

Configuring & Tuning Cluster Networks

- Node connectivity
- Node visibility
- Networking Services
- Security
- Performance Enhancement

DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 1

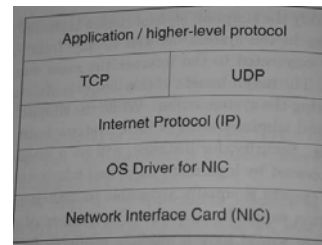
Network Designs

- Impact of Network Design
 - Security from outside attack
 - Usability
 - Single application : tune for it
 - General purpose : intuitiveness and ease of use
 - Application performance
- Network Designs
 - Fully Internet connected – all nodes visible from outside
 - Only front-end machine Internet visible
 - User logs on front-end and can access all nodes
 - Only front-end machine Internet visible
 - Can compile and test there
 - Computational nodes only accessible through scheduler
- Single System Image

DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 2

Internet Protocol Stack and Parameters

- NIC/OS Driver
 - Maximum Amount of Data in Frame : MTU (Maximum Transmission Unit)
 - Ethernet 1500 bytes
 - GE Jumbo Frame 9000 bytes
- IP (Internet Protocol)
 - Header includes length of datagram (incl header) field
 - 16 bits field -> 65,535 bytes
 - IP datagrams larger than MTU are fragmented and reassembled by receiver
 - Datagrams may also be fragmented by intermediate routers if MTU on outgoing link is smaller
- IP only identifies machines
 - Unreliable, unordered, connectionless (stateless)



DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 3

UDP and TCP

- User Datagram Protocol (UDP)
 - Unreliable, unordered, connectionless (stateless)
 - Allows identification of end-points of communication (ports)
 - Multiple flows per machine
- Transmission Control Protocol (TCP)
 - Reliable communication, bidirectional connection, state maintained
 - Data is segmented and each segment, plus TCP header, embedded in an IP datagram.
 - Maximum Segment Size (MSS) is size of segment
 - To avoid segmentation MSS is advertised as
 - $MSS = MTU \text{ of network} - \text{size of } (TCP + IP \text{ headers})$
 - On LAN MTU is MTU of NIC

DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 4

TCP

- On WAN determining MSS more difficult
 - MTU of all intermediate networks not known
 - TCP/IP assume MTU of 576 bytes unless sysadm specifies
 - Wide area MTU discovery is then used to determine the maximum MTU acceptable to all networks
- TCP reliability uses
 - positive acknowledgements (ACKs),
 - sliding windows to permit multiple unACKed segments
 - data buffering : receiver can advertise amount available
 - Provides flow control

DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 5

IP addresses

- 32 bits (4 octets) usually written 131.123.42.51
- Consists of network and host parts
- Netmask is used to indicate network part
 - AND with netmask
 - Ex: 255.255.255.0 means first 3 octets network, last host
- Reserved Host Addresses
 - 0 network itself
 - 255 broadcast
- Routers use network part of IP address to choose network link
- Sender must know address of local router – gateway router

DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 6

Non-Routable IP Addresses

- Reserved for private networks:
 - 10.0.0.0 - 10.255.255.255
 - 172.16.0.0 – 173.31.255.255
 - 192.168.0.0 – 192.168.255.255
- Can be used for clusters that
 - Don't need to communicate on the Internet
 - Are behind a firewall that does NAT (Network Address Translation)

DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 7

Hostnames

- Unique name - really an interface name
 - long version including domain (f01.fianna.cs.kent.edu)
 - Short version (f01)
- Recommendation for Clusters
 - Give nodes names like f01 – f32 and address 192.168.1.1 to 192.168.1.32
 - Gateway must follow nodes – try maximum available host address e.g. 192.168.1.254

DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 8

Name Resolution

- /etc/hosts
 - File with list of IP address, long and short hostnames, one per line
 - Also loopback device address
127.0.0.1 localhost.fianna.cs.kent.edu localhost
- Can have a master copy on one node
 - Push to other nodes using e.g. scp (see 5.3.5)
 - Use cluster administration tools (see Ch 6)
- Alternatives
 - NIS (Network Information Service)
 - Also can do /etc/passwd, /etc/group etc
 - DNS (Domain Name Service)

DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 9

File Sharing

- NFS (Network File System)
- Allows installation of software package on one machine and use on all
- Allows users to have access to programs and data on all machines
- Also useful on clusters
 - Also ensure shared libraries available on every node
- Users need to take care if they are writing from the processes executing on each node to ensure they do not overwrite data

DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 10

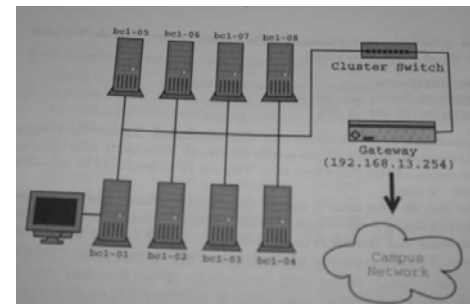
Remote Access

- telnet, rlogin, rsh, rcp
 - telnet security issue – send data including passwords in clear
- rlogin, rsh, rcp
 - Can use host based authentication
 - File of hosts authorized to connect without password
 - /etc/hosts.equiv
 - ~/.rhosts
 - Need to control physical access to network and to these files
- Secure shell ssh, slogin, scp, sftp
 - Public-private key based authentication
 - Can verify that host is expected host using keys
 - More later

DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 11

Cluster Configuration Example RH9

- Network address : 192.168.13.0
- Netmask: 255.255.255.0
- Gateway: 192.168.13.254



DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 12

Initial Install

- Default workstation install
- Add NIS
- Will run NIS and NFS servers on bc1-01
- Assume DNS server on 192.168.1.1 for hosts outside cluster network
- **Hostname and gateway**
 - /etc/sysconfig/network
 - NETWORKING=yes
 - HOSTNAME=bc1-01.phy.myu.edu
 - GATEWAY=192.168.13.254
- Use long name
- Changes only take effect on reboot

DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 13

Network Interface Configuration

- Disable the interface using
 - ifdown eth0
- /etc/sysconf/network-scripts/ifcfg-eth0

```
DEVICE=eth0
BOOTPROTO=static
IPADDR= 192.168.13.1
NETMASK=255.255.255.0
GATEWAY= 192.168.13.0
BROADCAST= 192.168.13.255
ONBOOT=yes
```
- Enable the interface using
 - ifup eth0

DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 14

Name Resolution

- Enter names of all hosts and localhost in /etc/hosts on bc1-01
- Use NIS to make available
- On other hosts only add localhost entry
- Nameserver
- /etc/resolv.conf

```
nameserver 192.168.1.1
search phy.myu.edu
```
- Search path is appended to short names

DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 15

Accounts

- Use adduser
 - Create /etc/passwd and /etc/shadow entries
 - Adds group for user in /etc/group
 - Creates home directory
- Run passwd to set password
- To create on other nodes
 - Do it all again, and again and ...
 - Automate with scp, ssh and perhaps NIS
- Note NIS and NFS treat root differently
 - NIS does not publish root info
 - NFS does not mount the root's home directory
 - Enhanced security

DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 16

Packet Filtering

- Default in RH9 blocks ssh, NFS, NIS
- Need to change
- /etc/sysconfig/iptables
 - Insert before first –A INPUT line
 - A INPUT –p tcp –m tcp –dport 22 - - syn –j ACCEPT
 - A INPUT –p tcp –s 192.168.13.0/24 –j ACCEPT
 - A INPUT –p udp –s 192.168.13.0/24 –j ACCEPT
- Allows ssh connections from everywhere and all tcp and udp packets from any cluster machines
- Execute /etc/rc.d/init.d/iptables restart

DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 17

Secure Shell

- RH9 has OpenSSH package
 - Ssh, scp, sshd
 - With packet filter rules, root can replicate files over nodes
 - Need NIS, NFS to enable for non-root users
 - Sshd generates authentication keys when first started
 - Client needs to know *public key* to validate
 - First connection to a host ssh asks user to authorize
 - Then stores name and public key in ~/.ssh/known_hosts
 - Alternatively sysadm can create system wide list in /etc/ssh/ssh_known_hosts
 - Can be done automatically using ssh-keyscan
 - Uses file (say scanhosts) with IP address, all names and addresses
- ssh-keyscan –t rsa,dsa,rsa1 –f scanhosts > /etc/ssh/ssh_known_hosts

DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 18

Ssh – user authentication

- Users must still be authenticated
 - Default by password
 - Alternative: host based authentication
 - Need to change configuration in /etc/ssh/sshd_config
 - HostbasedAuthentication yes
 - IgnoreUserKnownHosts yes
 - IgnoreRhosts no
 - Last allows authorization file ~/.shosts to be used
 - Needed for root for which /etc/ssh/shosts.equiv not used
 - Both need to contain list of full hostnames of all nodes
 - To make effective need to restart sshd
 - /etc/rc.d/init.d/sshd restart
 - The client config file /etc/ssh/ssh_config needs to have
 - HostbasedAuthentication yes
 - added

DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 19

Ssh – key generation

- Users can generate authentication keys using
 - ssh-keygen –t rsa
- Public and private keys put in
 - ~/.ssh/id_rsa.pub and ~/.ssh/id_rsa
- Adding public key to ~/.ssh/authorized_keys on any machine allows password free access
- Keys can be generated with or without a passphrase to protect the private key
 - If generated with passphrase, this must be entered for each remote operation
 - Alternative: use ssh-agent to manage private pass keys and ssh-add to add them to set of managed keys (see book)
 - See www.openssh.org for more information

DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 20

Network Information Service (NIS)

- Originally a Sun product (yellow-pages)
- On each node add to `/etc/sysconfig/network`
`NISDOMAIN=bc1.phy.myu.edu`
- NIS domain being administered
- NIS server run on bc1-01. Add to `/var/yp/securenets`
host 127.0.0.1
255.255.255.0 192.168.13.0
- Replace all: line in `/var/yp/Makefile` with list to be exported
all: passwd group hosts networks services protocols rpc

DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 21

NIS (ctd.)

- Need to build maps (databases) of services
`echo "loopback 127" >> /etc/networks`
`/etc/rc.d/init.d/ypserv start`
`/etc/rc.d/init.d/ypxfrd start`
`/etc/rc.d/init.d/yppasswdd start`
`cd /var/yp`
`make`
- Need to have start automatically at boot time
`chkconfig --level 345 ypserv on (etc)`
- Configure ypbind to run on all clients
`/etc.rc.d/init.d/ypbind start`
`chkconfig --add ypbind`
`chkconfig --level 345 ypbind on`

DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 22

NIS (ctd.)

- To make ophys use NIS update `/etc/nsswitch.conf`
`passwd: files nis`
`hosts: files nis dns etc`
- When new accounts or groups are added need to rebuild maps
- To make changed passwords available on all nodes use `yppasswd`
- If `sshd` is started before `ypbind` then `ssh` will not use NIS, so need to restart `sshd`

DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 23

Network File System

- On bc1-01 export users home directories by adding to `/etc/exports`
`/home 192.168.13.0/24(rw)`
- Enable and start NFS
`chkconfig --level 345 nfs on`
`/etc/rc.d/init.d/nfs start`
- Configure client nodes to mount `/home` directories from bc1-01 - add to `/etc/fstab`
`192.168.13.1:/home /home nfs rw,hard,intr,bg,rsize=8192,wsiz=8192 0 0`
- To cause remote file system to be mounted
`/etc/rc.d/init.d/netfs restart`

DiSCoV KENT STATE 25 April 2006 Paul A. Farrell
Cluster Computing 24

Scripting It

- For larger cluster need to automate process
- Can use, for example, kickstart in RH9
- A config file is automatically generated during installation
- This can be used to create a kickstart file for other nodes
- This is copied to a floppy and used together with the RH CD to create the other nodes
- See book for details