the Availability Digest

Calculating Availability – Hardware/Software Faults

January 2007

The Diminishing Effect of Hardware Failures

In our previous articles, we have discussed the basic availability equation for active/active systems, how the equation is affected by different repair strategies, and how it is affected by the requirement to recover a node and subsequently restore it to service once it has been repaired.[1] In all of these analyses, we assumed that the active/active system failure was caused by the failure of *s*+1 nodes due to hardware faults, where *s* is the number of spare nodes in the application network. Some component needed to be repaired in at least one node, and that node had to be recovered so that the active/active system could be restored to service.

However, in actual practice today, active/active system failures are rarely caused by dual hardware failures. Hardware has simply become much more reliable than software and system operators. If a system fails, it is most likely that one or more nodes failed due to a software fault or an operator error.

For instance, if the probability of a node failure due to a hardware fault is 10%, the probability of a system failure caused by two node failures, each taken down by a hardware fault, is only 1%. Other factors contribute to the other 99% of all system failures.

Consequently, an active/active system failure is highly likely to have been caused by at least one node having been taken down by a failure other than hardware – perhaps a software fault, an operator error, or an environmental fault of some kind. In this case, the system will be restored to service without the need to repair a hardware component. Only one of the nodes needs to be recovered in order to allow the active/active system to be restored to service.

Node recovery may require rebooting the system, reloading applications, opening the database, and testing that the node is, in fact, operational. System restoration may require resynchronizing the databases in the application network. These activities are likely to be much faster than a hardware repair – hours rather than days, resulting in a much faster system restoration than if a hardware component needs to be repaired.

This article shows the modifications to be made to the availability equations to account for the fact that only some system failures require a hardware repair.[2]

[1] Calculating Availability – Redundant Systems, *Availability Digest*; October, 2006.
Calculating Availability – Repair Strategies, *Availability Digest*; November, 2006.
Calculating Availability – The Three Rs, *Availability Digest*; December, 2006.

[2] The impact of hardware faults on system availability is dealt with in detail in *Breaking the Availability Barrier: Achieving Century Uptimes with Active/Active Systems*, by Paul J. Holenstein, Dr. Bill Highleyman, and Dr. Bruce Holenstein.

Availability Analysis Reviewed

The equations which we developed earlier that considered repair strategies, the recovery of nodes, and the restoration to service of an active/active system are

$$F = \frac{r/(s+1) + R}{r/(s+1)} f(1-a)^{s+1} \quad \text{for parallel repair} \qquad (1)$$

$$F = \frac{r+R}{r} f(1-a)^{s+1} \quad \text{for sequential repair} \qquad (2)$$

Parallel repair implies that there is a different service technician simultaneously working on each failed node. Sequential repair means that there is only one service technician working on one node at a time. The parameters in these equations are

F is the probability of failure of the active/active system.
a is the availability of a node.
r is the repair time for a node (hardware repair plus recovery).
R is the system restoration time.
s is the number of spare nodes provided for the system.
f is the number of ways that *s*+1 out of *n* nodes can fail.
n is the number of nodes in the active/active system.

System availability is, of course, 1-*F*.

The Impact of Occasional Hardware Faults

In the above equations, *r* is the time to repair the node and to recover it so that it can be returned to service. Should a node fail due to a problem other than a hardware fault, there are still activities that must be performed before a node can be recovered and returned to service. These activities might include the following:

- the time to decide that the node has really failed.
- the time to determine the cause of the problem – is it a hardware fault that needs to be repaired or a software or operator error that has done no other damage to the system.
- rebooting the node and restarting its applications.
- opening the database by the applications.
- testing the node to ensure that it appears to be working properly.

The time to perform these tasks is what we have referred to as the recovery time of the node.

Let us modify our definition of the above parameters. Let *r* be the repair time, if any, and *r’* be the recovery time:

r is the time to repair a hardware fault.
r’ is the time to recover a node (including the initial decision times).

Let us further define *h* as being the probability that a node will be taken down by a hardware fault:

h is the probability of a node failure due to a hardware fault.

All node failures require a recovery time of *r'*. In addition, *h* of those node failures require a hardware repair time of *r*. Thus, the nodal mean time to recover, mtr, is

$$\mathrm{mtr} = r' + hr \qquad (3)$$

In our earlier equations expressed by Equations (1) and (2) above, the nodal mtr was the repair and recovery time, *r*. All we need to do is to update that value with our new value of mtr as expressed in Equation (3). This yields

$$F = \frac{(r'+hr)/(s+1)+R}{(r'+hr)/(s+1)} f(1-a)^{s+1} \quad \text{for parallel repair} \qquad (4)$$

$$F = \frac{(r'+hr)+R}{(r'+hr)} f(1-a)^{s+1} \quad \text{for sequential repair} \qquad (5)$$

Note once again the differences between *r*, *r'*, and *R:*[3]

r is the time to repair a hardware fault in a node.

r' is the time to recover a node (including the initial decision times) excluding repair time, if any. It is the time spent by personnel at the node site from the time that the nodal fault occurred to the time that the node is ready to be returned to service in the active/active network, exclusive of hardware repair.

R is the time to restore the active/active system to service once one of its failed nodes has been recovered. This time might include resynchronizing the database copies in the application network, entering backlogged transactions, and any other activity that must be performed at a full system level before service is returned to the users.

The Point of No Return

As hardware becomes more and more reliable, there comes a point of no return beyond which further improvements in hardware availability provide no improvement in system availability. This is because the rate of system failures caused by software faults and operator errors is orders of magnitude greater than the rate of such failures caused by hardware faults. The NonStop triple modular redundancy (TMR) NSAA configuration which offers seven 9s of hardware availability is a case in point.[4]

As hardware becomes more reliable, the probability of a hardware fault, *h*, approaches zero; and the term *r'+hr* approaches *r'*. In the limit, Equations (4) and (5) approach

$$F = \frac{r'/(s+1)+R}{r'/(s+1)} f(1-a)^{s+1} \quad \text{for parallel repair} \qquad (6)$$

$$F = \frac{r'+R}{r'} f(1-a)^{s+1} \quad \text{for sequential repair} \qquad (7)$$

[3] Equations (4) and (5) have been derived rather intuitively. A formal analysis shows that these relationships are valid provided that the repair time is much larger than the recovery or restore times. For the detailed analysis, contact editor@availabilitydigest.com. Caution: For math nuts only.

[4] R. Buckle, W. Highleyman, The New NonStop Advanced Architecture: A Massive Jump in Processor Reliability, The Connection; September/October, 2003.

Equations (6) and (7) represent the limiting failure probabilities that can be achieved if there is no probability of a hardware fault. These limits are reached if the contribution of hardware repair time, *hr*, to nodal mtr is very much less than the contribution of nodal recovery time, *r'*.

For instance, assume that the nodal recovery time is 4 hours and that the repair time is 24 hours. If the probability of a hardware failure is 0.1% (.001), then hardware faults add only .024 hours to the four hours of recovery time for a total mtr of 4.024 hours. Reducing the probability of hardware failure even further will have little impact on system availability.

Summary

Not all node failures are caused by hardware faults. In fact, with today's hardware reliability, hardware faults generally contribute much less to node failures than do software faults or operator errors.

As hardware becomes more and more reliable as it has over the last several years, there comes a point at which further improvements in hardware reliability will have little impact on system availability. For instance, the hardware reliability of the triple modular redundancy configuration for HP's new NonStop servers has reached this limiting point.