the Availability Digest

www.availabilitydigest.com

Simplifying Failover Analysis – Part 2

June 2011

In our article, Simplifying Failover Analysis – Part 1,[1] we discussed the impact of failover time and failover faults on redundant systems. In a two-node redundant system, users are down if:

- both nodes fail, or
- one node fails, and the users are in the process of being failed over, or
- one node fails, and a failover fault occurs.

We showed that failover can be modeled as a two-node redundant system with the availability of one node being reduced by the effects of failover time and failover faults.

In this article, we extend the results of Part 1 to accommodate two additional complexities when modeling failover:

- What if the redundant nodes are different with different availability characteristics?
- How do we handle a redundant production node that is in the process of failing over internally? Even though it is technically down, it will not fail over to its backup node.

A Review of Failover Analysis

The Impact of Failover Time and Failover Faults

In Part 1, we defined the following parameters:

a	availability of a node
mtbf	nodal mean time between failure (the average time between failures for a node)
mtr	nodal mean time to recover (the average time to restore a node to service)
mtfo	time to fail over (the average failover time)

[1] http://www.availabilitydigest.com/public_articles/0510/failover_analysis.pdf

d	probability of a failover fault
A	availability of a system (probability that system is up)
F	probability that a system is down

We showed that the probability of downtime when failover is considered is

$$p(\text{downtime}) = F = (1-a)^2 + \frac{\text{mtfo}}{\text{mtbf}} + (1-a)d \qquad (1)$$

where the first term is the probability that both nodes will fail, the second term is the probability that the system will be in the process of failing over, and the third term is the probability that there will be a failover fault.

Using the relationship a = 1 – mtr/mtbf, or mtbf = mtr/(1- a), Equation (1) can be written as

$$p(\text{downtime}) = F = (1-a)^2 + (1-a)\frac{\text{mtfo}}{\text{mtr}} + (1-a)d \qquad (2)$$

Rearranging terms, we have

$$F = (1-a)\left[(1-a) + \frac{\text{mtfo}}{\text{mtr}} + d\right] = (1-a)\left[1 - \left(a - \frac{\text{mtfo}}{\text{mtr}} - d\right)\right] \qquad (3)$$

Let us define a modified node availability, *a'*:

$$a' = a - \frac{\text{mtfo}}{\text{mtr}} - d \qquad (4)$$

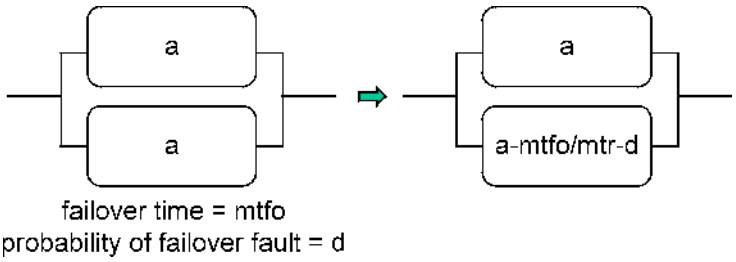
Equation (3) can then be written as

$$F = (1-a)(1-a') \qquad (5)$$

and

$$\text{system availability} = A = (1-F) = 1-(1-a)(1-a') \qquad (6)$$

Thus, the system behaves as a two-node system, a first node with an availability of *a* and a second node with an availability of *a* reduced by the effects of failover:



This observation provides a simple way to calculate the availability of a redundant system when it is impacted by failover times and failover faults.

The Impact of Failover

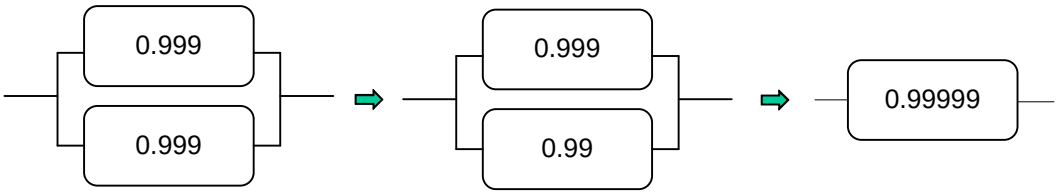
The impact of failover on system availability can be shown through a simple example. Consider a redundant system comprising two servers, each with three 9s availability (*a*) and a mean time to restore (mtr) of four hours. From Equation (6), if it were not for failover, this system would have an availability, A, of six 9s. In other words, the only failure mode is if both nodes should fail. Since we have no control over nodal availability (that is up to the manufacturer), we call this the *inherent availability* of the system. We can do no better than this.

Let us consider the impact of failover on this system. Assume that the failover time is one minute, and the probability of a failover fault is 0.5%. From our above analysis resulting in Equation (4), the effective availability, *a'*, of the second node is

$$a' = a - \frac{mtfo}{mtr} - d = 0.999 - \frac{1}{4x60} - 0.005 \approx 0.990$$

The system availability is therefore

$$A = 1 - (1 - 0.999)(1 - 0.990) = 0.99999$$



mtfo = 0
d = 0
$F = 10^{-3}x10^{-3} = 10^{-6}$
A = six 9s

mtr = 4 hours
mtfo = 1 minute
d = 0.005
$F = 10^{-3}x10^{-2} = 10^{-5}$
A = five 9s

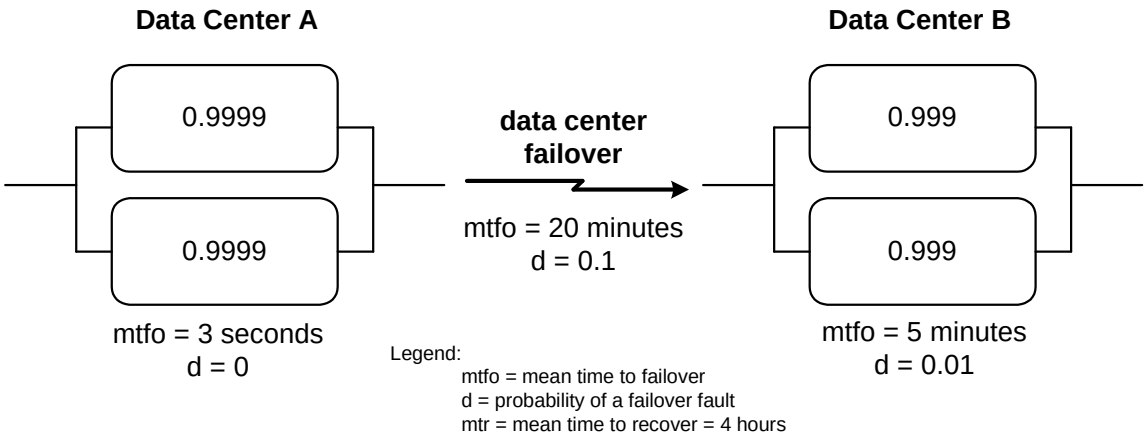
The system availability has been reduced from six 9s to five 9s. A failover time of only one minute and a failover fault rate of one in 200 has increased the amount of downtime by an order of magnitude! Failovers do count.

An Example System

The above analysis was then applied in Part 1 to a complex configuration of two data centers, each with a redundant set of systems.

Though both data centers are actively engaged in their own processing activities, Data Center A is running a particularly critical application that is backed up by less expensive systems in Data Center B.

In Data Center A, the application is running in an active/active system[2] comprising two fault-tolerant nodes. Each of the fault-tolerant nodes has an availability of four 9s[3] (each is up 99.99% of the time). Being active/active, users on a failed node can be failed over to the surviving node in three seconds. There are no failover faults since it is known that the surviving node is properly operating – after all, it is currently processing transactions.



The active/active system is backed up by a more economical cluster in Data Center B. Industry-standard servers are used with a nodal availability of three 9s (each node is up 99.9% of the time). Failover time is five minutes, and the probably of a failover fault is 1% (that is, 99 out of 100 failovers will be successful).

Should the active/active system in Data Center A fail, it takes twenty minutes on the average to fail over to Data Center B. Failover testing is limited due to the complexity and risk of failover, but what testing has been done indicates that one out of 10 failovers will be unsuccessful (the probability of a failover fault is 10%).

---

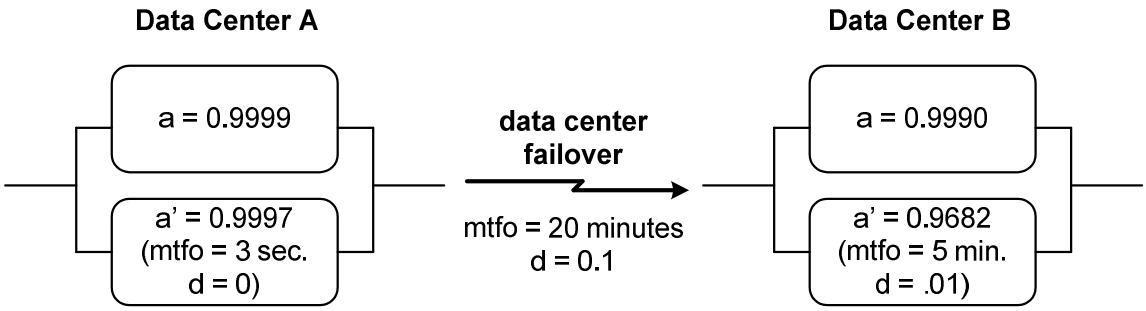
[2] What is Active/Active?, *Availability Digest*; October 2006.
[3] W. H. Highleyman, P. J. Holenstein, B. D. Holenstein, Chapter 1, The 9s Game, *Breaking the Availability Barrier: Survivable Systems for Enterprise Computing*, AuthorHouse; 2004.

Regardless of the node type – fault-tolerant or standard servers – the time to repair a node – the nodal mtr - averages four hours.

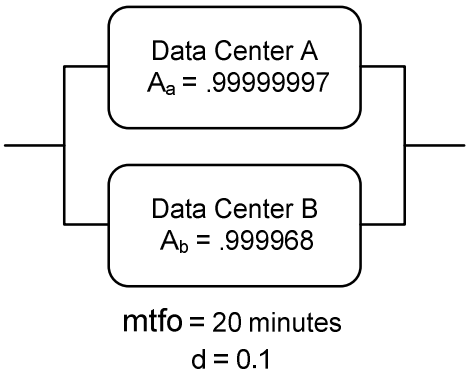
What is the system availability from the user's viewpoint?

Step 1 – Calculate the Availability of each Data Centers

The first step of this analysis was carried out in Part 1. In that step, the data-center nodes were replaced with nodes with availabilities modified by the failover parameters. The following configuration resulted:



The availability of each data center now can be easily determined, resulting in the following configuration:



Step 2 – Calculate the Availability of the Data Center Complex

We left Part 1 at this point with the observation that we now have two problems that require some modifications to our earlier analysis:

- A failover from Data Center A to Data Center B occurs only if Data Center A "fails hard" (i.e., both of its nodes fail, or it suffers an internal failover fault). There will not be a

failover from Data Center A to Data Center B if Data Center A is down because it is in the process of failing over internally.

- The two nodes have different availabilities.

Failover to Data Center B Does Not Happen If Data Center A is Failing Over Internally

The first problem is solved by separating out the Data Center A failover time. If failover time is ignored, the only impact of failover on Data Center A is a failover fault. However, since Data Center A uses an active/active system, the probability of a failover fault is 0 (d = 0). The availability of Data Center A when considering only hard failures is therefore [see Equation (3)]:

$$1-(1-a)[1-(a-d)] = 1-(1-0.9999)(1-0.9999) = 0.99999999 \text{ (eight 9s)}$$

Let us call this value the "hard" availability of Data Center A, A_a. A_a is the probability that Data Center A will be up unless because both of its nodes are down or because it suffered an internal failover fault:

$$A_a = 0.99999999$$

The probability that Data Center A will be down because it is failing over is, from Equation (1):

$$\text{p(A is failing over)} = \frac{\text{mtfo}}{\text{mtbf}} = (1-a)\frac{\text{mtfo}}{\text{mtr}} = (1-0.9999)\frac{3}{4\text{x}3600} = 2\text{x}10^{-8}$$

Note that subtracting the probability of failover of Data Center A from its "hard" availability gives the Data Center A availability calculated earlier:

$$0.99999999 - 2 \text{ x } 10^{-8} = 0.99999997$$

Nodal Availabilities are Different

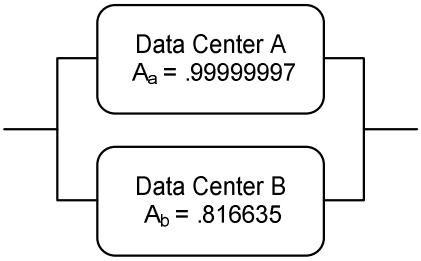
So far as the second issue is concerned (the two data centers have different availabilities), we can refine Equation (3) for this purpose by noting that it is only the failure of the production node that will cause a failover. Using the notation A_a for the production site hard availability and A_b for the backup site availability (including failover time), Equation (3) can be recast as

$$F = (1-A_a)\left[1-\left(A_b - \frac{\text{mtfo}}{\text{mtr}} - d\right)\right] \qquad (7)$$

In this example, as calculated above,

$$A_a = 1 - (1 - 0.9999)(1 - 0.9999) = 0.99999999$$
$$A_b = 1 - (1 - 0.999)(1 - 0.9682) = 0.9999682$$

Thus, the availability of the system is that of the production node backed up with a node whose availability is its inherent availability reduced by the effects of failover. This yields the following configuration:



where $A_b' = A_b - mtfo/mtr - d = 9999682 - 20/(4x60) - 0.1 = 0.816635$.

The overall availability for the dual data centers is therefore (ignoring Data Center A's failover time):

Dual data center availability = 1 – (1 - .99999999)(1 - .816635) = 0.9999999982
(almost nine 9s)

From this, we have to subtract the probability that the system will be down due to a Data Center A internal failover:

Dual data center availability = 0.9999999982 - 0.00000002 = 0.9999999782 (almost eight 9s)

Note that the three-second failover time of Data Center A has reduced total system availability by one 9. In fact, the probability of total system failure is primarily due to the three-second failover time of Data Center A!

The five-9s cluster backup in Data Center B has increased the overall availability by only 40% because the predominant factor in the system's availability is the failover time in Data Center A – no amount of data-center redundancy can make that any better. More precisely, the probability of downtime has been decreased from $3x10^{-8}$ to $2.18x10^{-8}$, a reduction of downtime of about 1.4. $2x10^{-8}$ of this probability is due to Data Center A failover time.

Is the cost of this remote cluster worth an increase of 40% in the reliability of the application? That, of course, depends upon the application. But this technique provides a simple way to calculate the availability of complex systems so that this judgment can be made.

It is worth noting that if there were a modest probability of a failover fault in Data Center A – say in the order of 1%, a similar computation would show that the backup data center does, in fact, significantly improve reliability.

Of course, beyond the need for high availability, the remote cluster may be justified in order to guarantee recovery from a disaster that might destroy the primary site. That is a different consideration.

MTRs are Different

One case we have not considered is what if the mean time to repair is different for both data centers. In Equation (7), which mtr do we use?

Note that the mtr term in Equation (7) comes about through rearrangement of the term mtfo/mtbf. In this term, mtfo is the failover time from Data Center A to Data Center B; and mtbf is the time interval between failovers. Therefore, we are dealing with the mtbf of Data Center A (it is its failure that triggers a failover). Consequently, mtr in Equation (7) is that of Data Center A.

To account for this, Equation (7) should be written as

$$F = (1 - A_a)\left[1 - \left(A_b - \frac{mtfo}{mtr_a} - d\right)\right] \qquad (8)$$

where mtr_a is the mean time to repair Data Center A. This is the mtr of its nodes (a node must be brought back into service in order to bring Data Center A back into service).[4]

Summary

Failover in redundant systems is a fact of life with which we have to deal. Even fast and reasonably reliable failovers can have a dramatic effect on downtime and availability. This analysis has shown how the failover characteristics of a complex redundant system can be replaced with an equivalent two-node system to facilitate simple and quick analysis.

[4] This is conservative if there are multiple maintenance technicians working on both failed nodes simultaneously rather than a single technician working only on one node. In this case, the average repair time for the data center will likely be shorter than the average repair time of a single node.