IPCI
(IP as a Cluster Interconnect)

Akila B.
OpenVMS Engineering
Germany, Sep, 2009

Europe 2009 Technical Update Days

Agenda

- Introduction to OpenVMS Clustering Technology
- Disaster Tolerant Clusters with OpenVMS
- Cluster Communication Architecture
- Need for IPCI solution
- IPCI Solution details
- Salient features of IPCI
- Customer advantage

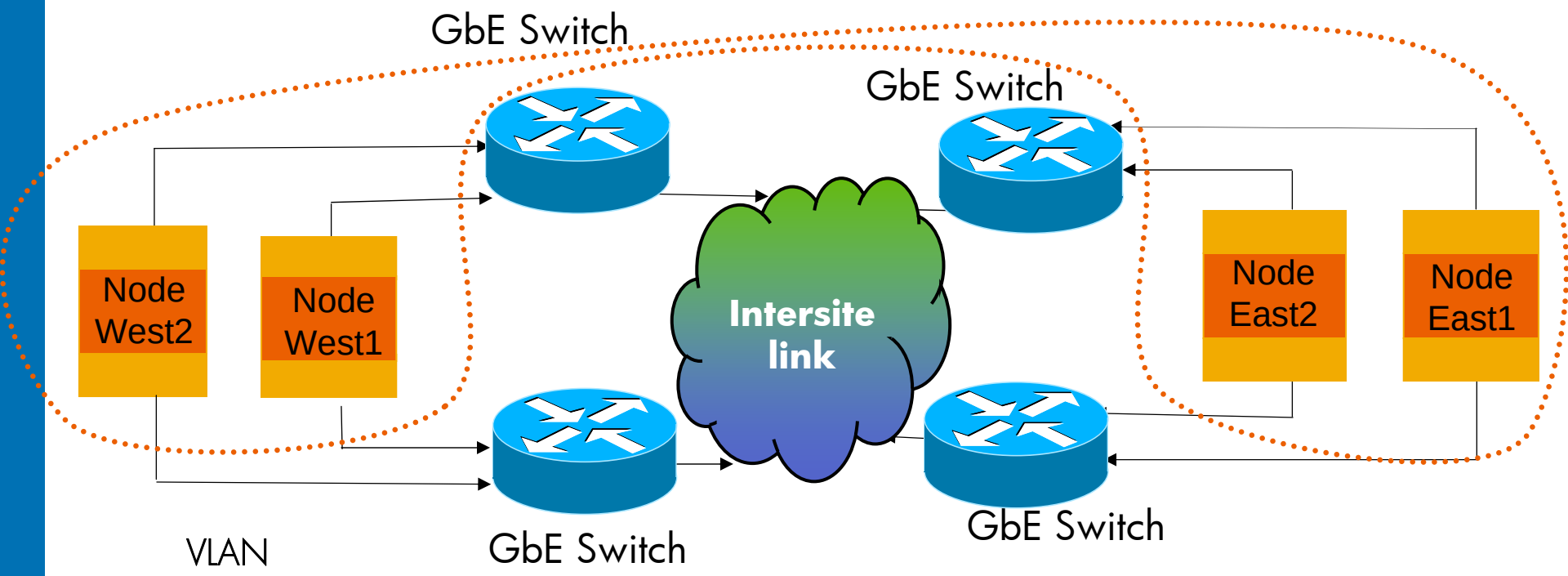
Introduction to OpenVMS Clustering

- Reliability
  - HP OpenVMS is known as "**gold standard**" in disaster tolerance
- Scalability
  - Qualified for 96 nodes and also mixed architecture configuration (IPF-Alpha, Alpha-VAX)
  - OpenVMS supports clusters with up to 500 miles apart
- Manageability
  - Shared-Everything model with Cluster wide file system
  - Single System Image, Cluster wide management facility

Disaster Tolerant (DT) Clusters with OpenVMS

- High Availability
  - Can be configured to withstand multiple points of failure
  - Supports configurations with applications active at multiple sites with automatic load balancing and failover
  - Automatic and transparent cluster failover
  - Host Based Volume Shadowing
- Uses industry standard storage and interconnects (SAN, FCIP, Ethernet)
- Distributed Lock Management
- Quorum site to avoid partitioned cluster.

Disaster Tolerant /Long Distance OpenVMS Clusters



LAN bridging/Extended LANs using switches

Nodes East1, East2, West1, West2 belong to same LAN/VLAN

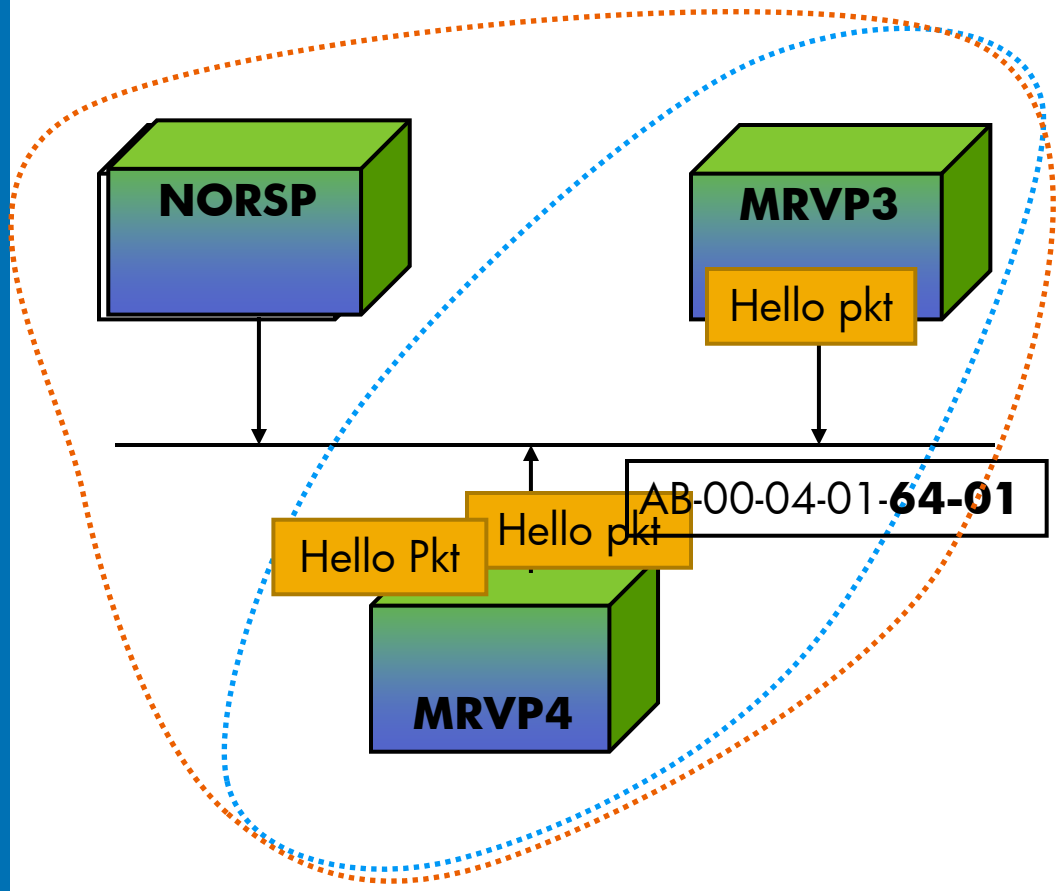
Cluster traffic is non-routable

- Introduction to OpenVMS Clustering Technology
- Disaster Tolerant Clusters with OpenVMS
- **Cluster Communication Architecture**
- Need for IPCI solution
- IPCI Solution details
- Salient features of IPCI
- Customer advantage

Current OpenVMS Cluster Interconnect Solution

- SCA (alias SCS) – System Communication architecture
  - Cluster communication protocol
- Cluster Interconnect
  - Alpha : LAN, Memory Channel, Shared Memory, CI
  - IA64 (Integrity) :LAN

OpenVMS Cluster Communication in LAN



MRVP3 and MRVP4 are in cluster

Cluster group number 100

The multicast address is

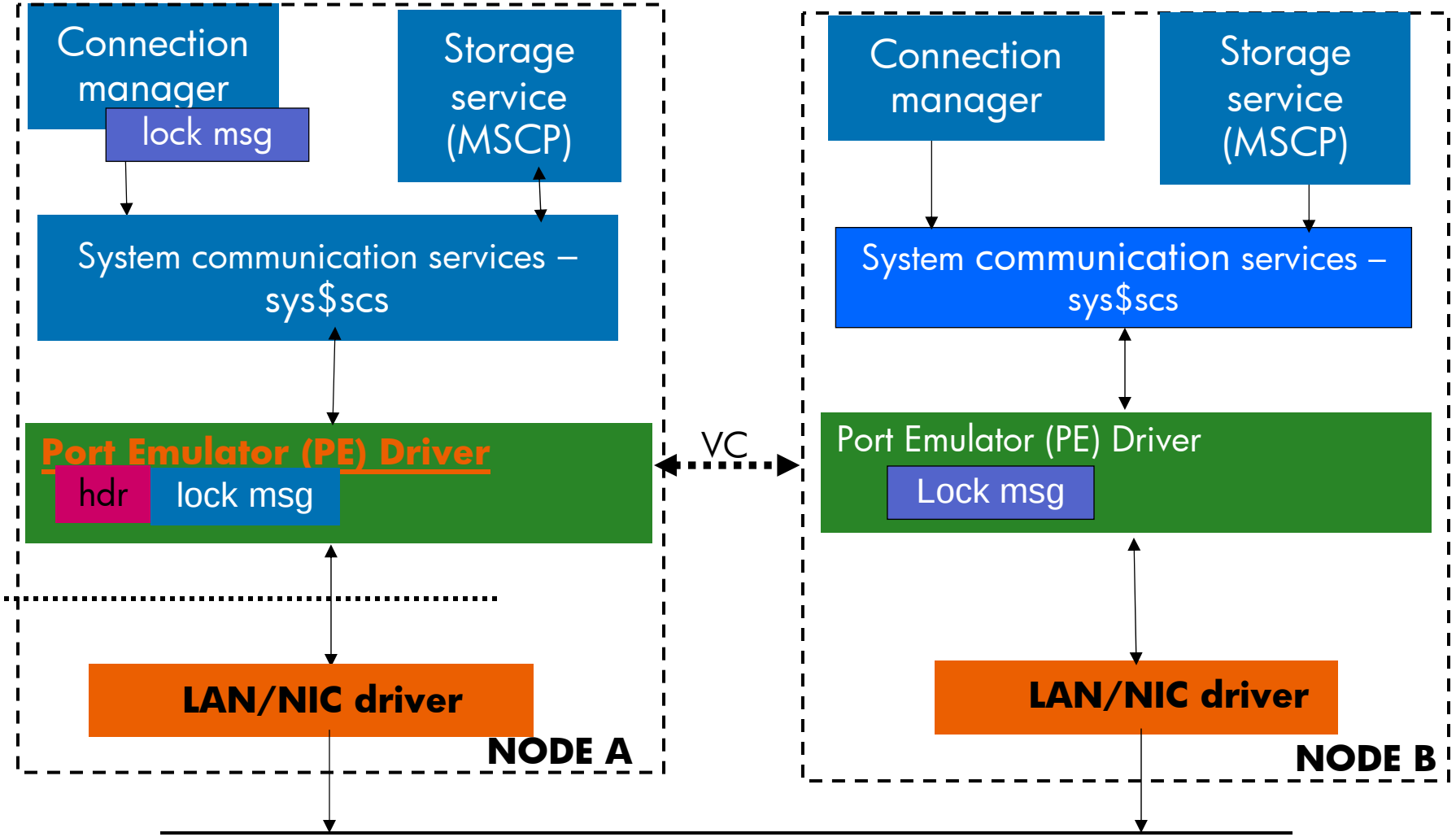
AB-00-04-01-00-01 +

the cluster group number

AB-00-04-01-**64-01**

Configure NORSP as cluster and reboot NORSP

OpenVMS Cluster Communication Architecture



- VC – Virtual Circuit consists of LAN Channels

Port Emulator (PE) Driver

- Component that implements OpenVMS Cluster communications in LAN (NISCA aka Network Interconnect System Communication Architecture)
- Transmits and receives datagram, sequenced messages and block transfer of data
- Multilayered architecture
- Consists of Transport layer, Channel Control Layer and Data Exchange layer

Current Solution - Summary

- OpenVMS customers use LAN interconnect primarily for Cluster communications
- NISCA protocol which is used for Cluster communication is LAN based
- OpenVMS Nodes should be in same LAN or VLAN for cluster communications
- Bridging and Extended LAN used for inter-site cluster communication

- Introduction to OpenVMS Clustering Technology
- Disaster Tolerant Clusters with OpenVMS
- Cluster Communication Architecture
- **Need for IPCI solution**
- IPCI Solution details
- Salient features of IPCI
- Customer advantage

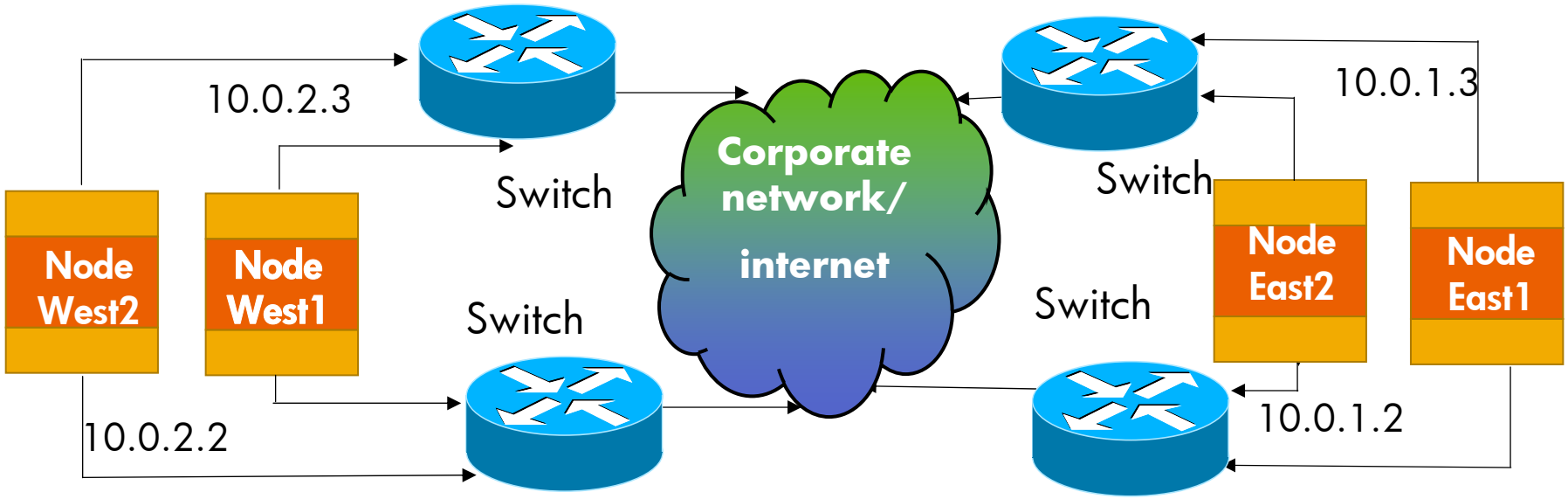
IP as a Cluster Interconnect (IPCI)

- Ability to use IP for OpenVMS clusters communications
- Enable OpenVMS cluster communication module (PE driver) to use IP stack
- Coexist with LAN interconnect for Cluster communication
- IP unicast and optionally IP multicast for node discovery

Motivation for IPCI

- Network switch during higher loads give priority to IP traffic than cluster (SCS) traffic
- Cluster instability during periods of heavy IP usage
- Ability to work without need of special network setting/devices
- Corporate policies restricting non-IP protocols
- Switch vendor dropping support for bridging
- IP is de-facto industry standard
- Leveraging benefits of improvements in IP technology

Cluster using IPCI

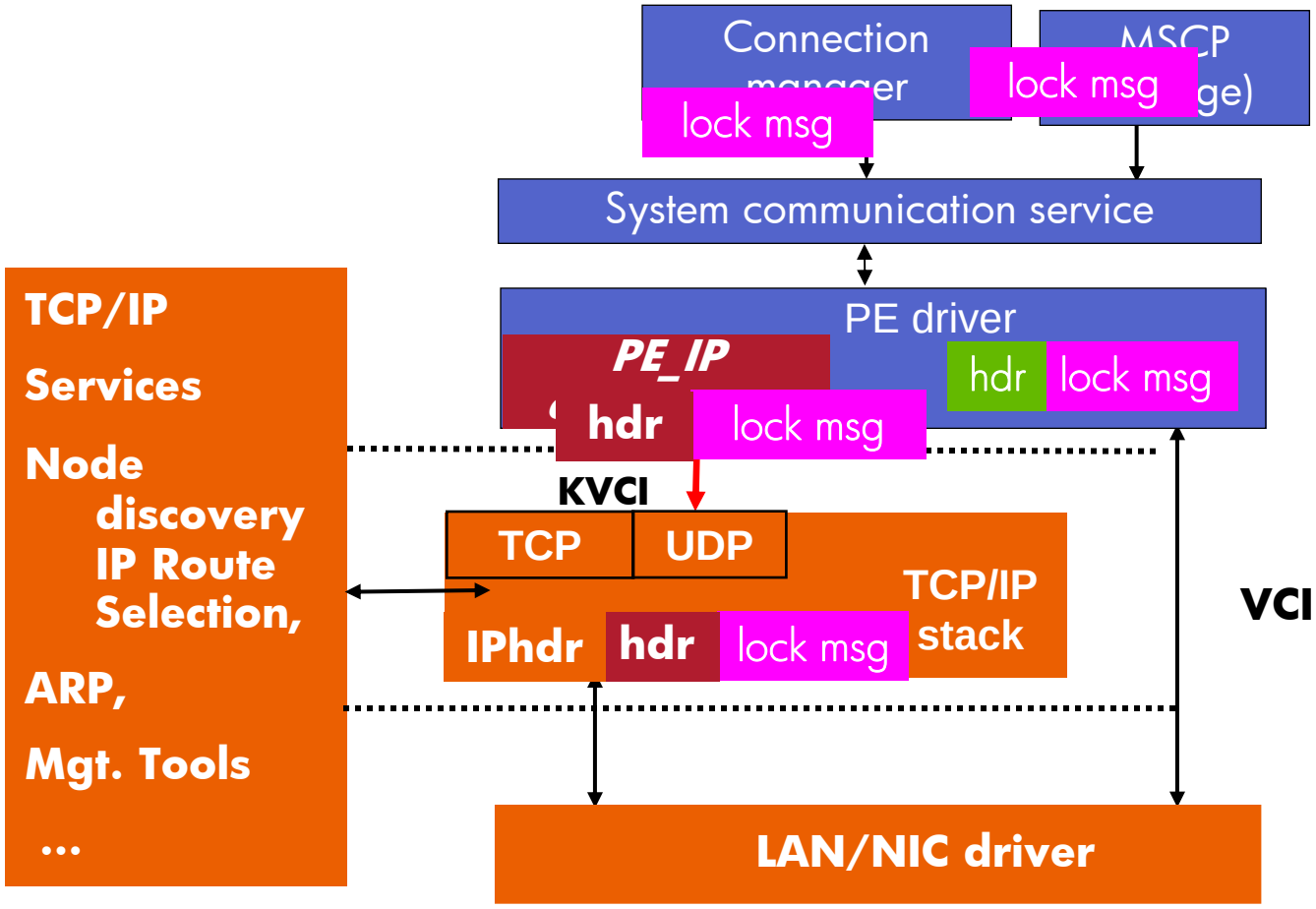


• Node East1, East2, West1,West2 can be part of the same or **different LAN** for cluster communications using IPCI. ***Cluster traffic is routable***

•East1 and West 2 has a Virtual Circuit (VC) VC consists of IP channels for SCS traffic

- Introduction to OpenVMS Clustering Technology
- Disaster Tolerant Clusters with OpenVMS
- Cluster Communication Architecture
- Need for IPCI solution
- **IPCI Solution details**
- Salient features of IPCI
- Customer advantage

IPCI solution – PE driver over UDP



Existing Cluster Component

New Component-Component interaction

NEW PEdriver component

Existing VMS/TCP/IP component

IPCI – Major Parts

- PEdriver over UDP Support
- TCP/IP Services boot time loading & initialization
- Availability Manager Support

PE driver over UDP

- The IP UDP service has the same packet delivery characteristics as 802 LANs
- PEdriver uses the IP based UDP datagram service as another LAN device
- Only the lowest layer of PEdriver has extension to locate and connect to the TCP/IP
- Uses Kernel mode Interface to talk to IP stack
- Node discovery using IP Unicast through configuration file
- Alternate Mechanism: IP Multicast

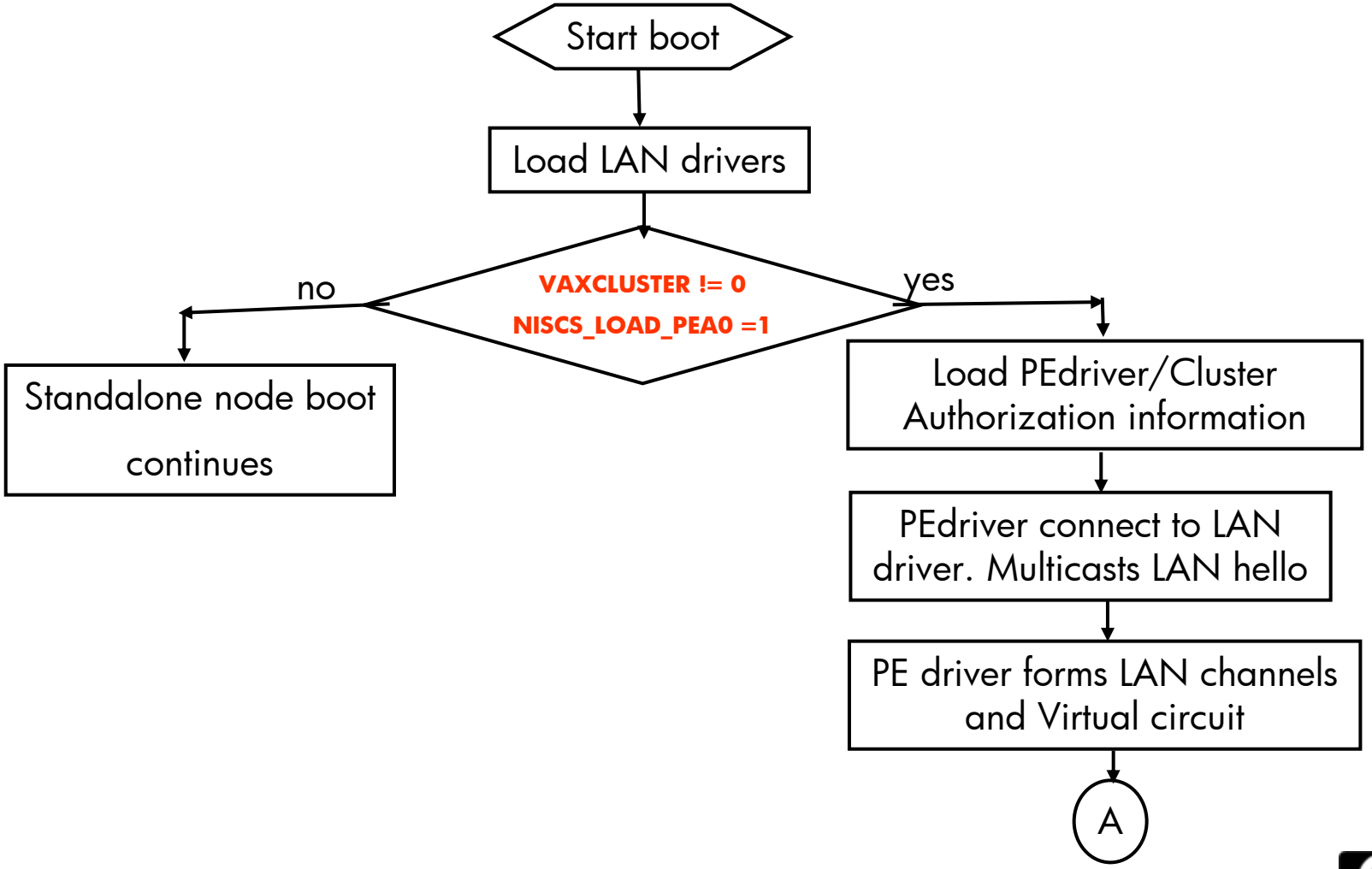
TCP/IP Services boot time loading and Initialization

- Cluster communications are available in an IP only network environment
- Loading of TCP/IP drivers during boot
- Cluster formation in a IP only network
- Existing boot sequence – LAN,PE driver, TCP/IP
- Boot Sequence with IPCI – LAN,TCP/IP, PE driver
- Ability to make use of boot time configuration information to initialize TCP/IP services

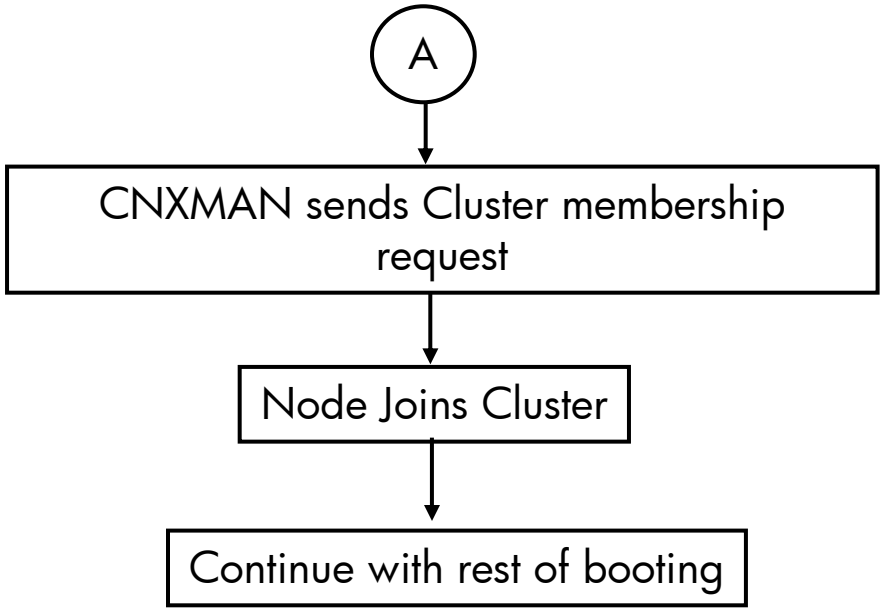
Availability Manager Support

- Support for IPCI in Availability manager
- Monitor cluster with nodes beyond single LAN

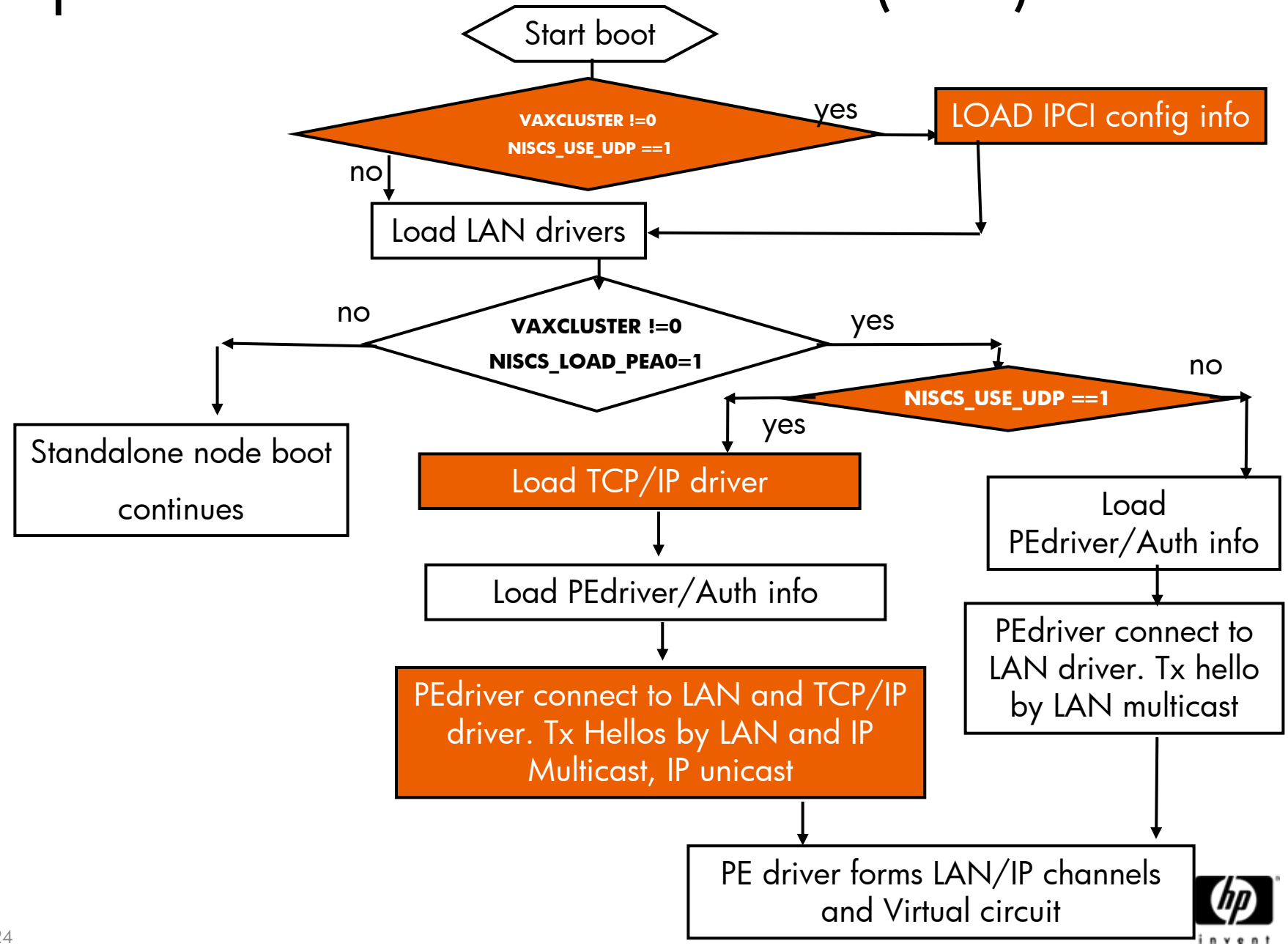
OpenVMS Cluster formation (LAN)



OpenVMS Cluster formation (LAN)



OpenVMS Cluster formation (IPCI)



Node Discovery – Unicast and Multicast

- Node discovery and hello packets
  - IP unicast address (IPv4 address of node)
  - IP multicast technique (Administratively scoped address)
- Unicast can be used if IP multicast is not enabled in network
- Efficient and reliable multicasting
- Basic Cluster principle: All nodes must see all

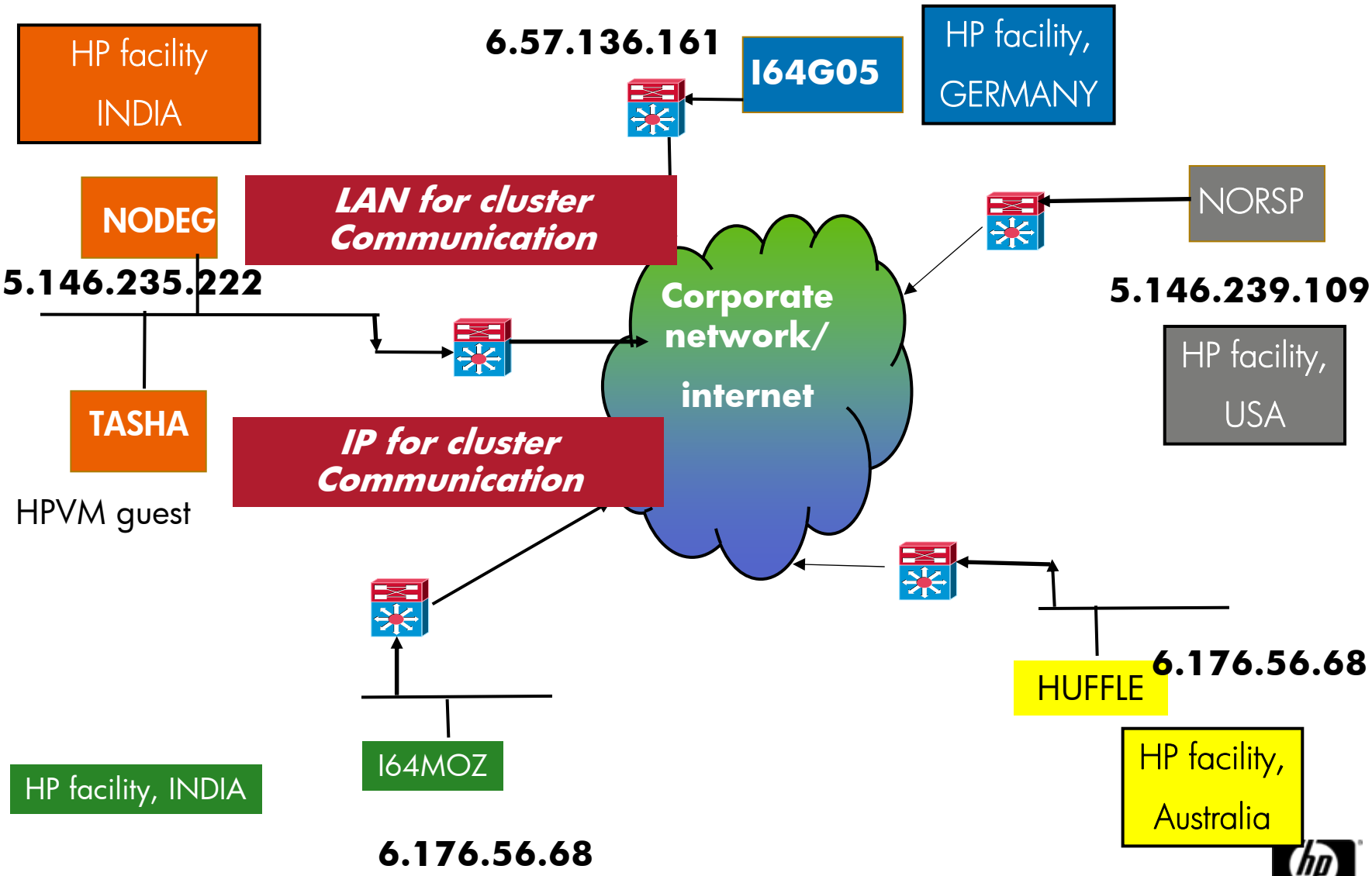
Cluster using IPCI

- 3 nodes in HP, India
- 1 node in HP, USA
- 1 node in HP Germany , 1 node in HP Australia
- 1 HPVM guest node
- Distance between Bangalore facilities <50 miles

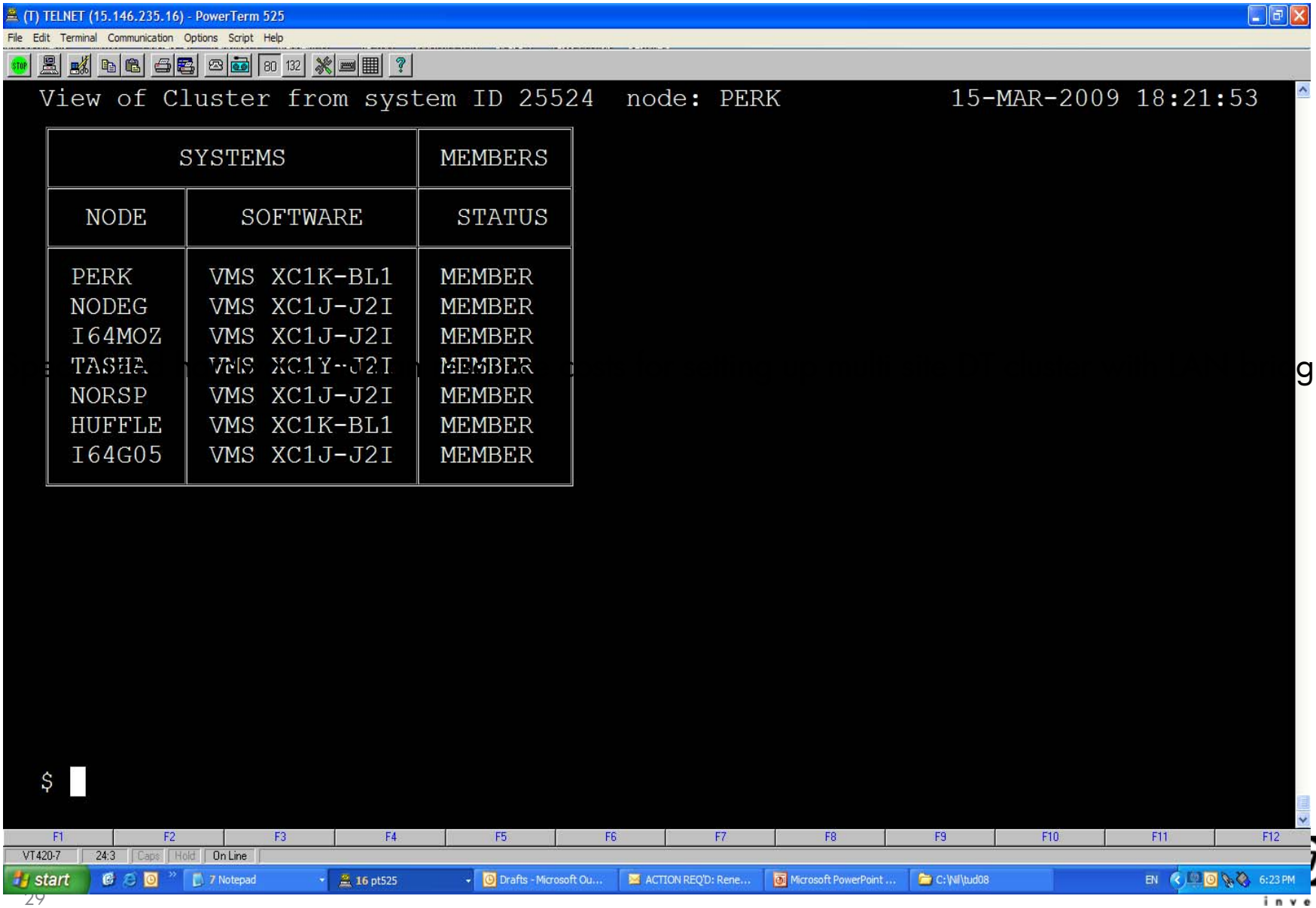
  (PEdriver Latency same as Ping Latency , approx 4ms)
- Distance between India and US - 8000 miles (Latency approx 350ms)

IPCI Cluster

Geographical Cluster using IPCI



SHOW CLUSTER



- ging

Add a node into a IPCI cluster

$ @SYS$MANAGER:CLUSTER_CONFIG_LAN

Cluster Configuration Procedure
CLUSTER_CONFIG_LAN Version V2.80
Executing on an IA64 System

DECnet Phase IV is installed on this node.
IA64 satellites will use TCP/IP BOOTP and TFTP services for downline loading

TCP/IP is installed and running on this node.

Enter a "?" for help at any prompt. If you are familiar with
the execution of this procedure, you may want to mute extra notes
and explanations by invoking it with "@CLUSTER_CONFIG_LAN BRIEF".

This IA64 node is not currently a cluster member.
MAIN Menu

1. ADD I64MOZ to existing cluster, or form a new cluster.
2. MAKE a directory structure for a new root on a system disk.
3. DELETE a root from a system disk.
4. EXIT from this procedure.

1. ADD I64MOZ to existing cluster, or form a new cluster.
2. MAKE a directory structure for a new root on a system disk.
3. DELETE a root from a system disk.
4. EXIT from this procedure.

Enter choice [4]: 1
Is the node to be a clustered node with a shared SCSI/FIBRE-CHANNEL bus (Y/N)? n

IA64 node, using LAN for cluster communications. PEDRIVER will be loaded.
No other cluster interconnects are supported for IA64 nodes.
Enter this cluster's group number: 1985
Enter this cluster's password:
Re-enter this cluster's password for verification:
ENABLE IP for cluster communications (Y/N)? Y
UDP port number to be used for Cluster Communication over IP[49153]?
Enable IP multicast for cluster communication(Y/N)[Y]?
What is IP the multicast address[239.242.7.193]?
What is the TTL (time to live) value for IP multicast packets [32] ?
Do you want to enter unicast address(es)(Y/N)[Y]?
What is the unicast address[Press [RETURN] to end the list]? 6.118.162.109
What is the unicast address[Press [RETURN] to end the list]? 6.50.12.169
What is the unicast address[Press [RETURN] to end the list]? 6.176.56.68
What is the unicast address[Press [RETURN] to end the list]? 5.146.235.222
What is the unicast address[Press [RETURN] to end the list]? 6.138.182.6
What is the unicast address[Press [RETURN] to end the list]?

What is the unicast address[Press [RETURN] to end the list]?

```
********************************************************************
    Cluster Communications  over IP  has  been  enabled.  Now
    CLUSTER_CONFIG_LAN will run the  SYS$MANAGER:TCPIP$CONFIG
    procedure. Please select the IP interfaces to be used for
    Cluster Communications  over IP  (IPCI). This can be done
    selecting "Core Environment"  option from the main menu
    followed by the "Interfaces"  option.  You may also use
    this opportunity to configure other aspects.
********************************************************************

  TCP/IP Network Configuration Procedure

    This procedure helps you define the parameters required
    to run HP TCP/IP Services for OpenVMS on this system.
```

- %TCPIP-I-IPCI, TCP/IP Configuration is limited to IPCI.
- -TCPIP-I-IPCI, Rerun TCPIP$CONFIG after joining the cluster.

Press Return to continue …

- HP TCP/IP Services for OpenVMS Interface & Address Configuration Menu

Hostname Details: Configured=Not Configured, Active=Not Configured

Configuration options:

0 - Set The Target Node (Current Node: I64MOZ)

1 - IE0 Menu (EIA0: TwistedPair 100mbps)

2 - IE1 Menu (EIB0: TwistedPair 1000mbps)

[E] - Exit menu

Enter configuration option: 1

* IPCI Address Configuration *

Only IPCI addresses can be configured in the current environment.
After configuring your IPCI address(es) it will be necessary to
run TCPIP$CONFIG once your node has joined the cluster.

IPv4 Address may be entered with CIDR bits suffix.
E.g. For a 16-bit netmask enter 10.0.1.1/16

Enter IPv4 Address []: 6.138.182.6/24

Requested configuration:

Node : I64MOZ
Interface: IE0
IPCI : Yes
Address : 6.138.182.6/24
Netmask : 255.255.255.0 (CIDR bits: 24)

Is this correct [YES]:

Updated Interface in IPCI configuration file:
SYS$SYSROOT:[SYSEXE]TCPIP$CLUSTER.DAT;

HP TCP/IP Services for OpenVMS Interface & Address Configuration Menu

Hostname Details: Configured=Not Configured, Active=Not Configured
Configuration options:
0 - Set The Target Node (Current Node: I64MOZ)
1 - IE0 Menu (EIA0: TwistedPair 100mbps)
2 - 6.138.182.6/24 *noname* IPCI
3 - IE1 Menu (EIB0: TwistedPair 1000mbps)
[E] - Exit menu

Enter configuration option: E
Enter your Default Gateway address []: 6.138.182.1
The default gateway will be: 6.138.182.1. Correct [NO]: YES
Updated Default Route in IPCI configuration file: SYS$SYSROOT:[SYSEXE]TCPIP$CLUS
TER.DAT;
TCPIP-I-IPCIDONE, Finished configuring IPCI address information.
Will I64MOZ be a boot server [Y]

Configuration Files

- SYS$SYSTEM:PE$IP_CONFIG.DAT
- SYS$SYSTEM:TCPIP$CLUSTER.DAT

SYS$SYSTEM:PE$IP_CONFIG.DAT

- Generated by CLUSTER_CONFIG_LAN.COM
- Read early in the boot sequence
- Provides information to PEdriver
- Can be common through out cluster
- Remote node IP address should be present in local node PE$IP_CONFIG.DAT in order to allow remote node join the cluster
- Best practice for IP unicast: Include all IP address and have one copy throughout the cluster
- "$MC SCACP reload" to be used to refresh IP unicast list on a live system

SYS$SYSTEM:PE$IP_CONFIG.DAT

Configuration File for IPCI

```
! CLUSTER_CONFIG_LAN creating for CHANGE operation on  8-NOV-2008 10:46:19.26
multicast_address=239.242.7.193
ttl=32
udp_port=49153
unicast=6.118.162.109
unicast=6.50.12.169
unicast=6.176.56.68
unicast=5.146.235.222
unicast=6.138.182.6
```

SYS$SYSTEM:TCPIP$CLUSTER.DAT

- Generated by TCPIP$CONFIG which is invoked by CLUSTER_CONFIG_LAN.COM
- Read early in the boot sequence
- Provides information to PEdriver to use the correct TCP/IP interface (WE0 OR WE1) for Cluster traffic
- Provides information to TCP/IP stack to initialize the interface with IP address and default route

SYS$SYSTEM:TCPIP$CLUSTER.DAT

Configuration File for IPCI

- default_route=6.138.182.1
- interface=IE0,EIA0,6.138.182.6,255.255.255.0

Console Messages

HP OpenVMS Industry Standard 64 Operating System, Version XBXH-J2I

%DECnet-I-LOADED, network base image loaded, version = 05.16.00

%VMScluster-I-LOADIPCICFG, loading the IP cluster configuration files

%VMScluster-S-LOADEDIPCICFG, Successfully loaded IP cluster configuration files

%SMP-I-CPUTRN, CPU #1 has joined the active set.

%SYSINIT-I- waiting to form or join an OpenVMS Cluster

%VMScluster-I-LOADSECDB, loading the cluster security database

%EIA0, Auto-negotiation mode assumed set by console

Console Messages

%EWE0, Link up: 1000 mbit, full duplex, flow control (txrx)
%EWD0, Link up: 1000 mbit, full duplex, flow control (txrx)
%PEA0, Configuration data for IP clusters found
%PEA0, IP Multicast enabled for cluster communication, Multicast address, 239.242.7.193
%PEA0, Cluster communication enabled on IP interface, IE0
%PEA0, Successfully initialized with TCP/IP services
%PEA0, Remote node Address, 6.138.185.6,!INDIA added to unicast list of IP bus, IE0
%PEA0, Remote node Address, 5.146.235.222, !INDIA added to unicast list of IP bus, IE0
%PEA0, Remote node Address, 5.146.239.109,!USA added to unicast list of IP bus, IE0
%PEA0, Remote node Address, 5.146.235.224,INDIA added to unicast list of IP bus, IE0
%PEA0, Remote node Address, 6.176.56.68,AUSTRALIA added to unicast list of IP bus, IE0
%PEA0, Remote node Address, 6.50.12.169 !GERMANY, added to unicast list of IP bus, IE0
%PEA0, Remote node Address, 5.146.238.251 ! HPVMGUEST, added to unicast list of IP bus, IE0
%PEA0, Hello sent on IP bus IE0
%PEA0, Cluster communication successfully initialized on IP interface , IE0
%CNXMAN, Sending VMScluster membership request to system NORSP
%CNXMAN, Now a VMScluster member -- system I64MOZ
%STDRV-I-STARTUP, OpenVMS startup begun at 29-OCT-2008 15:20:41.10

SCACP Enhancements

- SCACP – Command line Management Utility for OpenVMS Cluster
- New commands to manage and view Cluster communications using IP
- $ “MC SCACP reload” command to refresh IP unicast list
- $ “ MC SCACP SHOW channels/IP” for IP channels summary

SCACP commands

$ MC SCACP SHOW CHANNEL PERK

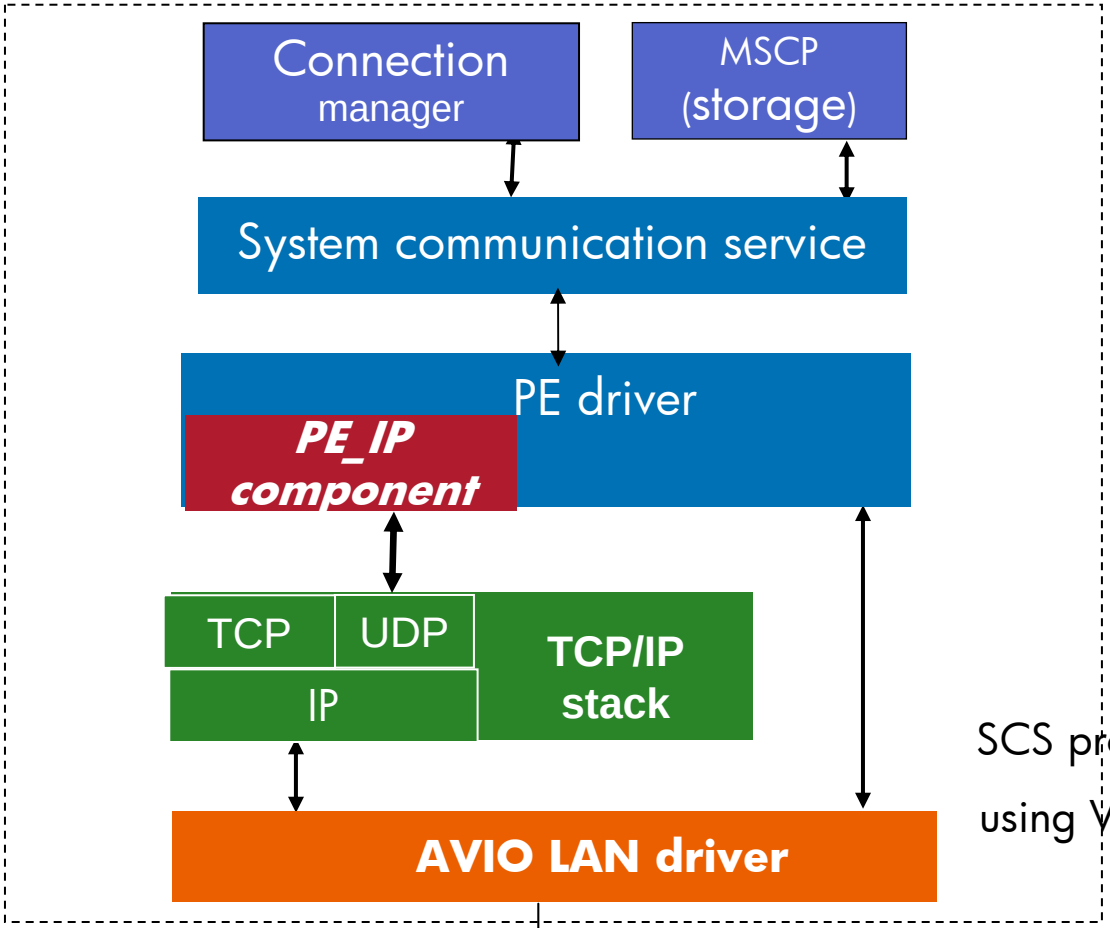
```
NODEG PEA0 Channel Summary 10-MAY-2008 05:09:51.38:

Remote LAN/IP Dev Channel ECS Buffer Delay  Packets
Node   Loc Rmt    State  state      Size  uSec  (S+R)
-----------------------------------------------------
PERK    WE0 IE0  Open  N(T,I,F) 1394   708.9 30410576
PERK    EWA EIA  Open  Y(T,P,F) 1426   551.9 26586732
PERK    WE0 IE1  Open  N(T,I,F) 1394   784.2 38475399
PERK    EWA EIB  Open  Y(T,P,F) 1426   572.4 23780101
PERK    EIA EIB  Open  Y(T,P,F) 1426   694.1 15288091
```

IP channels between nodes

LAN channels between nodes

OpenVMS Cluster Architecture in HPVM



OpenVMS guest cluster node in HPVM

SCS protocol transported using Virtual switch with AVIO

Distance and Latency

- Current Distance limitation will still be applicable
- Speed of Light causes approx 1 millisecond for 50 mile roundtrip
- Distance more than 500 miles require site specific configuration and we suggest contact HP DTCS (HP Disaster Tolerant Cluster services) or Product management

Security

- Normal intranet and Internet Security principles
- VPN (virtual Private Network)
- TTL (Time to live)
- Firewalls

Performance

- Engineering has conducted initial performance test
- TCP/IP stack overhead in μs
  - As distances increase this overhead becomes negligible
  - Better performance for clusters beyond a single LAN
- Recommended FASTPATH CPU configuration
  - LAN,TCP/IP and PE on CPU
  - Ensure headroom in CPU and no saturation

- Introduction to OpenVMS Clustering Technology
- Disaster Tolerant Clusters with OpenVMS
- Cluster Communication Architecture
- Need for IPCI solution
- IPCI Solution details
- **Salient features of IPCI**
- Customer advantage

Feature details

- Available with OpenVMS V8.4
  - Will be available with next Field Test
  - No Prior version support
  - Support for HPVM guest nodes
- Requires HP TCP/IP services for OpenVMS V5.7
  - Not available with other TCP/IP stacks
  - Initial release supports IPv4 only; no IPv6
  - Requires static IP addresses; optionally IP Multicast

Feature details

- Coexists with LAN interconnect for Cluster communication
- Support for Satellite nodes included
- Existing intra-node distance/latency limitation (500 miles) applies

- Introduction to OpenVMS Clustering Technology
- Disaster Tolerant Clusters with OpenVMS
- Cluster Communication Architecture
- Need for IPCI solution
- IPCI Solution details
- Salient features of IPCI
- **Customer advantage**

Customer Advantage

- Only IP services (No LAN bridging) are provided by some Telco Vendors. Customers can now use OpenVMS clusters with IPCI
- Lower infrastructural and Operational costs
  - No extra license/cost for LAN bridging (Layer 2 service)
- Leverage the benefits from the improvements in IP and LAN interconnect technology
- One network infrastructure in Data center for all purposes

- Contact:
  - vivasvan.shastri@hp.com – Product Manager for Clusters.

Any Questions?

Thank You