of support vs. 30 seconds of support will benefit the user at 125%, or if at 150% overload 60 seconds of support vs. 10 seconds of support provide any additional protection for the critical load.

Clearly any overload that is sustained for 10 seconds is not due to a transient event, like an inrush, motor start or static switch energization. Even 10 seconds of continuous 125% to 150% overload strongly indicates that there is a problem in the critical load that is not just temporary. These problems fall into one of the following categories:

1) A fault (the most likely possibility). An example would be a tool dropped inside an equipment rack, or a server or other piece of IT gear that suffered a short circuit. In the event of any fault, the desired outcome is always the same: clear the fault (open the circuit breaker) as quickly as possible, hopefully before the fault brings the bus voltage low enough to shut off other connected critical equipment. In these events, the fault current increases rapidly, well beyond the 125% or 150% load level of the UPS, so it does not matter how long the UPS supports overloads at this level. The fault current either causes the UPS to enter the 200% current limit, immediately causing a transfer to bypass on output under voltage, or it exceeds the rating of the branch breaker, causing it to trip and effectively remove the fault. Either way, the UPS overload time limit is unimportant.
In a real world application, most data center loads are connected to a remote power panel (RPP) containing 20A or 30A circuit breakers. Referring to the trip curves for this type of breaker (see Figure 1), it is evident that seven to 20 times the “handle rating” of the breaker is required in order to trip it instantaneously (within one cycle). This instantaneous trip is critical to ensure that the fault does not bring down the rest of the connected equipment. So, to trip a 20A breaker quickly, it requires between 140A and 400A of available current from the UPS. In practical terms, this means that any UPS rated less than 80 kVA and loaded at 50% will not be able to trip the branch breaker; it cannot produce enough instantaneous current. This is why a transfer to bypass must occur immediately, not after an arbitrary time delay. If the fault occurs on a 30A branch breaker, any UPS less than 100 kVA would be unlikely to trip the breaker instantaneously. In fact, in the case of a low impedance fault the UPS will transfer immediately to bypass on output under voltage. Clearly, a longer overload time provides no value, no protection, and would actually delay the isolation of a high impedance fault.

2) Inadvertent addition of load: An example would be a technician plugging in or energizing a new set of servers, or multiple racks of new equipment, without realizing that this action would overload the UPS. In this case, depending on the amount of unintentionally added load, the UPS could reach its 150% overload point. In this situation, ANY UPS at 150% overload is operating near its design limit for both current and temperature. It is unwise to artificially extend this risky condition, by letting the UPS operate in this state for any amount of time.

The correct response by the UPS is to verify that the 150% overload is sustained (not just an inrush), and transfer to bypass as soon as possible. Any delay in transferring means unnecessary extension of the time the UPS spends under duress. This is not good for any UPS, because it extends over temperature and over current beyond the design intent, but more importantly, it adds unnecessary risk for the critical load.

3) Progressively failing load device: An example would be a server or other IT device that is experiencing a progressive degradation of its internal insulation (a failing filter capacitor or EMI device, for example). This type of failure may not draw a large current immediately, but gradually increases the fault current over hours to days. The first indication of the problem could be an overload alarm on the UPS. As in the above cases, the best response from the UPS would not be to simply “cook” in an overloaded condition. This overload is not going to just “go away,” so the UPS will always eventually transfer to bypass. The amount of overload time provided by the UPS gives no benefit in this case, the end result is the same: the load goes on bypass.
As the previous cases show, the correct action is to let the comparatively stronger utility bypass handle these sustained overloads. Utility power, with its inherently lower impedance, is more likely to successfully clear faults, and support overloads without reduction in bus voltage, than is any UPS. The available fault current from the utility is large, only limited by source impedance, and immediately available. Contrast this with any UPS where the output current is limited to only about 200% of nominal. (See Figures 2 and 3)

**Figure 2:** Utility power available to clear downstream fault.

**Figure 3:** UPS attempting to clear a downstream fault.

So what is the value of an overload timer in a UPS? It simply allows the UPS to verify that a sudden load increase is not due to a transient event like a motor starting or static switch activating in the data center. In these very brief events, we would, of course, want the UPS to remain online, without producing nuisance transfers to bypass. These transients are over in a few cycles, so a “longer” overload timer is unnecessary and adds risk.

These factors also help determine if a longer overload support time indicates a “stronger” UPS or more robust design. Analysis of various UPS vendors’ equipment shows that semiconductor ratings, fuse sizes and thermal capabilities like fans and heat sinks are often quite similar for the same rated products. It is rare to find a UPS with larger semiconductors or extra fans or oversized heat sinks because each of these dramatically affect cost, and could negatively impact the ability to price the product competitively. Using vastly over rated components in the UPS would have the side benefit of allowing a higher maximum operating temperature, that is, 50°C instead of the ubiquitous 40°C found in most UPS specifications today. It is significant that UPS vendors do not offer this wider temperature range. The engineers that design these machines maintain adamantly that if one requires, for example, a 100 kW UPS that will support 150% reliably, one must design a 150 kW UPS, with its inherent increase in cost. There is no free lunch and no safe shortcut; more power costs more money. Over stressing a UPS beyond its design intent is discouraged for any application, just as running a sports car beyond the redline is considered both risky and potentially expensive. There is just no good reason to do it.
Summary
In this discussion, we have stated the general specifications for UPS overload performance and noted that specifications vary in the length of time a UPS is expected to support overloads. The amount of tolerable overload is quantified as 150%, and the previous paragraphs describe why this UPS overload condition should be limited in time. The mission-critical nature of equipment and enterprises supported by power protection equipment demands that potentially damaging overloads be handled properly. The use of extended overload timers to demonstrate the ruggedness of the UPS can be misleading to the user at best, and at worst, detrimental to the optimal protection for the data center loads.

For more information on Eaton’s UPS and power quality products, please visit www.eaton.com/powerquality.