Stretching A Wolfpack Cluster Of Servers For Disaster Tolerance

Dick Wilkins
Program Manager
Hewlett-Packard Co.
Redmond, WA
dick_wilkins@hp.com

Motivation

- WWW access has made many businesses 24 by 7 operations.
- Critical sales and support functions commonly take place online.
- Downtime is likely result in losses of many millions of dollars an hour.
- A disaster that results in a week or two of critical application unavailability is likely to result in business failure.

Disaster Causes

- Environmental (fire, tornado, earthquake, flood, major power outage)
- Civil unrest, terrorist actions
- Major operational errors
- Etc.

Traditional Recovery Methods

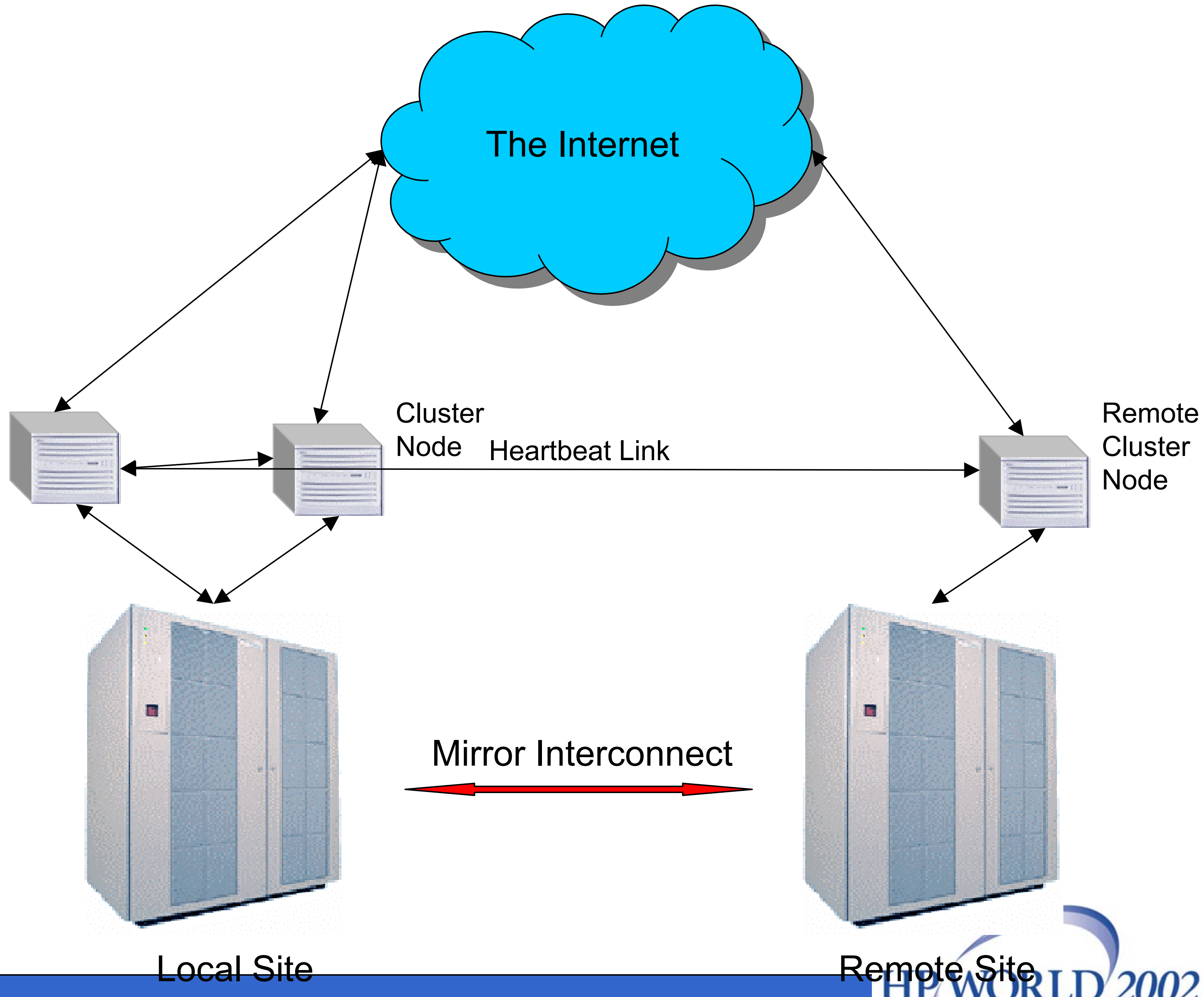
- Recovery from offsite backups
- Warm/hot site restoration
- Remote mirroring
- Clusters (HA only, not disaster tolerant)
- Fault tolerant architectures (ditto)

How to obtain Disaster Tolerance

Combine solutions

- In addition to routine backups, etc.

- Use remote mirroring
- . . . . and clustering
- . . . . and place cluster nodes at each site


The Internet
Cluster Node
Heartbeat Link
Remote Cluster Node
Mirror Interconnect
Local Site
Remote Site

What makes this hard?

1) Mirroring is unidirectional
2) The quorum disk is special
3) Total loss of communications between sites can present protocol problems

Unidirectional mirroring

- Mirroring was not designed as a general purpose communications path.
- It just moves data from primary to secondary disk copies.
- MSCS is a “shared nothing” clustering system (vs. “shared disk”).
- When ownership of a disk is changed, mirror direction may need to be swapped.

The Quorum disk is special

- Used as a tie breaker to prevent “split-brain”
  - Where two (or more) nodes think the others are down and try to provide the same service
- Low level SCSI commands may be delivered at any time
- Therefore must accurately follow the SCSI Specification semantics for Reserve, Release and Bus/Device Reset

Preventing Split-Brain

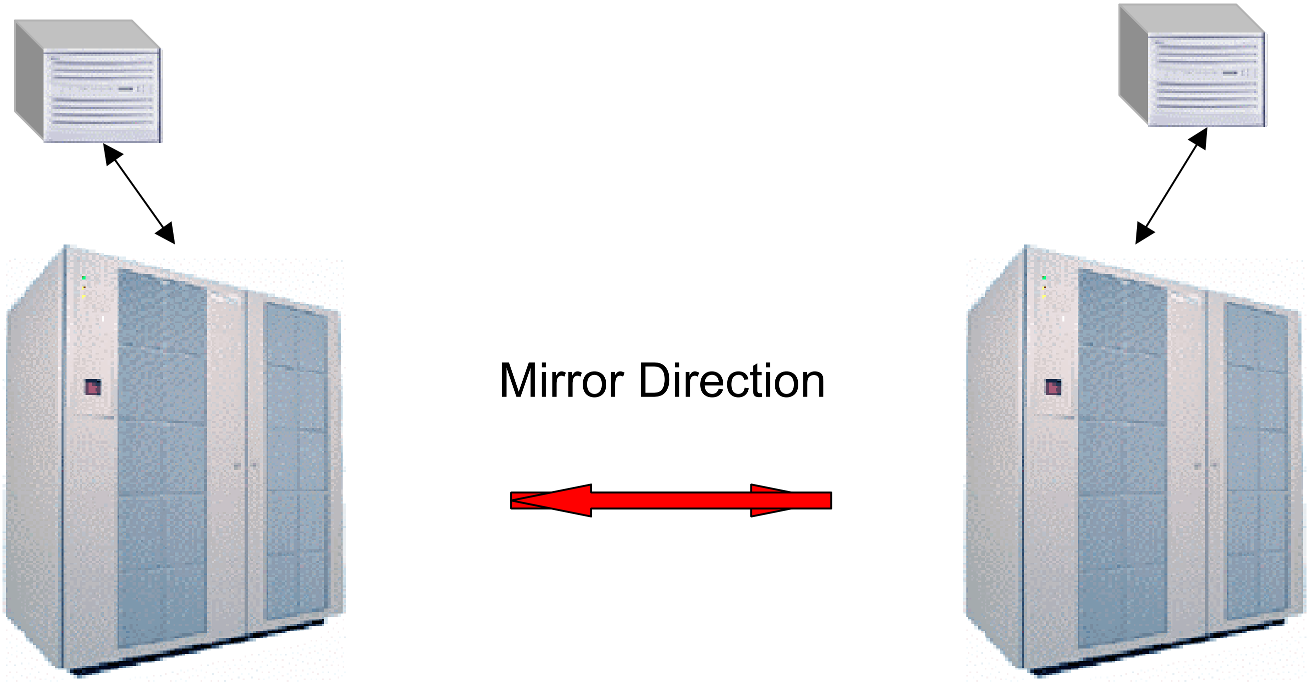
- When all inter-site communications are lost…
- It’s hard to run a protocol between sites to decide to move control of the cluster
    1. The remote site may be down (“disaster”)
    2. It may have lost network connectivity
    3. It may be up and providing service
- How to decide?

Wolfpack Background

- Two approaches to application data disks
  - “Shared Data”, all nodes see all disks
  - “Shared Nothing”, one node at a time
- Applications and their “resources” are packaged into “groups”
  - Disk(s)
  - IP address
  - Network name
  - The application itself

The “Resource DLL”

- Inserted into group bring up prior to disks
- Checks to see the mirroring status and direction
- Swap direction it if needed



Other resource DLL functions

- The disk array allows for sync and async I/O to mirror volumes (LUN by LUN)
- Numerous options are available based on data-safety requirements
- The DLL must preserve and protect the selected level of data-safety (600+ rules)
- User supplied scripts can be run
- GUI is provided to setup and manage

Dealing with the Quorum disk

- We want the quorum disk mirrored in two places, but…
- It must continue to operate as a single drive in the face of failures
- We didn't want to create more communications paths to rely on

Disk control protocol (simplified)

1. A node wanting to control a disk issues a Bus/Device Reset to the disk
2. It waits INTERVAL+ time
3. It issues a SCSI Reserve command
4. If the reserve fails, it assumes another node has defended its ownership
5. If it succeeds, it is the new disk owner and starts routine re-reservations every INTERVAL time

Disk protocol (cont.)

- MSCS also uses reads/write to probe the current reservation state
- We place a “filter driver” in the SCSI stack to intercept SCSI commands for the quorum disk
- A service program cooperates with the filter driver to insure correct behavior of the quorum disk pair
- We use three special disk pairs to “communicate” between servers and sites

The special disks

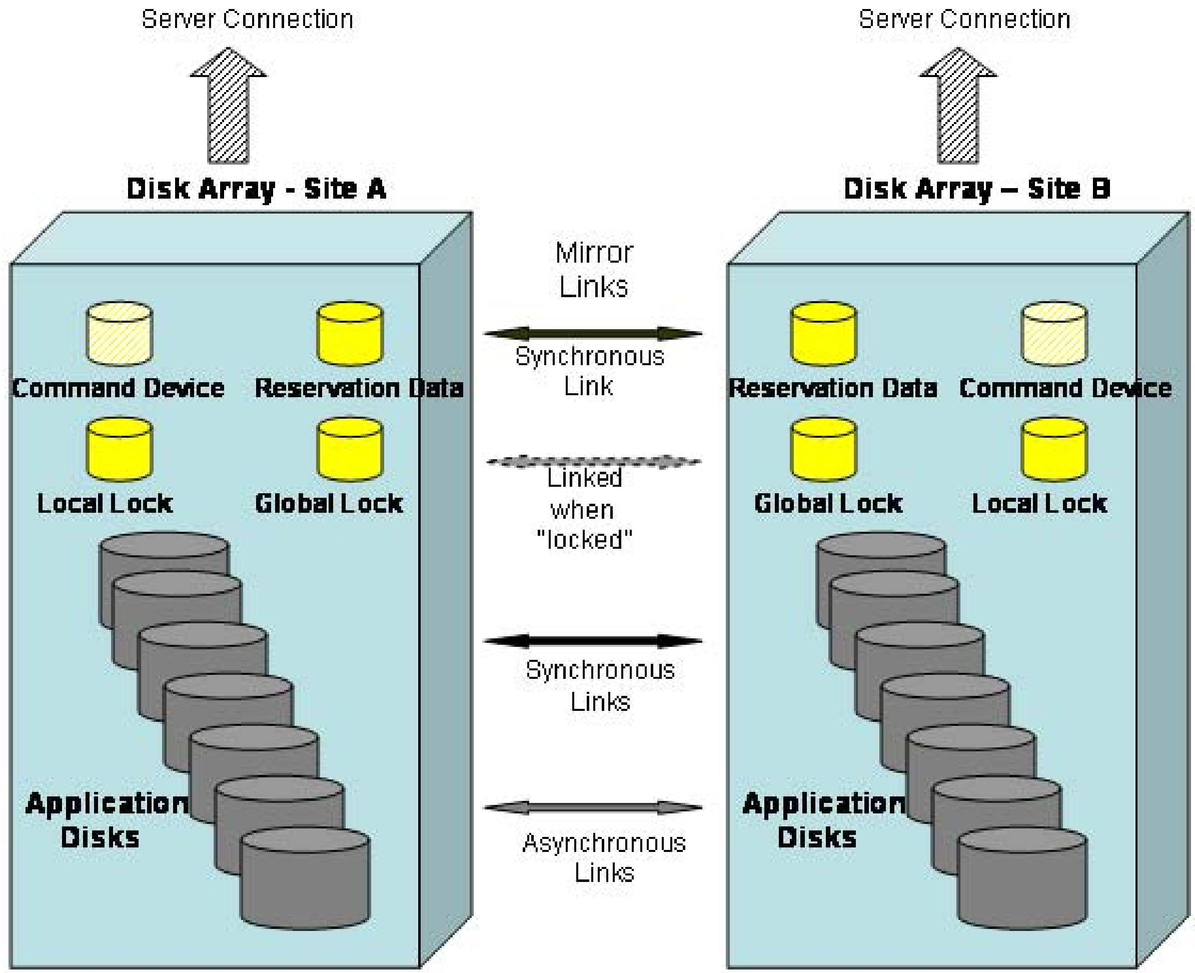
1. The first disk pair, synchronously mirrored, holds the current reservation state of the quorum disk
2. A second disk “pair” (not mirrored) is used to hold a exclusionary lock between local nodes
3. The third disk pair is used as a global lock. It’s pair state and direction indicates the ownership of the global lock.

The special disks (cont.)

- These disks are small (the smallest that can be created in the array)
- They are not part of the cluster configuration
- They are not assigned drive letters
- They do not contain file systems

- Also required are “Command Device” disks on each array used for array control commands

The Disks


Server Connection
Server Connection
Disk Array - Site A
Disk Array – Site B
Mirror Links
Command Device
Reservation Data
Synchronous Link
Reservation Data
Command Device
Local Lock
Global Lock
Linked when "locked"
Global Lock
Local Lock
Synchronous Links
Application Disks
Application Disks
Asynchronous Links

One type of failures is hard!

- If all communications are lost between sites you need help to recover
- You need the quorum, but we "stretched" it
- An external arbitrator can help
  - Located with the application users
  - Is prepared to allow one and only one site proceed with cluster ops
- Arbiter location available from Active Directory (or config files)

How it works

- For the node that currently holds the quorum device
  - It detects the loss of communications
  - It creates a process
  - The process contacts the arbitrator
  - If it makes contact, it knows all is OK and exits
  - If it is unable to make contact, it assumes the communications failure has isolated it and shuts down clustering (also assuming the other site will restart clustering)

How it works (cont)

- When MSCS asks a node to take control of the quorum disk and the normal protocol finds a communications failure
  - It causes clustering to shut down
  - It creates a new process to see if recovery is needed
  - That process contacts the arbitrator
  - If the arbitrator tells it to proceed, it cleans up the metadata and restarts clustering locally, if not, it just exits

How the arbitrator decides

- Upon request for arbitration, the arbitrator process attempts to contact the cluster’s general IP address
  - If the “cluster” responds (a operating node that holds the quorum device) it reports that the cluster is running
  - If it cannot contact the cluster’s IP, it reports that the cluster is down or isolated
  - If multiple nodes request this service, it only responds to the first

Software Installation

- Hardware configuration is done first
- Typical Microsoft Installer/InstallShield script setup is used
- One node at a time (quorum is moved ahead of each node's setup)
- Same script installs the resource DLL
- Also installs the arbitrator on a node external to the cluster

Summary

- A stretched cluster significantly increases disaster tolerance
- It is possible to quickly and automatically recover from major failures
- Inter-site link technology doesn’t matter
- Cost is a small increment when clustering and distance mirroring are already present
- This implementation is hardware specific but the technology is portable to other distance mirroring solutions
- This implementation is Microsoft WHQL certified

Future Work

- New arrays (distance mirroring is all that is necessary)
- Windows .NET and Win64
- Redundant Arbitrator
- New quorum technologies from Microsoft (SCSI persistent reservations, new quorum algorithms)
- Transportability to Unix clusters?