Liebert®

IT White Paper

THE ALWAYS-ON NETWORK: STRATEGIES FOR ACHIEVING HIGH AVAILABILITY OF IT SYSTEMS

EMERSON™
Network Power

Introduction

Managing the availability of mission critical systems requires an understanding of the risks and costs of losing access to business critical information or services balanced against the cost of achieving a certain level of availability.

According to the Uptime Institute, 25 percent of all information downtime results from the interaction of computer hardware with its physical environment.

That balance is shifting toward higher levels of availability as network services becomes essential to business continuity and the cost of downtime escalates. A February 2004 study by Infonetics Research revealed that interruptions in enterprise availability cost large companies an average of 3.6 percent of annual revenues.

As a result, IT managers face growing pressure to drive network availability to unprecedented levels. At the same time, Voice over Internet Protocol, radio frequency identification, just-in-time inventory, lean manufacturing and point-of-sale integration are placing new demands on networks – and the systems that support them.

This pressure is not limited to the data center. For an increasing number of organizations, the network itself is mission critical. It may not always be possible for remote systems to achieve the same levels of availability as those in the data center, but the gap can be closed by applying the strategies and technologies used in the data center to systems outside it.

According to the Uptime Institute, 25 percent of all information downtime results from the interaction of computer hardware with its physical environment.

Continuous availability of mission-critical systems rests not only upon flawless operation of the systems themselves, but on the infrastructure that supports those systems. Achieving “five nines” network availability requires installation and management of an infrastructure that supports continuous availability.

This requires four key components:

- Mission critical power
- Mission critical cooling
- Monitoring and management
- Proactive maintenance

Power Availability

IT systems have two connections to the outside world: a connection to electrical power and a connection to the network. Without power, information technology becomes worthless. Unfortunately, the modern power grid does not adequately support IT operations – and it appears to be getting worse.

Electricity demand has jumped 30 percent over the past 10 years, while transmission capacity has increased only half that amount. As seen in August 2003, a strain in one location can cause the whole system to buckle. That outage represented the fourth major failure of the central power grid within the last decade. Consequently, each organization must create a power infrastructure that can deliver the power quality and reliability IT systems require. Accomplishing this involves selecting the right equipment and system architecture.

UPS Technology

UPS systems can use either passive standby, line-interactive or double-conversion technology. When availability is a high priority, true online double-conversion technology is the clear choice. Double-conversion systems can tolerate long power fluctuations without transferring to battery power. They also completely isolate mission critical systems from the power source, while line-interactive UPSs can allow equipment-damaging power anomalies to pass through.

This is exactly what occurred at a Manhattan data center that used both line-interactive and double-conversion UPS systems before the August 2003 blackout. During the blackout, batteries in the line-interactive UPS units lacked the capacity to support equipment while backup generators started up. Many shut down before their rated backup time and dropped loads unexpectedly. Power supplies on seven servers connected to line-interactive UPS units were destroyed.

System Architecture

Power system architecture plays a major role in achieving appropriate — and cost effective — availability. When evaluating a power architecture, two questions should be kept in mind:

1. Have single points of failure been identified and eliminated everywhere it is cost-effective to do so?
2. Is this the simplest architecture to achieve the availability requirements of the organization?

Power system architectures can be classified into four tiers (see Figure 1). Moving up the tiers increases availability, but requires increased investment.

The first tier provides only protection against surges, therefore delivering the same level of availability as the utility source. The second tier adds a UPS, but without redundancy, thus requiring downtime for UPS service. Tiers three and four introduce redundancy to eliminate single points of failure and are used in high-availability data centers.

When availability is a high priority, true online double-conversion technology is the clear choice.

The highest availability is achieved in a tier four dual bus system that creates parallel power paths to eliminate single points of failure from the point power enters the room to point at which it is used.

The N + 1 architecture used in some tier three systems — in which N is the number of UPS units required to support the load and "+1" is an additional unit for redundancy — can be used to balance availability, cost and scalability; however, as modules are added to the system, complexity goes up and reliability goes down. In statistical analysis of N + 1 systems, 3 + 1 appears to be the threshold at which the risks to network availability outweigh the benefits of scalability.

Protecting Remote Applications

Outside the data center, many of the same principles and solutions apply. However, these environments typically don't have access to a backup generator, so UPS battery capacity determines how long systems can run in the event of an outage. Where 10 or 15 minutes of battery capacity is typical in the data center, this is usually not enough outside the data center, particularly as new applications like IP telephony are deployed. These systems often require as much as four hours of battery backup to ensure phones can continue to operate in the event of an outage.

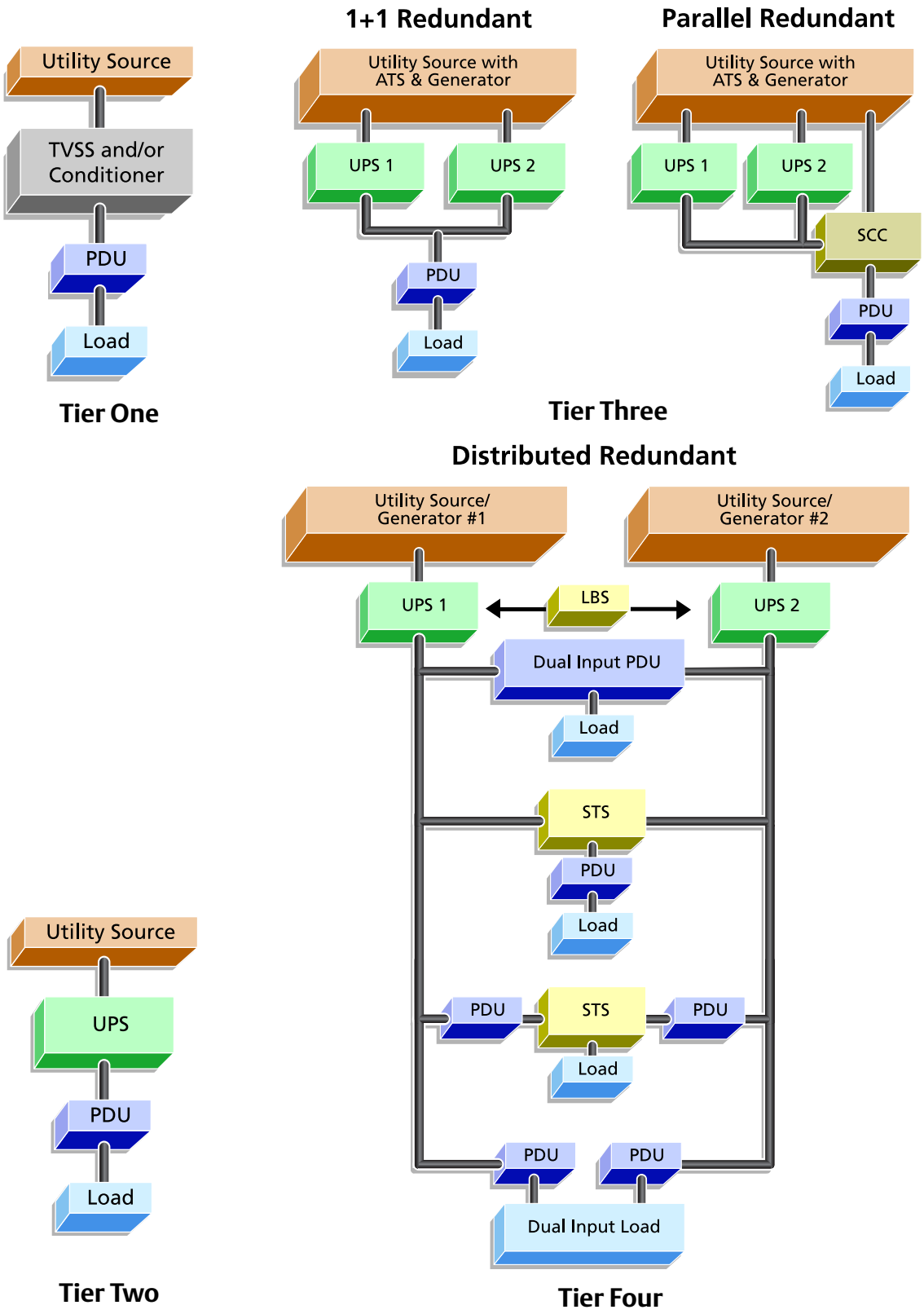

Figure 1. As you move from Tier One to Tier Four higher levels of availability are supported.

In terms of topology, systems outside the data center are becoming more powerful and more critical, and are outgrowing the protection provided by line-interactive systems. As in the data center, double-conversion UPS should be specified for any application considered mission critical.

Because the UPS system is often located in the same rack as the equipment being protected, redundancy is simpler in network closets than it is in the data center, where the UPS system is supporting multiple racks spread across the room. And, with the reduced size of today's rack-mount UPS systems, 1 + 1 redundancy is feasible in remote environments and can deliver very high availability. At minimum, the UPS should be configured with an external bypass to allow power to be routed around the UPS during service or in the event of a failure.

Heat Removal

Heat removal is also critical to ensuring network availability. Heat removal becomes even more critical as the trend toward server consolidation accelerates. Packing more processing power into a smaller space and packing more servers in each rack can test and even exceed the limits of raised-floor cooling.

Facilities have dealt with this situation by increasing rack spacing, essentially distributing the heat from the equipment over a larger space. This offers, at best, an interim solution. Increasing rack spacing quickly consumes available data center space while reducing the number of racks that the data center can support. Combining room-level precision cooling with supplemental high-density cooling offers the most economical and energy-efficient solution.

Systems outside the data center are becoming more powerful and more critical, and are outgrowing the protection provided by line-interactive systems.

Consider a typical 10,000-square-foot data center. Assuming average rack power densities of 1 kW, racks take up approximately 35 percent of the floor space. This leaves 65 percent of space for aisles and support systems. If a typical rack consumes about 7 square feet of floor space, this facility can support up to 500 racks of 1 kW each.

When average power density increases to 10 kW per rack, the data center must increase rack spacing to spread the heat load across the room. Now racks can use only 3.5 percent of the space in the room and the facility can support only 50 racks.

Assuming a data center shell cost of $200 per square foot and cooling costs of $925 per kW with a cooling system designed for densities of 50 Watts per square foot, a 10,000-square-foot facility costs $2.46 million. At rack densities of 1 kW, the cost is $4,920 per rack.

When rack densities increase to 10 kW, shell and cooling costs stay the same, but now the same infrastructure can support only 50 racks. Consequently, the cost to support each rack grows by a factor of 10 to $49,200.

Although supplemental cooling boosts cooling-system costs, it also supports closer equipment arrangement. With supplemental high-density cooling, the 10,000-square-foot data center can sup-

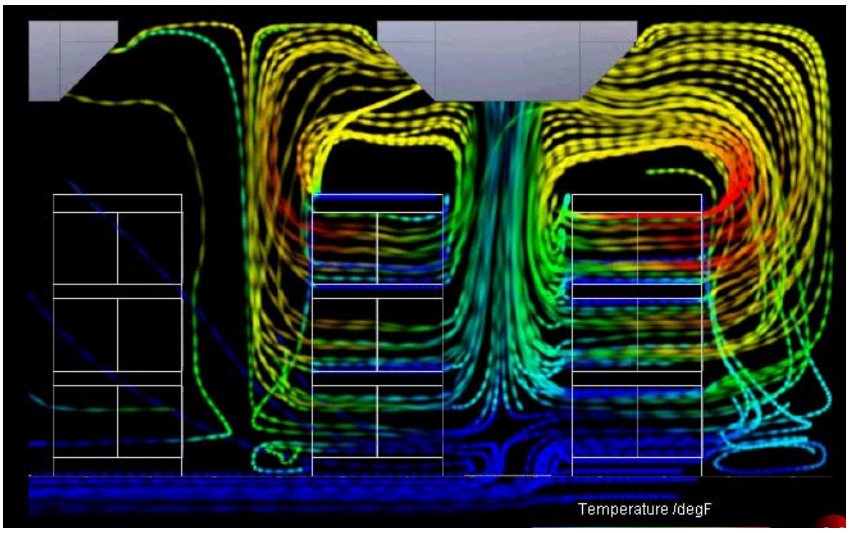

Figure 2. New cooling solutions supplement air coming from under the floor with focused cooling mounted on the rack or above the cold aisle.

port 386 racks of 10 kW each – a 700 percent increase in capacity. This drives support cost per rack down to $7,128 – less than 15 percent of the cost of a solution that relies on rack spacing.

Outside the data center, sensitive electronics too often rely on building air conditioning for heat removal. This approach does not provide adequate cooling, particularly during off hours or winter months. As equipment densities and power requirements outside the data center rise, problems created by this approach will become more prevalent.

Two solutions now exist for these environments. Compact, ceiling-mounted cooling units can provide dedicated, year-round cooling for small room and closet environments. Special enclosures are also available that integrate cooling within a sealed rack to create a protected environment for equipment outside the data center.

Another challenge that is being presented by rising densities is the rate at which rooms heat up when an outage occurs. UPS-protected systems continue to operate and generate heat, but the cooling system shuts down, causing heat to accumulate in the room or rack. At densities of 150 Watts per square foot, it takes a room approximately 13 minutes to go from 70 degrees F to 95 degrees F. When densities increase to 300 Watts/square foot, that time is reduced to eight minutes. At 450 Watts per square foot, the room heats to 95 degrees F in about two minutes.

This problem caused wireless service failures during the Great Blackout of 2003. Critical systems shut down while they still had ample battery capacity because of heat buildup in the room in which they were operating.

In the data center, precision cooling systems should be supported by a backup generator whenever possible Outside the data center new enclosure systems are now available with battery protected fans to exhaust heat out of the rack during an outage.

Monitoring and Management

Despite increasing pressure to boost network availability, support system management remains an under-used strategy for maintaining and elevating availability. By collecting data that can serve as the basis for a preventive maintenance program and providing early warnings to the right people at the right time, monitoring solutions can

help avoid disaster. At the least, they support faster recovery.

A basic approach to support system monitoring should cover mission critical power and cooling systems. As availability needs rise, monitoring can be extended to other support equipment, such as generators, automatic transfer switches, physical security systems and leak detection systems.

For systems operating outside the data center, critical system management provides essential visibility and control. Communication cards in remote UPS systems provide a constant update of UPS status and enable remote restart of connected equipment. Temperature monitors can be installed in racks or small rooms to ensure heat does not reach dangerous levels. These inputs can be integrated into the centralized monitoring system in the data center or into a network management system.

Because the IT and facility management departments often share responsibility for vital support systems, an effective monitoring strategy provides each with the information they need in a manner that is consistent with the way they monitor and maintain other systems within their responsibility. This requires that monitoring components support both the Modbus and BACnet protocols used by building management systems and SNMP, which is standard for network management systems.

Service

Maintaining support-system availability requires a well-coordinated preventive maintenance and emergency service strategy that addresses both mechanical and electrical systems.

Organizations that choose to handle maintenance themselves should work with support-system vendors to identify an effective preventive maintenance schedule and to obtain instructions for handling failures. They must also obtain a list of parts to keep on hand to aid prompt recovery in the event of a failure. That inventory should be included in total ownership cost calculations. The wait time to receive parts not kept on hand should also go into evaluating the do-it-yourself approach.

For systems operating outside the data center, critical system management provides essential visibility and control.

When striving for high availability, working with a professional service organization offers several advantages. Most importantly, it ensures required uptime through 24-hour service support. In addition to reducing the frequency of failures and speeding recovery time, professional service also removes the burden of maintenance from IT staff members, allowing them to focus on managing the network. Systems outside the data center should be integrated into an overall proactive service strategy that includes regular preventive maintenance.

Conclusions

Any significant change in an organization's physical structure, technology deployment or business objectives should trigger a reevaluation of the four

elements of its physical support plan – power, heat removal, monitoring and service management.

The most reliable critical power systems minimize or eliminate single points of failure in the simplest way possible and have the flexibility to adapt to future requirements.

Outside the data center specify double conversion UPS systems, carefully evaluate battery requirements and make sure that taking the UPS offline doesn't require connected equipment to be shut down.

From a cooling perspective, be aware of the increased heat loads being created by new high density systems. New solutions to cooling have been developed specifically for these systems. Increasing densities will also drive increased use of dedicated precision cooling systems outside the data center.

Systems across the network should be monitored for both alarm conditions and to collect data on system operation. A service strategy that focuses on keeping systems running, rather than one that deals with problems after a failure has occurred, ensures systems achieve the levels of availability they are capable of.

Liebert®

LIEBERT CORPORATION
1050 DEARBORN DRIVE
P.O. BOX 29186
COLUMBUS, OHIO 43229
800.877.9222 (U.S. & CANADA ONLY)
614.888.0246 (OUTSIDE U.S.)
FAX: 614.841.6022
www.liebert.com

Printed in U.S.A. 0405 WP406

EMERSON™
Network Power