HP Dutchworld - OpenVMS sessions - November 2008

Networking with OpenVMS systems

(some of the things you always wanted to find out about networks but never got around to trying them out to see what really happens)

Colin Butcher

Context

Part 1:
- Basic principles of data networks

Part 2:
- OpenVMS network protocols (SCS, TCPIP, DECnet, "DECnet over IP", LAT, MOP, AMDS etc.)

Part 3:
- Network infrastructures - putting it all together
- Examples
- Discussion

Part 1:

- Basic principles of data networks and storage networks

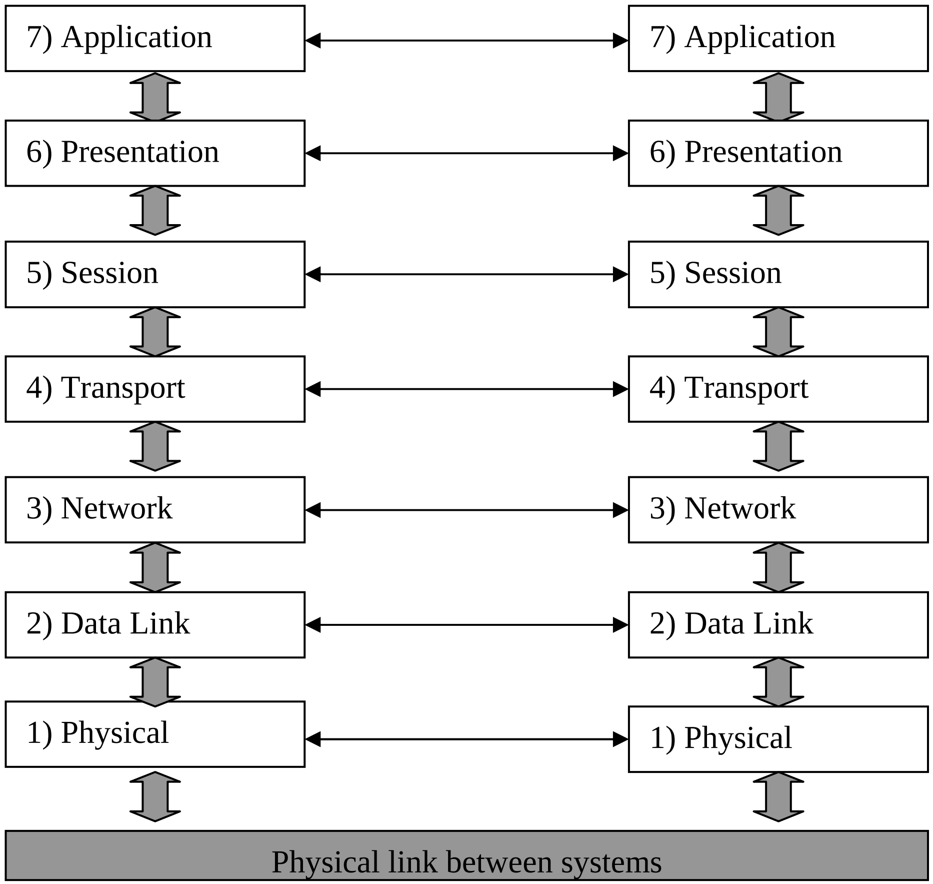
Data networks

- Ethernet technologies
- Physical components and cabling
- Protocols and addressing
- Network Interfaces
- Network Segmentation
- Wide-Area networks (WANs)

OSI 7 layer model

7	Application	Provides for distributed processing and access, contains application programs and supporting protocols (eg FTAM)
6	Presentation	Coordinates conversion of data and data formats to meet the needs of the individual applications
5	Session	Organises and structures the interactions between pairs of communicating applications
4	Transport	Provides reliable transparent transfer of data between end systems with error recover and flow control
3	Network	Permits communication between network entities
2	Data link	Specifies the technique for moving data along network links between defined points on the network, and how to detect and correct errors in the Physical layer (layer 1)
1	Physical	Connects systems to the physical communications media

Data flow between the layers



Cabling

- Transmission properties are important - a square wave at in one end needs to be recognisable as a square wave coming out at the other end
- Copper:
  - Co-axial (thick-wire, thin-wire)
  - Twisted pair (Category 5, 5E, Category 6 etc.)
- Fibre-optic:
  - Monomode (typically 9 micron)
  - Multimode (typically 50 or 62.5 micron)

LAN - Ethernet and derivatives

- 10 Mbit/sec
- 100 Mbit/sec (Fast ethernet)
- 1,000 Mbit/sec (Gigabit ethernet)
- 10,000 Mbit/sec (10Gigabit ethernet)
- Copper / fibre (different transmission characteristics)

- Wireless ethernet (2Mbit / 11Mbit / 54Mbit / 108Mbit)
  - Note: WAP, GPRS, HSCSD, Bluetooth etc. are not wireless ethernet.
  - Access control and data privacy are major issues

NICs (Network Interface Cards)

- Provide connection between IO subsystem and network
- Copper / fibre / wireless physical interfaces
- On-NIC processing:
  - Packet creation
  - Address filtering
  - Encryption
  - Protocol processing (TCP/IP offload - TOE)

Ethernet addressing and packets

- Hardware MAC address
- Physical MAC address
- Broadcast address
- Multicast addresses
- Point to point addresses
- Ethernet packet format v IEEE802.3 packet format
- Packet size (normal frames and jumbo frames)

LANCP> SHOW DEVICE /CHAR
SDA> SHOW LAN

LAN - Segmentation

Why segment a network?

- Performance
- Security
- Availability

How can you segment a network?

- Multiple NICS
- Repeaters
- Bridges
- Switches
- VLANs
- Routers

Repeaters

- Layer 1 devices ("flat" network)
- Provide electrical fault isolation
- Simply re-time and re-transmit signal
- No control of bandwidth
- Beware of cumulative end to end delay exceeding maximum permissible frame timing – which leads to 'folklore' such as the "three repeater rule"

Beware of the generic term "hub"

Bridges

- Packet content based (Layer 2)
- Store and Forward
- Easy to use and configure
- Poor control of bandwidth (filtering)
- Spanning tree algorithm
- Provides an extended LAN
- Not all protocols can tolerate the inherent delays in working over an extended LAN
- Remote booting (MOP, BOOTP etc.) will absorb bandwidth

Switches

- Introduces parallelism
- Speed of chipsets (latency & bandwidth)
- Full duplex operation on a single device per port basis
- Traffic monitoring (mirror ports)
- Link aggregation
- Bandwidth control
- “Store and forward” versus “Cut through” switching
- Layer 2, Layer 3 etc. switching
  - Layer 3 generally refers to TCP/IP routing layer
  - Layer 4 generally refers to TCP/IP port numbers, eg: HTTP port 80 traffic)

VLANs are another way to segment a network for performance and security

- Implemented within core switches
- Also implemented in some NICs / device drivers (LLdriver)
- Port based VLANs
- Protocol based VLANs
- Connectivity between VLANs
- VLAN tagging of packets (802.1Q)
- VLAN tagging of packets out of NICs
- QoS (Quality of Service) and bandwidth reservation

Routing

- Routers do not need to be involved in the normal inter-node traffic within a LAN, other than keeping track of who's where and making themselves known
- Routers build knowledge of address (node or interface) reachability on a per-protocol basis
- Protocol address based (Layer 3)
- Need to design addressing scheme
- Bandwidth control
- Design routing paths
- Routing table updates are propagated between routers

Routers

Routers are generally used to interconnect LANs over one or more WAN links

- Separate devices or can be integrated into the core
- Need to design protocol addressing scheme and areas
- Good control over bandwidth
- Layer 3 devices – protocol address based
- IPV6 is common in big core routers
- Rare to find DECnet routing in modern routers – it's a TCP/IP dominated world in the WAN
- Can set up OpenVMS systems as dedicated multiprotocol routers if you need both DECnet and TCP/IP routing

Part 2:

- OpenVMS network protocols (SCS, TCPIP, DECnet Phase IV and DECnet-Plus, "DECnet over IP", LAT, MOP, AMDS etc.)

Typical network protocols:

- SCS (clustering)
- TCP/IP (and all it's components)
- DECnet-Plus (NSP,OSI and DECnet over IP) or DECnet Phase IV (NSP only)
- DECdns (not to be confused with TCP/IP's DNS/BIND)
- LAT (DECserver terminal access etc.)
- MOP + Remote Console (DECserver, LAVC boot etc.)
- DTSS (can be disabled)
- LAD + LAST (Infoserver etc.)
- AMDS (quorum adjustment)

OpenVMS networking features (1)

- VAX VMS V4.x introduced SCS for LAVC
- Infoserver introduced LAD / LAST for serving remote disc containers. Also used by RSM. Available in OpenVMS V8.2-1 onwards for Integrity to provide network upgrade.
- Pathworks (Advanced Server) introduced DECnet for PC operating systems and LANmanager functionality for OpenVMS systems. Replaced by CIFS (based on SAMBA)
- Galaxy introduced SMCI pseudo-LAN interconnect
- V8.3 introduced PEdriver compression

OpenVMS networking features (2)

- OpenVMS V7.1 introduced LANCP / LANACP for MOP loading without DECnet (needed to load cluster satellites)
- OpenVMS V7.3-2 introduced "LAN failover" for improved LAN availability (all protocols)
- TCP/IP V5.4 introduced "failsafe IP" for improved TCP/IP availability within a cluster

- LLdriver: LAN failover, VLAN tagging
- Jumbo frames (frame size varies with device type)
- PEdriver: compression, checksumming

LAN failover

- Presents single high-availability LAN device for protocols that don't handle multi-path LANs well (TCP/IP V4, AMDS, etc.)
- Equivalent to Proliant NIC teaming in failover mode
- No load balancing
- Inter-operability with switches
- Use LANCP to set up the failover group of NICs
- Booting is interesting – the cluster protocol has to restart on the virtual interface after the cluster has first formed before LLdriver was loaded

TCP/IP V4 layers

- Layer 4 – port or socket layer (eg: HTTP = port 80, "well known" ports allocated by convention)
- Layer 3 – IP addressing and routing layer (eg: 192.168.0.n/24, DNS/BIND resolver user to convert IP hostnames to interface IP addresses)
- Layer 2 – MAC address layer (ARP used to convert IP interface addresses to MAC addresses, cached locally)
- Layer 1 – Physical layer (transmission media)

TCP/IP services

- DNS and the BIND resolver
- DHCP address provision
- BOOTP services
- FTP / TFP file transfer
- NFS file serving
- Monitoring with SNMP
- SMTP / POP / IMAP
- Secure extensions: SSH, SSL, IPSEC
- Printing (LPR / LPD)

SYS$MANAGER:TCPIP$CONFIG.COM
Also BSD style commands as well as DCL commands

Failsafe IP

- Think about "service addresses", not "interface addresses"
- Have per interface addresses for manageability and path availability "ping tests"
- Have per machine addresses for systems management access (one failsafe IP address across all NICs running IP in that machine)
- Have per application addresses for user access and manual load distribution across NICs.

Use the IFCONFIG commands

DECnet Phase IV - addressing

- End Node
- Routing Nodes: Level 1 & Level 2 (Area) Routers
- MAC Address formed from Node address:
  - Area 1 - 63, Node: 1 - 1023
  - 16 bit address = (Area x 1024) + Node number
  - SCSSYSTEMID = same 16 bit value
  - AA-00-04-00-nn-mm
  - nn-mm = byte reversed hexadecimal 16 bit address

DECnet Phase IV – who's where?

- DECnet "hidden information":
  - End Node to Routers (end node hello packets)
  - Routers to Routers (routing updates)
  - Routers to End Nodes (router hello packets)

- DECnet Phase IV bases the MAC address on the node number, so no need for routers on LAN except for determining adjacencies.

Router on LAN will give fast "node unreachable" rather than slow "timeout" when attempting to connect to a node that is not on the LAN.

DECnet Phase IV - Limitations

- Number of nodes in a private network can exceed the address range (eg: Easynet)
- MOP loader needs fake node entries
- Sets MAC address on all LAN adapters based on DECnet node address, so cannot connect multiple LAN adapters to the same LAN (or extended LAN).

Can route between parallel LANs, but cannot bridge between them due to the risk of duplicate MAC addresses.

DECnet Phase IV to Phase V

- The obvious big difference - NCL in place of NCP
- Name Services
- DECnet over IP
- Permanent database is NCL script files (text)
- Time Synchronisation Service
- Routing algorithms (Phase V routers)
- Multiple path behaviour (multi-homed End System)
- Startup early in boot sequence
- Phase IV compatible addressing on first adapter only (by default)

Addressing

- Phase IV compatible addressing
  - The AA-00-04-00-xx-xx address
  - Multiple CSMA-CD adapters

- Synonyms and FullNames
- Address Towers
  - Transport selection (NSP or TP4)
  - Session Control version selection (SC2 or SC3)

NCL Entities

- session control
  - applications and ports
- transports
  - NSP and OSI (plus OSI templates)
- routing
- routing circuits
- csma-cd station (ethernet and FDDI)
- <datatype> links (HDLC, DDCMP etc.)
  - <datatype> link logical station
- modem connect lines

- Can disable DTSS by defining the NET$DISABLE_DTSS logical name in SYLOGICALS.COM.
- DTSS server can receive time from NTP (example provided)
- New procedures for changing DST zone rules (Alpha only), also see AUTO_DLIGHT_SAV system parameter
- Phase IV migration improvements (databases, FDDI)
- Improved NCL help
- Reduced NCL output on boot by default (NET$STARTUP_QUIET_NCL logical)

Relative time is more useful than absolute time

- Need to be able to order events across the network based on timestamps
- Timestamp format
  - Time value
  - Inaccuracy component
- DTSS Servers and Clerks
- External reference Time Providers

DECnet over IP (1 of 2)

- Preserves DECnet APIs for existing applications
- Performance and availability are determined by underlying IP network infrastructure
- DECnet uses TCPIP as a pseudo-transport layer
- Need to have RFC1006 and RFC1869 (aka RFC1006-Plus) OSI transport templates - ports 102 and 399
- Streams interface
- Need to have PWIP driver enabled

DECnet over IP (2 of 2)

- Need to have DNS/BIND name service in list for access to local name resolver:
  - @NET$CONFIGURE ADVANCED
  - Naming services: "LOCAL,DOMAIN"
  - use 127.0.0.1 as address of name resolver!
- Can enable DECnet over IP "on the fly":
  - change the naming (remove from LOCAL or DECdns naming database with DECNET_REGISTER and add to HOSTS database or DNS/BIND server)
  - NCL> FLUSH SESSION CONTROL NAMING CACHE ENTRY "*"

Phase V summary

- Don't be scared of Phase V - it works, generally at least as well as or better than Phase IV
- NCL is a much better tool than NCP
- Keep it simple - use Local naming, Phase IV compatible addressing and the NSP transport
- DECnet over IP lets you use applications with the DECnet APIs over a TCP/IP only infrastructure
- Multi-Homed End Systems get you load balancing for the price of an end-node licence

Part 3:

- Network infrastructures - putting it all together
- Discussion

Network link characteristics

- Bandwidth – determines throughput
  - Large packets shift more data with less overhead
- Signal path quality and reliability
  - Retransmits severely affect overall throughput
- Latency – determines round trip delay
  - Determines how much data is in transit at any given instant
  - Data in transit is at risk if there is a failure
- Jitter (“div latency” or variation of latency with time) – determines predictability of round trip delay
  - Understanding jitter is important for establishing timeout values
  - Severe latency fluctuations can cause system failures

Traffic flow and performance

- Traffic flow, end-to-end packet delivery, delivery failure notification and performance are key parts of the design of any network protocol, as are the addressing scheme and the naming scheme

- Multicast packets are inherently “fire and forget”
- Multiple paths – packets may no longer arrive in the order in which they were sent
- What happens when paths fail or are intermittent?
- How do we cope with bad latency or jitter?

- Time synchronisation across the infrastructure

- Node naming, addressing schemes and routing mechanisms

- Multiple NICs and multiple LANs

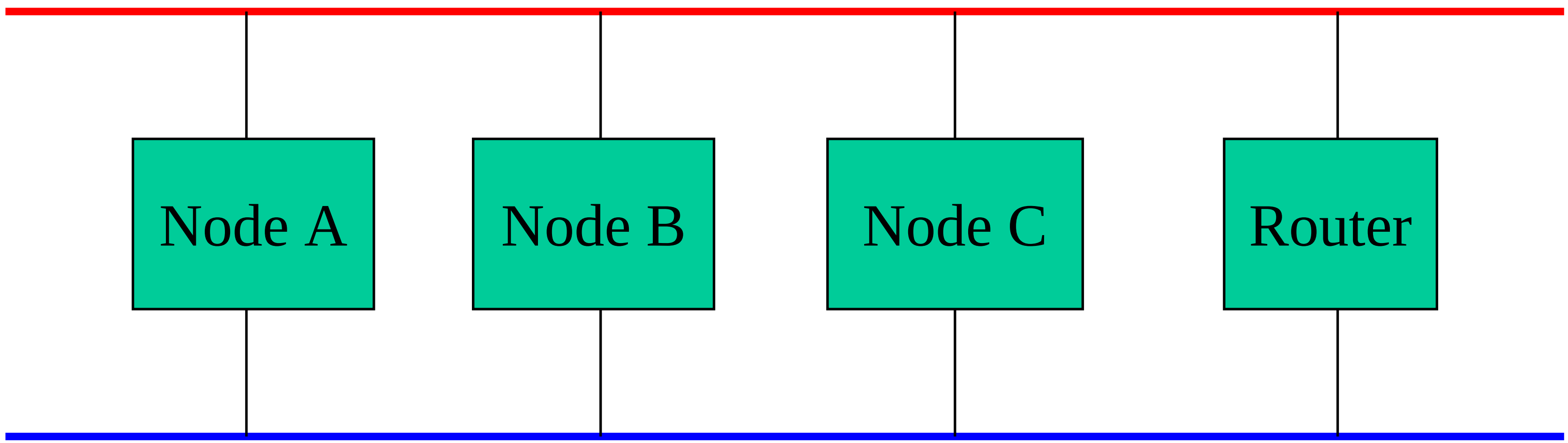
- Map functions to NICs:
  - Management (ILO, SAN appliance, etc.)
  - Clustering
  - Network backups
  - Data transfers (eg: FTP, NFS etc.)
  - Interactive users

Multiple LAN interface behaviours

- LAN failover (LLdriver)
- DECnet Phase IV and V - load balancing
- TCP/IP - Failsafe IP
- SCS – stopping and starting per adapter with SCACP or LAVC$START_BUS / LAVC$STOP_BUS

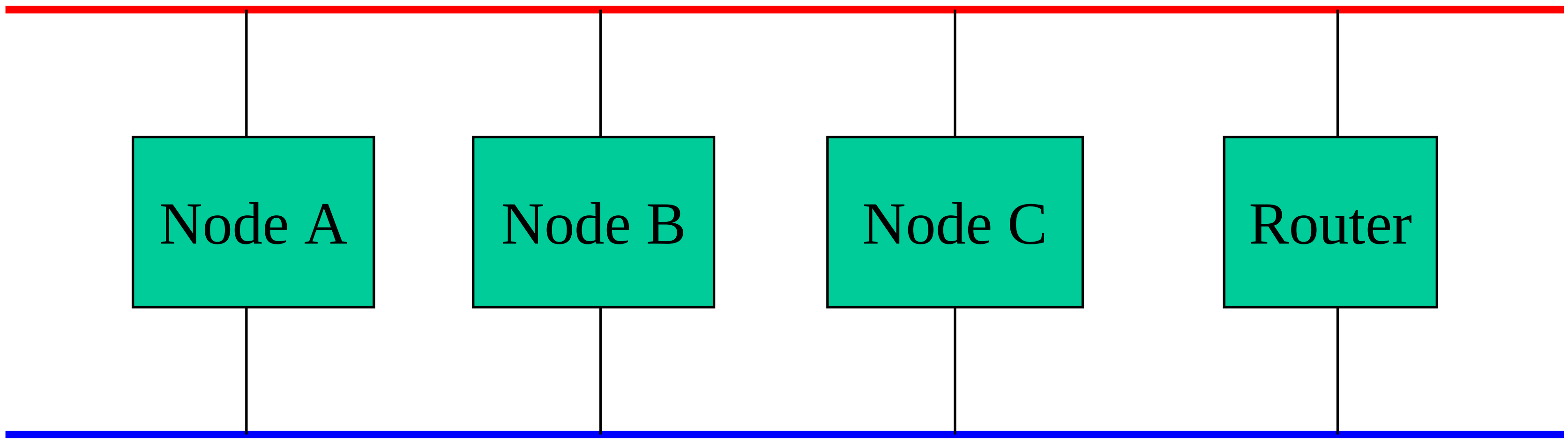
- MOP and LANCP (network booting)
- LAD / LAST (InfoServer)
- LAT (DECservers)

Dual LAN behaviour - DECnet



- Phase IV – different costs, all L1 routers
- Phase IV Plus – same costs, all L1 routers
- Phase IV Plus – same costs, end node failover
- DECnet/OSI – both active, Multi-homed End Systems
- DECnet-Plus – both active, Multi-homed ES or IS

Dual LAN behaviour – TCP/IP

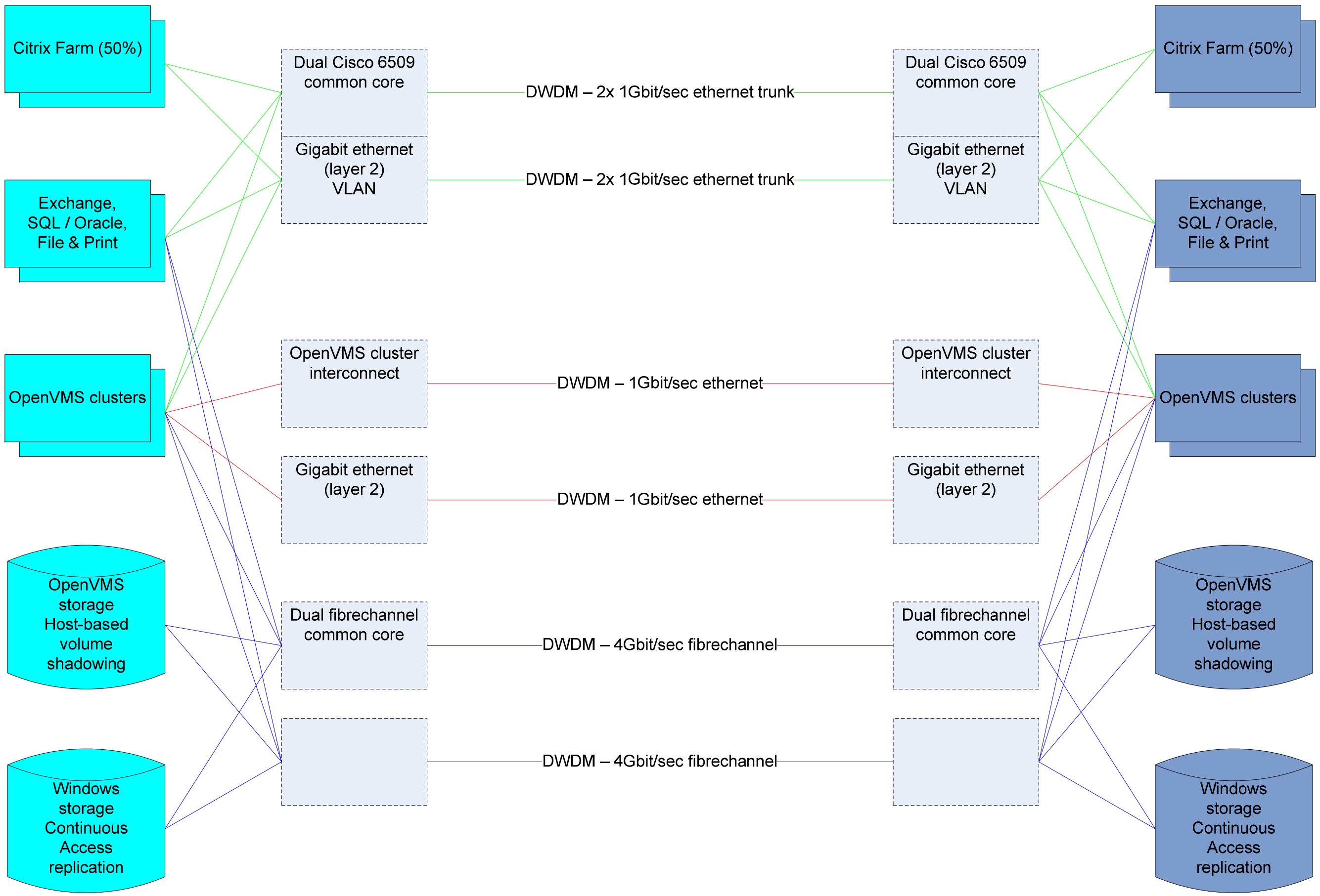


- Multiple NICs per subnet
- Different physical networks must be in different subnets
- Use LAN failover for high-availability dual connect

Examples and discussion

- Safety-critical and mission-critical system:
  - Migrate from Alpha to Integrity
  - Move from 3x regional clusters to single national cluster
  - Move from HSG80s to EVA4100s
  - Move to multiple NIC connectivity
- Similar principles apply in many other cases

Example overall infrastructure

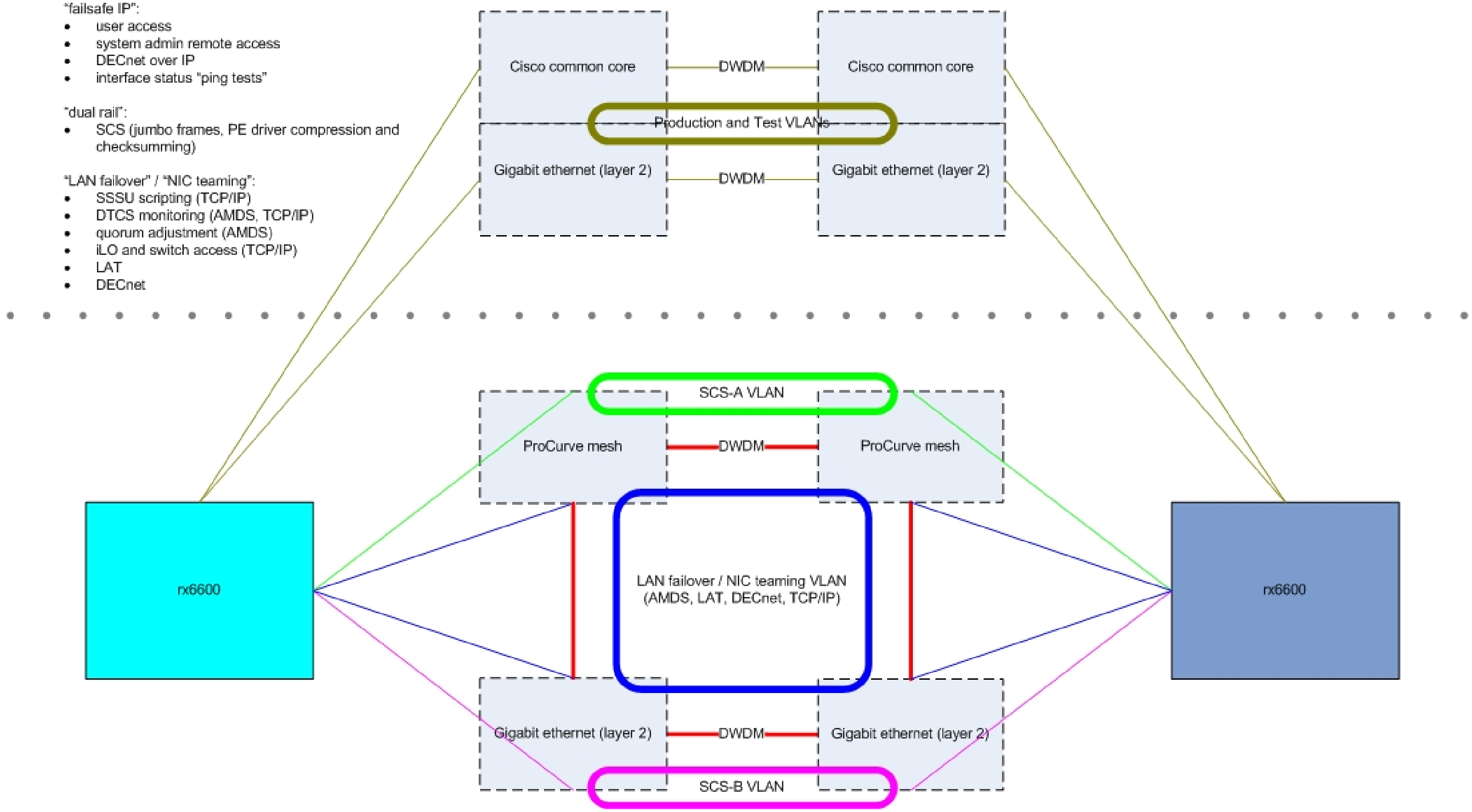


Common network infrastructure

"failsafe IP":
- user access
- system admin remote access
- DECnet over IP
- interface status "ping tests"

"dual rail":
- SCS (jumbo frames, PE driver compression and checksumming)

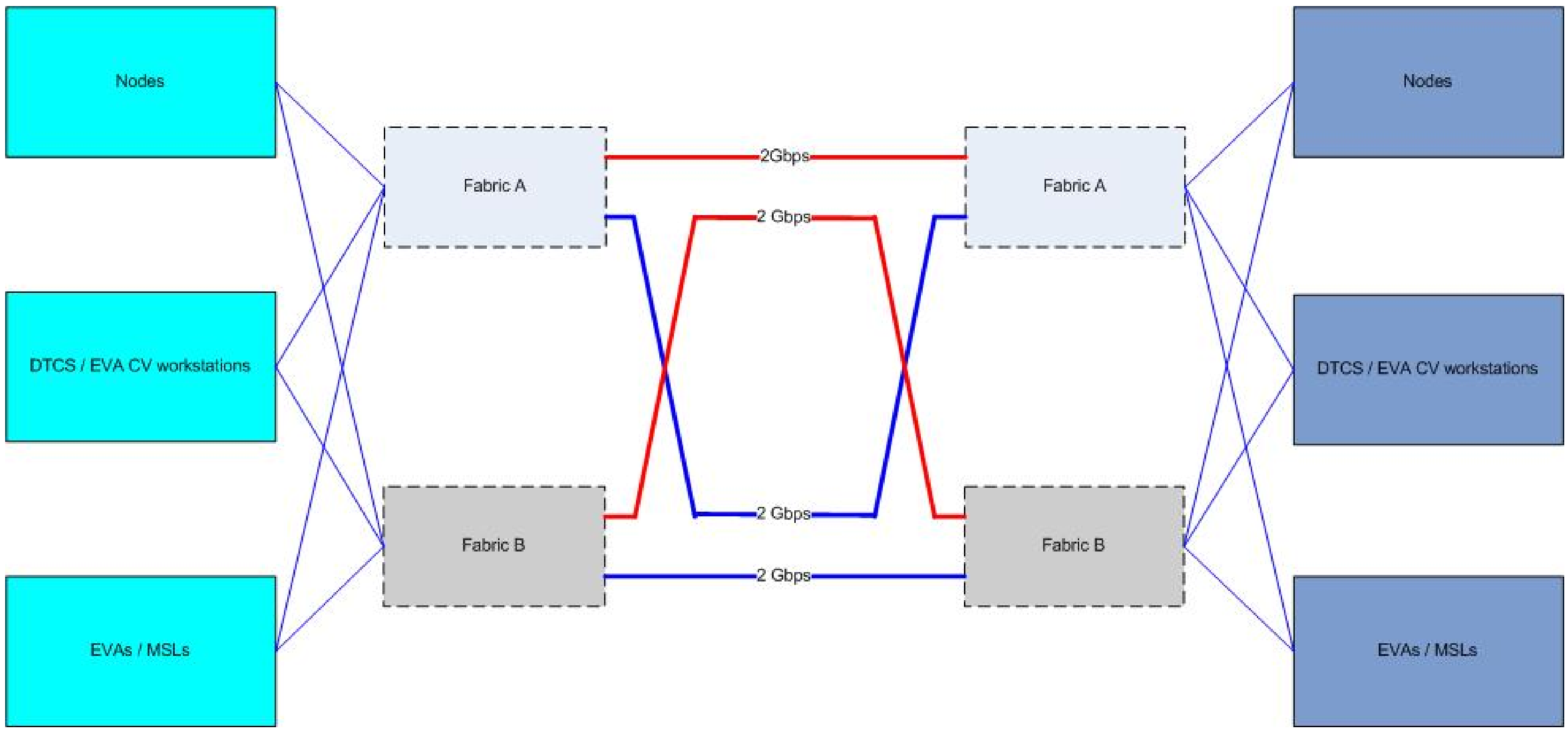
"LAN failover" / "NIC teaming":
- SSSU scripting (TCP/IP)
- DTCS monitoring (AMDS, TCP/IP)
- quorum adjustment (AMDS)
- iLO and switch access (TCP/IP)
- LAT
- DECnet



Web: www.xdelta.co.uk

Stuff that works

SAN fabric infrastructure



Converged ethernet and RNIC

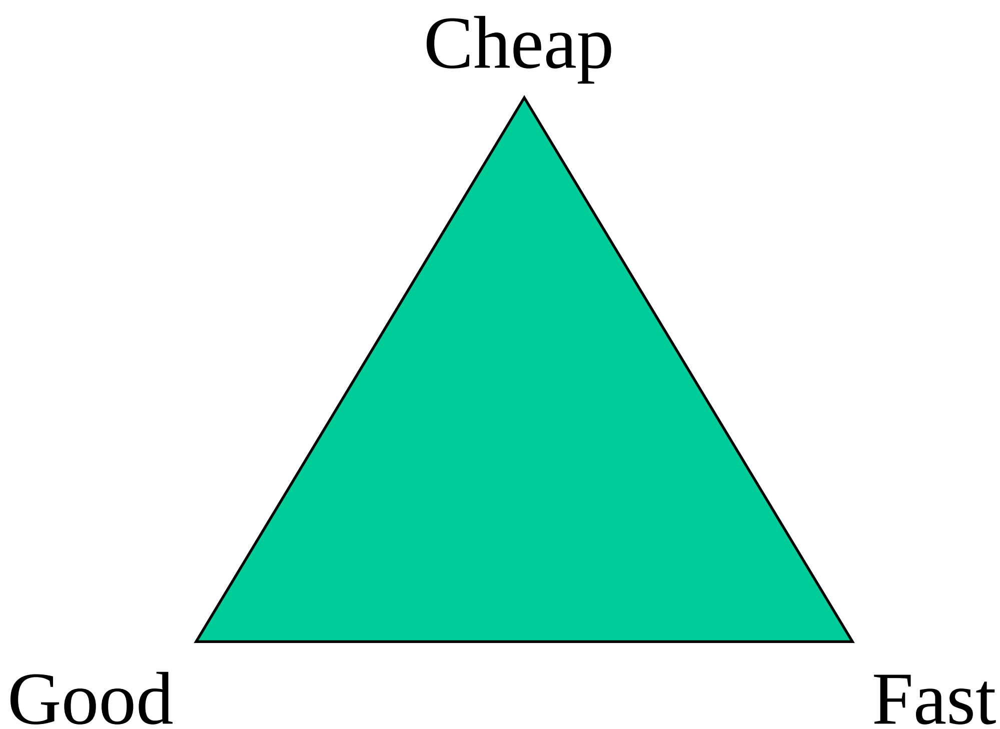
- "Converged ethernet" – fibrechannel and ethernet protocols on the same physical carrier with common interfaces and switching infrastructure

- 10GigE (and faster)

- Fibre, not copper (transmission characteristics matter)

- RNIC – offloading the bulk of the protocol handling to the NIC and minimising both CPU overhead and moving data around between the memory subsystem and the NICs

Compromises – pick any two!



Thank you for your participation

Q & A