the Availability Digest

Calculating Availability – Failover

February 2007

In the past several articles, we have concentrated on the calculation of availability of redundant systems. We have considered the failure of a system caused by the failure of all of its spare subsystems, leaving it exposed to the failure of just one more subsystem. We have analyzed the impact of different repair strategies and of system restoration time subsequent to nodal repair. We have looked at the case in which a subsystem fails for reasons other than a hardware failure and therefore needs no repair, just a node recovery.

In these analyses, we have assumed that once a single subsystem failed, the recovery from this failure was instantaneous as a redundant backup subsystem took its place. Therefore, the failure of a single system failure contributed nothing to the lack of availability.

However, failover is not instantaneous. It can take anywhere from milliseconds to days depending upon the system. Long failover times can contribute significantly to unavailability. In this article, we look at the impact of failover times on availability.

But first, we review the results of our previous analyses.

A Review of Availability Calculation

We define *A* as being the availability of a system (that is, the probability that the system is up), and *F* as being the system failure probability (that is, the probability that the system is down). Since the system is either up or down, then

$$A = 1 - F \tag{1}$$

If the average time that the system is up is MTBF (the mean time between failures), and if the average time that it is down is MTR (the mean time to repair), then

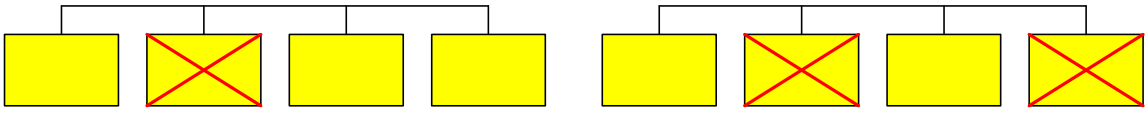
$$A = \frac{MTBF}{MTBF + MTR} \tag{2}$$

and

$$F = \frac{MTR}{MTBF + MTR} \approx \frac{MTR}{MTBF} \tag{3}$$

where the approximation in Equation (3) is valid if MTBF >> MTR.

We consider a system with *n* spare subsystems, *s* of which are spares. That is, the system will survive the failure of *s* subsystems. However, the system fails if *s*+1 subsystems should fail.

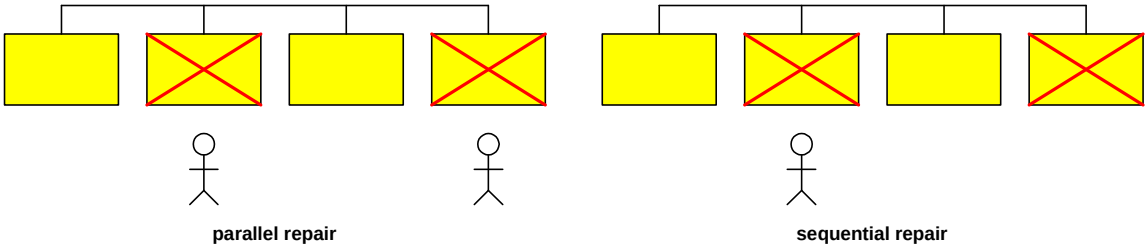


A Four-Node System With One Spare

There are two repair strategies that we have considered – parallel repair and sequential repair:

- Parallel Repair – When multiple subsystems have failed, repair proceeds independently via separate service technicians working on each subsystem. The return to service of one subsystem is independent of the others.

- Sequential Repair – Only one service person is available, and he works on one subsystem at a time. Therefore, one subsystem is repaired and returned to service; and then the next subsystem is repaired.



Repair Stategies

Letting *a* be the availability of a subsystem, *r* be the subsystem's repair and recovery time, and R be the restore time of the system once one subsystem has been returned to service, we showed that the system failure probability, *F*, for the two repair strategies is given by

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

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

where

F is the probability of failure of the system.
f is the number of failure modes for the system. It is the number of ways that the failure of *s*+1 subsystems out of *n* subsystems can cause a system failure..
n is the number of subsystems in the system.
a is the availability of a subsystem.
s is the number of spare subsystems in the system.
r is the repair and recovery time for a subsystem. That is, it is the time required to return the subsystem to service.
R is the restore time of the system. It is the time that it takes to perform system-wide functions required to return the system to service once it has a full complement of subsystems. For instance, these functions may include database resynchronization and the reentry of transactions that occurred during the system's downtime.

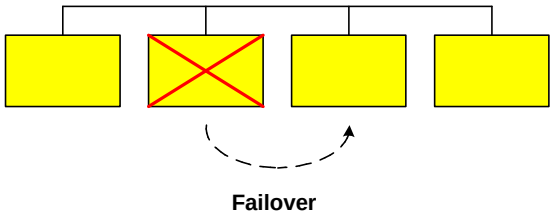
For a single-spared system (s = 1), these relationships become

$$F = \frac{r/2 + R}{r/2} \frac{n(n-1)}{2} (1-a)^2 \qquad \text{for parallel repair} \qquad (6)$$

$$F = \frac{r + R}{r} n(n-1)(1-a)^2 \qquad \text{for sequential repair} \qquad (7)$$

If a fault is due to a hardware fault, the hardware must be repaired and the subsystem recovered. However, if the fault is not caused by hardware, the subsystem needs only to be recovered. In this case, if only h of the faults are caused by hardware, the subsystem recovery time r in the above equations can be replaced with (r'+hr) where r' is the subsystem recovery time, h is the proportion of faults caused by hardware, and r is now the hardware repair time.

What is Failover?



In a redundant system, when a component fails, the system invokes a backup component to take its place. This process is called failover – the functions of the failed component are failed over to the backup component.

Depending upon the system, failover can be measured in milliseconds (process pairs in a NonStop system), seconds (an active/active system), minutes (a cluster), or hours or more (a cold backup system). During the failover period, some system functions may not be available; or some users may not be provided service. Once the failover is complete, all services are restored, albeit at perhaps a reduced responsiveness if components are more heavily loaded.

The Impact of Failover Time on Availability

We start by assuming the worst case – that the system is totally unavailable during the failover time. This is the case for active/standby computer configurations as well as for clusters if a cluster node is providing unique services. We will cover the case of partial unavailability during failover later. Active/active systems are examples of this case since a failure of one node out of n denies service to only 1/n of the users rather than all users.

When considering failover time, we have two sources of system down time:

- s+1 subsystems fail, or
- one subsystem fails, and the system must failover to a backup subsystem.

s+1 Subsystems Fail

The contribution to the probability of system failure due to a multiple subsystem failure is given by Equations (4) through (7) above.

Failover

The system will be down for a failover time every time there is a single subsystem failure. If each of the *n* subsystems has a mean time before failure of mtbf, an *n*-subsystem system will have a single subsystem failure every mtbf/*n* units of time. Let us define MTFO as the mean time for failover. Then the probability that the system will be down while it is failing over is

$$\text{probability that system is down due to failover} = \frac{\text{MTFO}}{\text{mtbf/n+MTFO}} \qquad (8)$$

That is, once a failover has completed, there will be an average time of mtbf/*n* until the next subsystem failure occurs (remember that mtbf is defined as the mean time before failure, not the mean time between failures). When that subsystem fails, a failover must take place, taking a time of MTFO. Therefore, the total time between subsystem failures is (mtbf/*n* + MTFO). The system is down MTFO of this time.

System Probability of Failure with Failover

The probability of system failure during failover time must be added to the system down time due to a multiple system failure. Taking the case of a single-spared system with parallel repair, the system failover probability becomes

$$\begin{aligned} F &= \frac{r/2+R}{r/2}\frac{n(n-1)}{2}(1-a)^2 + \frac{\text{MTFO}}{\text{mtbf}/n+\text{MTFO}} \\ &\approx \frac{r/2+R}{r/2}\frac{n(n-1)}{2}(1-a)^2 + \frac{\text{MTFO}}{\text{mtbf}/n} \end{aligned} \qquad (9)$$

where *F*, *r*, *R*, *a*, and *n* are defined above and

mtbf	is the mean time before failure of a subsystem.
MTFO	is the mean time for failover.

The approximation is good if mtbf/*n* >> MTFO, which will normally be the case.

The parameters for nodal recovery time, *r,* nodal mean time before failure, mtbf, and nodal availability, a, are related by Equations (1) and (3) as follows:

$$\text{mtbf} = r/(1-a)$$

Thus, Equation (9) can be rewritten as

$$F \approx \frac{r/2+R}{r/2}\frac{n(n-1)}{2}(1-a)^2 + \frac{\text{MTFO}}{r}n(1-a) \qquad (10)$$

Note that as a practical matter, MTFO should be no larger than *R*, the system restore time, since the time to bring up the backup system should be no longer than the time to bring up the entire system following a multiple subsystem failure. In fact, it will generally be much less (except, perhaps, for a cold standby).

Some Examples

A Cluster

Let us first take the case of a single-spared, four-node cluster which is made up of nodes with availabilities of .999. The recovery time for a node is two hours, as is the system restore time. The failover time is three minutes. Thus,

r	= 2 hours
R	= 2 hours
n	= 4
a	= .999
s	= 1
MTFO	= .05 hours (3 minutes)

The probabilities of failure for this case are:

Probability that the system is down due to a multiple node failure = 1.8×10^{-5}
Probability that the system is down during failover = 10×10^{-5}
Probability that the system is down = 11.8×10^{-5}

The inherent system availability has been reduced from .999982 to .999882, or from a little less than five nines to a little less than four nines, due to failover times. Failover time has reduced availability by about one 9. Put another way, failover time has increased downtime by more than a factor of six.

A Hot Standby

Let us now take a fault-tolerant system with an availability of four 9s. The system is backed up by a like system as a hot standby. The mean time to failover is 2 hours. All other parameters are the same:

r	= 2 hours
R	= 2 hours
n	= 2
a	= .9999
s	= 1
MTFO	= 2 hours

The results for this case are

Probability that the system is down due to a multiple node failure = 3×10^{-8}
Probability that the system is down during failover = 2×10^{-4}
Probability that the system is down = 2×10^{-4}

The inherent system availability has been reduced from over seven 9s to less than four 9s. Failover time dominates the system availability in this case.

Active/Active Systems

Active/active systems and systems with similar characteristics are somewhat different in that only a portion of users are affected by a failover. If users are evenly distributed across the *n* nodes in the system, the failure of a node affects only 1/*n* of the users.

In many applications, availability is taken as the availability of services to the user, not of the availability of the system as a whole. In this case, a failover affects only 1/*n* of the users; and therefore the probability that the system is unavailable due to failover should be reduced by that factor.

Equation (10) then becomes

$$\begin{aligned} F &\approx \frac{r/2+R}{r/2}\frac{n(n-1)}{2}(1-a)^{s+1} + \frac{MTFO}{r}\frac{1}{n}n(1-a) \\ &\approx \frac{r/2+R}{r/2}\frac{n(n-1)}{2}(1-a)^{s+1} + \frac{MTFO}{r}(1-a) \end{aligned} \qquad (11)$$

Let us take as an example an active/active system with the same parameters as the hot standby example above except that the failover time is 1 second. Then

r	= 2 hours
R	= 2 hours
n	= 2
a	= .9999
s	= 1
MTFO	= 1 seconds = .00028 hours

The results for this case are

Probability that the system is down due to a multiple node failure = 3×10^{-8}
Probability that the system is down during failover = 1.4×10^{-8}
Probability that the system is down = 4.4×10^{-8}

The probability of failure has been increased by about 50%, but the attribute of extreme availability has been maintained (over seven 9s).

Summary

Failover time plays a very important and sometimes dominant role in system availability. For some system configurations such as an active/standby system, failover times in the order of hours completely mask the system downtime due to dual system failures. In these cases, the resulting system cannot really be considered a high availability system. It is a disaster-tolerant system in that it can recover from failures of the active system but only at the cost of seriously reduced availability.

Clusters fare significantly better. Failover times contribute a significant but not an overwhelming contribution to the failure probability of a cluster.

Active/active systems still retain their attribute of extreme availability in the presence of failover times so long as these times can be kept short, measured in seconds rather than minutes.