the ***Availability Digest***

Calculating Availability – Redundant Systems
October 2006

ff^2

Our logo expresses the basic availability equation for an active/active application network in a somewhat stylized form (see What's That Nerd Logo?). But what is the real relationship between the various subsystem factors and system availability? Though the relationship can become quite complex when many factors are taken into account, the overriding availability equation is relatively simple. It is[1]

$$A = 1 - f(1-a)^{s+1}$$

In this article, we will show the origin of this equation and what it means. The relationship also leads to the concept of "9s" as a measure of availability as well as to some useful associated rules. We will explore these topics as well.

If your eyes glaze over at some of the algebra, skip the body of this article and go right to the end to read the simple but important availability rules that come out of the analysis.

The Availability Relationship

To simplify things a bit, we will dive into this topic in small increments. First, we will look at a simple two-node system with one spare. Next, we will look at a multinode system with one spare. Finally, we will look at a multinode system with multiple spares.

First, let us define the following terms:

A is the probability that the system will be up (its availability).
F is the probability that the system will be down.
a is the availability of a node.
n is the number of nodes in the system.
s is the number of spare nodes in the system.
f is the number of ways that all of the spares plus one other node can fail (that is, the number of node failures that will cause the system to fail).

Clearly, $A = 1-F$.

---

[1] A more extensive derivation of the availability equation may be found in the book entitled *Breaking the Availability Barrier: Survivable Systems for Enterprise Computing*, by Dr. Bill Highleyman, Paul J. Holenstein, and Dr. Bruce Holenstein, published by AuthorHouse; 2004
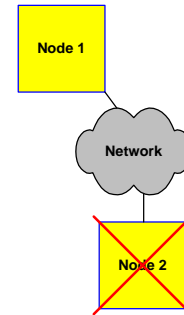
Dual Node, Single Spare

We first consider an active/active system with two nodes, only one of which need be operational for the system to be considered available. The availability of a single node is *a*. This is the probability that the node will be up. Therefore, the probability that it will be down is (1-*a*). The probability that both nodes will be down is $(1-a)^2$. This is the probability *F* for the failure of the system:

$$F = (1-a)^2$$

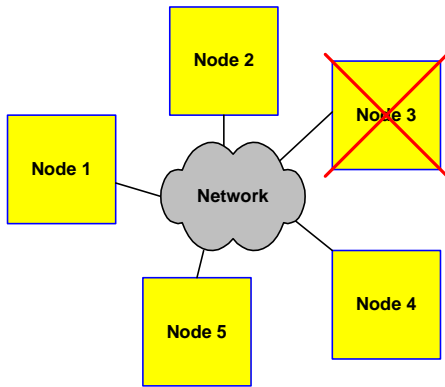Thus, the availability of the system, *A*, is

$$A = 1-F = 1-(1-a)^2$$

For instance, if the node availability is .99, the probability that it will be down is .01. The probability that both nodes will be down, thus causing a system failure, is $.01^2$, or .0001. Thus, the system availability is (1 – 0001), or .9999. The system has an availability of four 9s.

Multiple Nodes, Single Spare

In a multinode system with one spare, it will still take only the failure of two nodes to take down the system. However, there are many ways that we can have a failure of two nodes. For instance, if there are five nodes in the system, there are ten ways that two nodes can fail (count them). Thus, in this case, the number of failure modes, *f*, is ten.

Dual Node, Single Spare

Multinode, Single Spare

In general, if there are *n* nodes, there are *n* ways that one node can fail. Given a single node failure, there are (*n*-1) ways that a second node can fail. However, this reasoning has counted each failure mode twice; e.g., node 2 followed by node 5 and node 5 followed by node 2. Therefore, for an *n* node system, the number of failure modes, *f*, is

$$f = \frac{n(n-1)}{2}$$

The probability of failure of the system is the probability that any two nodes will fail times the number of ways that two nodes can fail:

$$F = \frac{n(n-1)}{2}(1-a)^2$$

Thus, the availability of the system is

$$A = 1-F = 1-\frac{n(n-1)}{2}(1-a)^2$$

For instance, consider a five-node system. Using our previous example for nodes with an availability of .99, the number of failure modes is ten; and the availability of a five-node system is $[1 - 10(.01)^2]$, or .999. This is three 9s of availability.
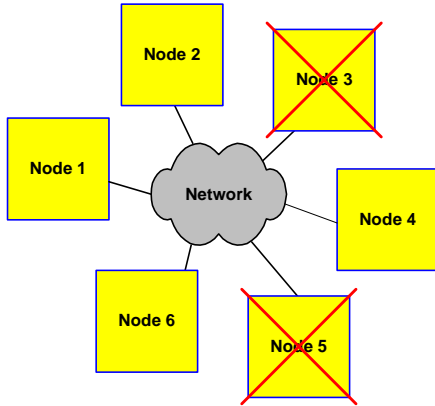
Note that this is less than the availability of four 9s for the two-node system. Here is an important rule to note:

As an active/active application network gets larger with no increased sparing, its availability goes down. This is because of the increase in the number of failure modes.

We talk about additional levels of sparing next.

Multiple Nodes, Multiple Spares

The next step to consider is the impact of having more than one spare. We have defined *s* as the number of spare nodes in the network. Therefore, it will take the loss of *s*+1 nodes to take down the network.

Multinode, Multiple Spares

Since the probability of losing one node is (1-*a*), the probability of losing *s*+1 nodes is (1-*a*)$^{s+1}$. Note that if there is a single spare (*s* = 1), this reduces to (1-*a*)^{2}, as used above for a single spared network.

The next question is how many ways are there for *s*+1 nodes to fail? This is the number of failure modes, *f*, for the network and is the number of ways that *s*+1 nodes out of *n* nodes can fail. The number of such combinations is given by the rather imposing expression

$$f = \frac{n!}{(s+1)!(n-s-1)!}$$

The symbol "!" means "factorial." For instance, 3! is 3x2x1 = 6.

Thus, the probability of failure for an *n* node system with *s* spares is

$$F = f(1-a)^{s+1}$$

and its availability is

$$A = 1 - f(1-a)^{s+1}$$

where f is given above for *n* nodes and *s* spares. This is the relationship that we promised you at the beginning of this article.

As an example, if there are two spares (*s* = 2), the number of failure modes, *f*, becomes

$$f = \frac{n(n-1)(n-2)}{6}$$ for two spares.

Consider a six-node system with two spares. That is, at least four nodes must be up and running in order for the system to be operational. Then *f* = 20 (that's right – there are twenty ways that three nodes out of six can fail – count them). Using our example above of a nodal availability of .99, the probability of failure of the system, *F,* is 20x(.01)^{3}, or .00002. This yields a system availability of .99998, or almost five 9s. This compares to the similar singly-spared system above that had an availability of three 9s.

What About Those 9s?

The measure of 9s for availability is a logarithmic measure. It is like the Richter scale. An earthquake of magnitude 6 is ten times more powerful than a magnitude 5 earthquake. Likewise,

a system with an availability of four 9s is ten times more reliable than a system with an availability of three nines.

Now there's a clue. Let's take the logarithm of the failure probability of an active/active system:

$$9s = -\log_{10}(F) = -\log_{10}[f(1-a)^{s+1}] = -\log_{10}(f) - (s+1)\log_{10}(1-a)$$

(We needed to add the minus sign since $A = 1-F$.)

For instance, consider a five-node system ($n = 5$) with one spare ($s = 1$). In this case, f is equal to 10. Log(f) is log(10), which is 1. Further, assuming that the nodal availability, a, is .99, log(1-a) = log(.01), which is -2. The system availability measured in 9s is then -1+2X2, or three 9s (.999).

Notice that the log of the nodal failure probability, (1-a), is the measure of that node's availability in 9s. For instance, if the node has an availability of .01 = 10^{-2}, the log of its availability is -2, equivalent to two nines of availability.

Noting that the log of the nodal availability is multiplied by (s+1), we come up with the following important rule:

> *Adding a spare node adds the number of nines associated with that node to the system availability but reduced by the increase in failure modes.*

That is, adding an additional spare node adds the number of 9s of that node to the system availability – almost. This improvement in availability is reduced a bit by the increase in the number of failure modes in the system. More nodes mean more failure modes.

Our Logo

Let us now return to our logo. It represents the failure probability of an active/active system with one spare node – the most common of active/active systems. The first "f" represents the number of failure modes – the number of ways that two nodes can fail in the system. The second "f" represents the probability of failure of any two nodes. Thus, the probability of failure of the system is ff^2 (if you will forgive the stylization).

Rules of Availability

We leave you with the following rules for the availability of an active/active application network:

> 1. *The more nodes in an active/active network, the less reliable it is for a given sparing level. This is because of the increase in failure modes.*
>
> 2. *Adding a spare node to an active/active network adds the number of nines associated with that node to the system availability – almost. The improved system availability is reduced somewhat by the increase in failure modes.*

As an example, using the relations we derived above, if the availability of a node is .99, the availability of a two-node system in which only one node is required to be up (i.e., there is one spare) is .9999, or four 9s.

If we add a third node to the system, maintaining still only one spare, the systems availability drops to .9994 (a little over three 9s).

If, however, that third node was an additional spare node, the system availability becomes .999994 (a little over five 9s).

We hope that this leaves you with a feeling of the impact on active/active system availability as a function of system size and its sparing level.