the Availability Digest

www.availabilitydigest.com

Simplifying Failover Analysis – Part 1

October 2010

A system is down if it is not providing service to its users. To improve service availability, it is common to provide redundancy in the system. Should a component fail, one or more other components (not necessarily identical) are available to take over the role of the failed component and to continue service to the users.

A redundant system is certainly down if all redundant components required to provide service fail. However, it is also down for all users that are in the process of being failed over following a single-node failure. Until those users are once again connected to a properly operating system, they must sit idle. Furthermore, failover is not always successful. If failover fails, users are down until one of the systems is returned to service. This is called a *failover fault*.

In this article, we look at the impact of failover on the availability of services to the user. The analysis leads to a surprisingly simple technique for computing the net availability of a redundant system when failover is considered. It also shows by example how reasonably fast and reliable failovers can still have a dramatic impact on system availability.

Redundant Systems

Typical system configurations that are used to provide high- or continuous availability through redundancy include active/passive systems and active/active systems. In an active/passive system, a second system is standing by to take over processing should the production system fail. If this should occur, the passive node is brought into service; and all users are switched to that system.[1]

In an active/active system,[2] one or more processing nodes are all active and are cooperating in the application. A transaction can be sent to any processing node in the application network and can be processed properly. Should a processing node fail, the users connected to that node are switched to a surviving node.

A major difference between active/passive systems and active/active systems is the process of switching users from a failed system to an operational system, a process known as *failover*. In an active/passive system, the backup system must have its database and applications loaded, the user network must be reconfigured, and the system must then be thoroughly tested before it can be put into service. This process could take minutes to hours (even days for large systems using magnetic tape backup).

[1] In this context, a cluster is an active/passive system. The cluster nodes share a common storage system, whereas the nodes in a classic active/passive system have their own databases.

[2] What is Active/Active?, *Availability Digest*; October 2006.

In an active/active system, all that needs to be done is to switch users or transactions from the failed node to surviving nodes, a process that often can be done in seconds.

Failover

A redundant system is certainly down if all redundant components required to provide service fail. However, it is also down for all users that are in the process of being failed over following a single node failure. Until they are once again connected to a properly operating system, they must sit idle. Furthermore, failover is not always successful. If failover fails, users are down until one of the systems is returned to service. This is called a *failover fault*.

Consider a two-node redundant system. Users are down if:

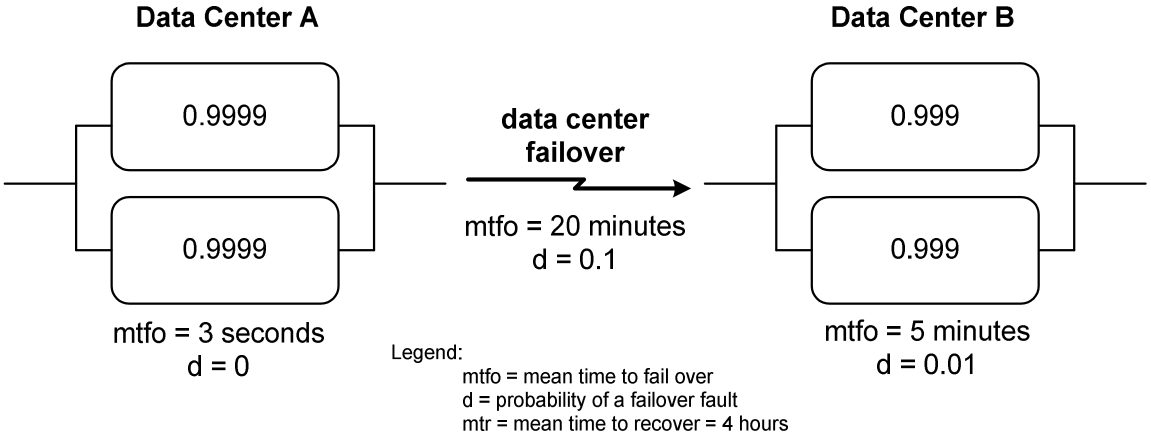
- both nodes fail, or
- one node fails, and the users are in the process if being failed over, or
- one node fails, and a failover fault occurs.

An Example System

Before proceeding with the analysis of the impact on availability of failover, let us look at an example of a fairly complex redundant system comprising both active/active and active/passive systems. We will use this example to calculate the overall system availability with the simplified technique to be developed.

In this example, there are two data centers. Though both 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 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

[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.

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%).

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? We will visit this question after the following analysis.

Failover Analysis

Availability 101

We define 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	mean time to fail over (the average failover time)
d	probability of a failover fault
A	availability of system (probability that system is up)
F	probability that system is down

The basic availability equation is

System availability = A = proportion of time system is up = uptime time/total time.

Using the above parameters, the nodal availability as a function of its mean time between failure, mtbf, and its mean time to repair, mtr, is (note that, by definition, mtbf is total time):

$$a = \frac{\text{mtbf} - \text{mtr}}{\text{mtbf}} = 1 - \frac{\text{mtr}}{\text{mtbf}} \qquad (1a)$$

This can be rewritten as

$$\text{mtbf} = \frac{\text{mtr}}{(1-a)} \qquad (1b)$$

an expression that we will use later.

Failover Analysis

Let p(x) be the probability of event x. From our earlier discussion, the probably that a user will experience downtime is

F = p(downtime) = p(dual-node failure) + p(failing over) + p(failover fault) (2)

Let us look at each of these components of downtime.

Dual-Node Failure

The probability that one node will fail is (1-a). Therefore, the probability that two nodes will fail is

$$p(\text{dual-node failure}) = (1 - a)^2 \qquad (3)$$

Failing Over

The amount of time that a user spends failing over is

$$p(\text{failing over}) = \frac{mtfo}{mtbf} \qquad (4a)$$

Note that the probability of failing over is the same for users on an active/active system as it is for users on an active/passive system. In an active/active system, only half the users are affected by a node failure. However, since there are two active nodes, a node failure will occur twice as often as a failure of the single production node in an active/passive system that takes down all users.

Using Equation (1b), Equation (4a) can be rewritten as

$$p(\text{failing over}) = (1-a)\frac{mtfo}{mtr} \qquad (4b)$$

Failover Fault

A failover fault occurs if the primary node fails and then the failover is unsuccessful. The primary node will fail with probability (1-a). The probability that the failover is unsuccessful is d. Therefore,

$$p(\text{failover fault}) = (1-a)d \qquad (5)$$

Downtime

Substituting Equations (3), (4b), and (5) into Equation (2) yields

$$p(\text{downtime}) = (1-a)^2 + (1-a)\frac{mtfo}{mtr} + (1-a)d \qquad (6)$$

Simplifying Failover Analysis

The key to the simplified method to which we earlier alluded is that Equation (6) can be rewritten as

$$p(\text{downtime}) = (1-a)\left[(1-a) + \frac{mtfo}{mtr} + d\right] = (1-a)\left[1 - \left(a - \frac{mtfo}{mtr} - d\right)\right] \qquad (7)$$

Let

$$a' = a - \frac{mtfo}{mtr} - d \qquad (8)$$

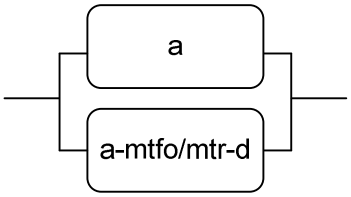
Using Equation (8), Equation (7) can be rewritten as

$$p(\text{downtime}) = (1-a)(1-a') \qquad (9)$$

and

$$\text{system availability} = A = 1 - p(\text{downtime}) = 1 - (1\text{-}a)(1\text{-}a') \qquad (10)$$

In other words, 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.



The Impact of Failover

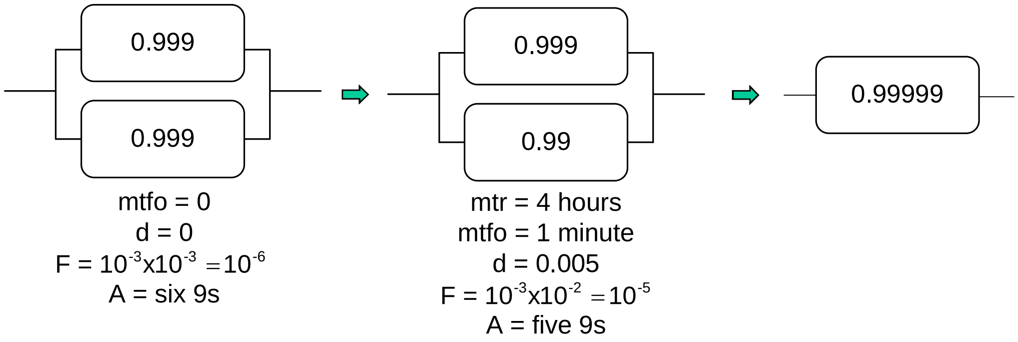
The impact of failover on system availability can be shown through a simple example. Consider a redundant system comprising two servers, each with three 9s availability (*a*) and a mean time to restore (mtr) of four hours. From Equation (10), if it were not for failover, this system would have an availability, 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.

Now 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, the effective availability, *a'*, of the second node is

$$a' = a - \frac{\text{mtfo}}{\text{mtr}} - d = 0.999 - \frac{1}{4\text{x}60} - .005 \approx 0.990$$

The system availability is therefore

$$A = 1 - (1 - 0.999)(1 - 0.99) = 0.99999.$$



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.

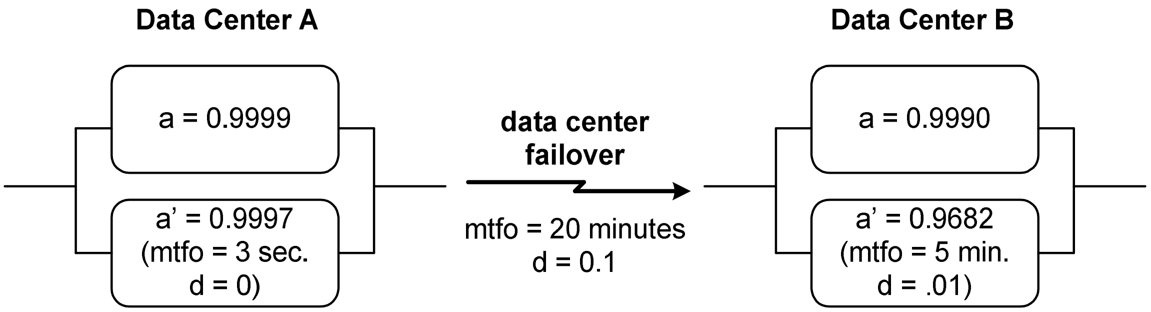
Curiosities

Equation (8) at first glance may seem strange in two regards. First, it says that if the failover time, mtfo, is greater than the recovery time, mtr, the effective availability of the second node is negative. How can this be? This is simply a statement of the intuitive fact that if failover time is greater than recovery time, then recover the failed node - don't fail over to the backup node. In fact, don't even bother with redundancy if this is the case (except, of course, for disaster recovery).

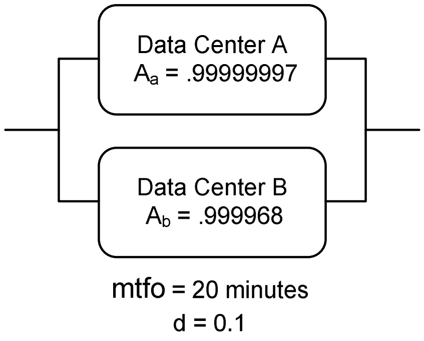
The other curious result is that the availability reduction is the same whether mtfo is one minute and mtr is four hours or if mtfo is ten minutes and mtr is forty hours. Remember, however, that downtime is mtr/mtbf. For a fixed availability, if mtr increases, mtbf also increases proportionately. Therefore, the proportion of downtime due to failover, mtfo/mtbf, stays the same.

An Example

Let us now return to our example that we posed in the beginning of this article. We can replace our data centers each with an equivalent node pair per Equation (8):



The availability of each data center can now be easily determined from Equation 10, resulting in the following configuration:



However, 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.

We will look at the solution to this more complex system in Simplifying Failover Analysis – Part 2.

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. In Part 1, we have shown how to evaluate the effects of failover on simple redundant systems. In Part 2, we extend this to more complex configurations.