Connect webinar – 12th Feb. 2010

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

www.downloads.xdelta.co.uk/2010/2010_02_12-openvms_networking-colin_butcher.pdf

Part 1:

- Basic principles of data networks and storage networks

Part 2:

- OpenVMS "on the wire" protocols (SCS, TCPIP, DECnet, "DECnet over IP", LAT, MOP, AMDS etc.)

Part 3:

- Network infrastructures - putting it all together
- Discussion

Part 1:

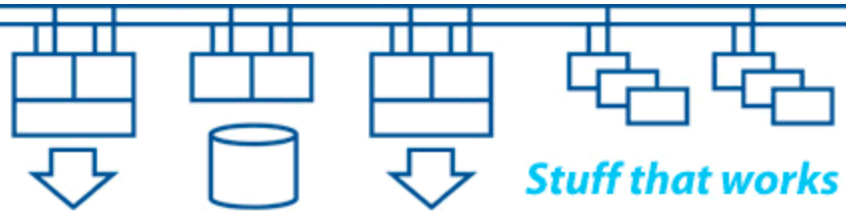
- Basic principles of data networks and storage networks

Data networks

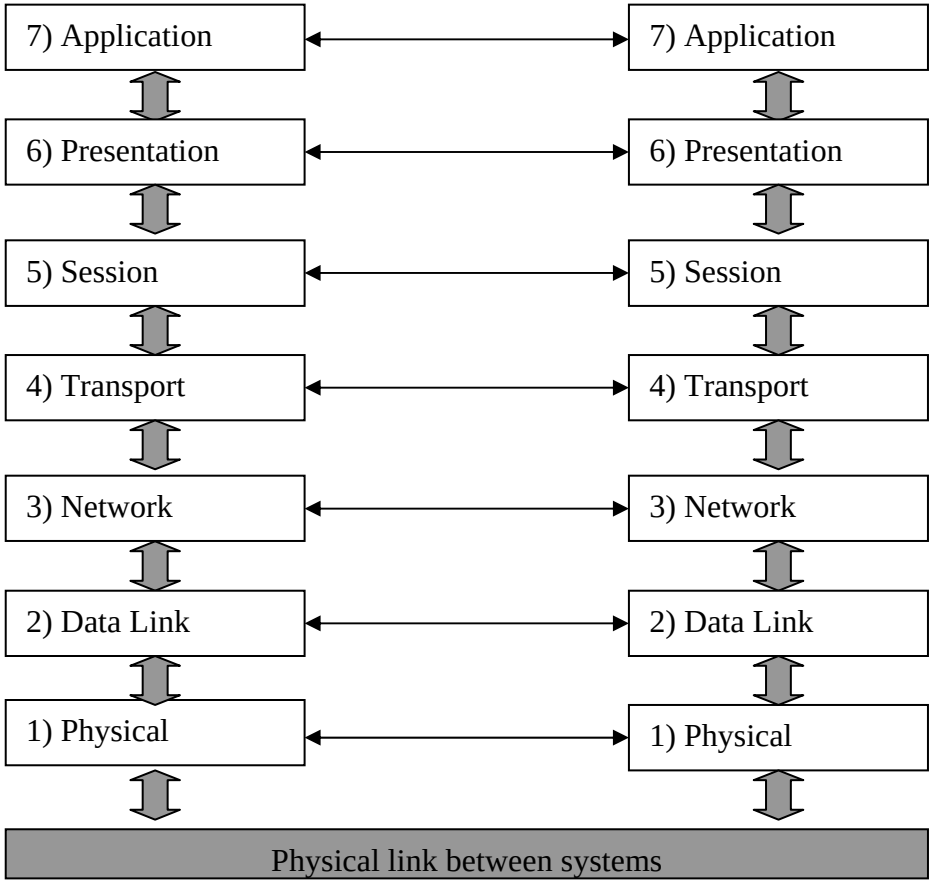
- Local-Area Networks (LANs)
  - Ethernet technologies
  - Physical components and cabling
  - Protocols and addressing
  - Network Interfaces
  - Network Switches
- LAN segmentation
- LAN extension
- Wide-Area networks (WANs)

OSI Seven-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



Four-layer model (predominantly TCP/IP)

- Layer 4 – Layer 3 protocol specific: TCP/IP port or socket (BG devices); DECnet 'object' or 'application'
- Layer 3 – Protocol specific addressing and routing layer. Needs protocol address to MAC address translation. TCP/IP 'routing'; DECnet 'circuit' or 'routing circuit'
- Layer 2 – MAC address layer, Ethernet V2 or IEE802.3 format packets. TCP/IP 'interface'; DECnet 'line' or 'csma-cd station';
- Layer 1 – Physical layer (transmission media)

Cabling

- Transmission properties, transmitter components and receiver components are important - a square wave fed at in one end needs to be recognisable as a square wave coming out at the other end

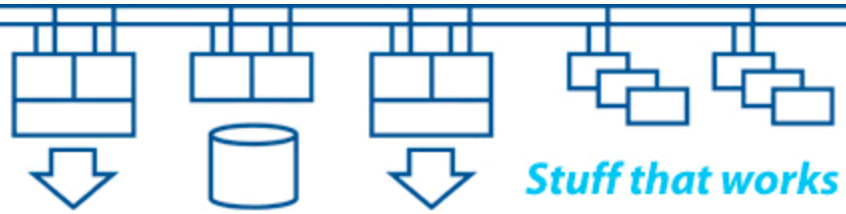
- Copper:
  - Co-axial (thick-wire, thin-wire)
  - Twisted pair (Category 5, 5E, Category 6 etc.)

- 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, HSPA, UMTS, 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

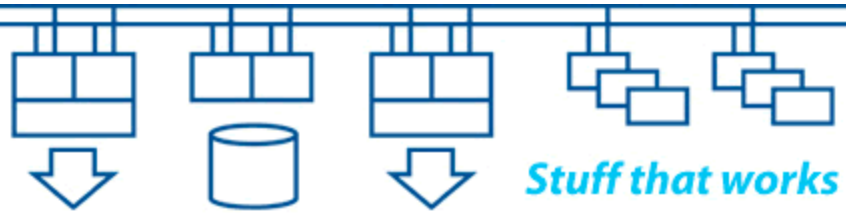
LAN - Segmentation

Why segment a network?

- Availability
- Performance
- Security

How can you segment a network?

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



Repeaters

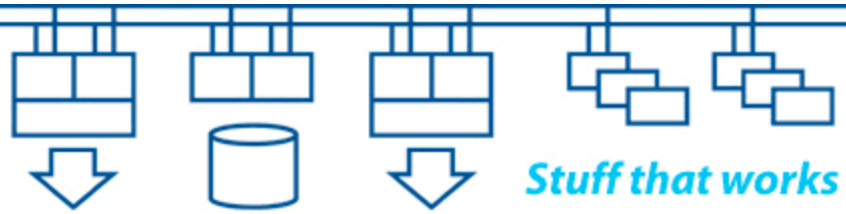
- Layer 1 devices ("flat" network)
- Provide electrical fault isolation
- Simply re-time and re-transmit signal
- No control of bandwidth
- Beware of cumulative end to end delay exceeding maximum permissible frame timing – which leads to 'folklore' such as the "three repeater rule"

- *TIP: Beware of the generic term "hub"*

- 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, Layer 4 switching
  - Layer 2 is protocol independent – MAC address based
  - Layer 3 generally refers to TCP/IP routing layer
  - Layer 4 generally refers to TCP/IP ports, eg: HTTP port 80



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

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

LAN extension between sites

- Different manufacturers (Cisco, HP, Extreme etc.) have slightly different terminology and features (eg: Cisco 'etherchannel'; Procurve 'meshing'; Extreme 'EAPS ring'; etc.)
- Inter-switch links can be "trunked" to provide sufficient bandwidth
- Link "glitches" will cause traffic disruption, so use routing (layer 3, not layer 2) to minimise disruption
- VLANs can extend across multiple switches
- Wave division multiplexing (DWDM, CWDM) can be used for extended distance inter-switch links

Wireless ethernet

- Shared bandwidth ("flat" network)
- Security issues (access control, authentication, data encryption)
- Roaming issues (multiple Access Points and MAC address migration between ports)
- Management issues
- Antennas (coverage and beam patterns)
- Wireless repeaters and bridges

Storage networks

- Fibrechannel technologies (1 / 2 / 4 / 8 … Gbps)
- Storage devices (disc arrays and tape drives)
- Host Bus Adapters (HBAs) and FC switches
- WWIDs and WWNs
- SAN Segmentation (switching, routing, zoning etc.)
- Storage subsystems and device presentation
- SAN extension (FC over IP, DWDM 'dark fibre')

Fibrechannel

- A switch based network optimised for shifting large quantities of data with high throughput and low latency
- All endpoints (HBAs, storage controllers etc.) uniquely identified with a WWID (World-Wide ID)
- Multiple switches can be interconnected
- Inter-switch links can be trunked
- The network between the storage devices and the systems is known as a fabric
- Large fabrics need to be segmented
- High availability typically uses a dual-fabric SAN

Device addressing and presentation

- WWIDs are unique
- Systems and storage controllers scan the fabric to build a list of paths between devices
- Storage devices (eg: EVA Vdisks) are presented to specific hosts (HBAs) by the array controller with a LUN (logical unit number) and (required by OpenVMS) a device identifier
- Device presentation can be controlled to limit access to specific paths (by WWID)

SAN segmentation

- Device paths and visibility can be controlled by zoning in the switches
- Zones can be port based, or WWID based (known as "soft zoning")
- Zones can overlap (think Venn diagrams)
- Current zoning best practice uses the "single initiator, multiple targets" model

- Systems (HBAs) need to have BIOS type support for booting from SAN devices

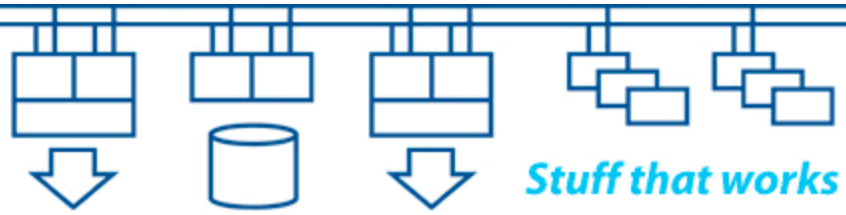
- See the HP SAN design reference guide

SAN extension between sites

- Inter-site links can be "trunked" (as with data networks) to provide sufficient bandwidth
- Link "glitches" will cause fabric resets and rescans, so use FC routing in large extended SANs to minimise disruption
- Zones can extend across multiple switches (as with VLANs)
- Wave division multiplexing (DWDM, CWDM) can be used for extended distance inter-switch links
- "FC over IP" can be used to link SANs over an IP data network (beware latency issues – use QoS techniques)

1 (c): Wide Area Networking (WAN)

- ISDN, POTS
- Leased Line (KiloStream, MegaStream, T1 etc.)
- Frame Relay
- ATM
- MPLS
- "Dark fibre" and Wave Division Multiplexing
- SONET / SDH etc.
- ADSL / SDSL
- VPNs
- Managed services (usually TCP/IP based)
- Encapsulation and tunnelling
- FC over IP, FC over Ethernet

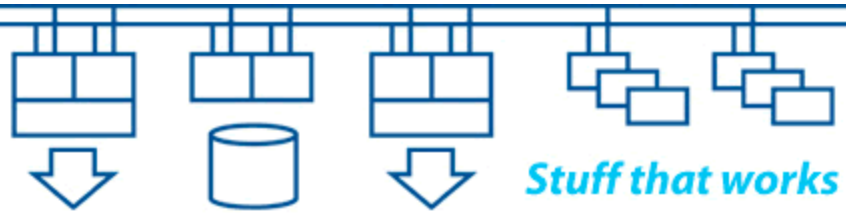


Routing

- 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 are generally used to interconnect LANs over a WAN

- 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:

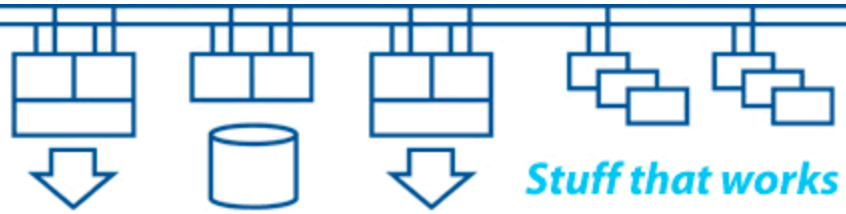
- OpenVMS "on the wire" network protocols (SCS, TCPIP, DECnet Phase IV and DECnet-Plus, "DECnet over IP", LAT, MOP, AMDS etc.)

Typical network protocols "on the wire":

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

OpenVMS networking features (1)

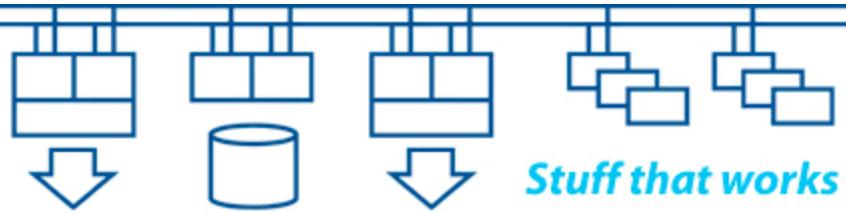
- VAX VMS V4.x introduced SCS for LAVC
- Infoserver introduced LAD / LAST for serving remote disc containers. Also used by RSM. Available in OpenVMS V8.2-1 onwards for Integrity to provide network upgrade.
- Pathworks (Advanced Server) introduced DECnet for PC operating systems and LANmanager functionality 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



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)

TCPIP$CONFIG
BSD style commands as well as DCL commands

DECnet Phase IV - addressing

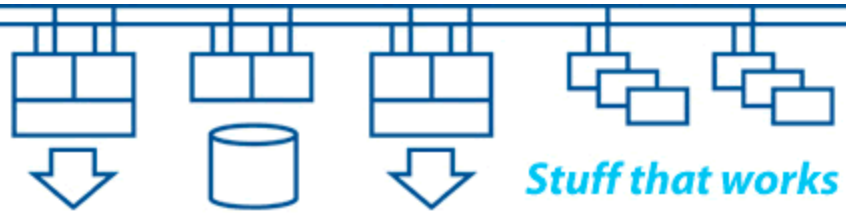
- End Node
- Routing Nodes: Level 1 & Level 2 (Area) Routers
- MAC Address formed from Node address:
  - Area 1 - 63, Node: 1 - 1023
  - 16 bit address = (Area x 1024) + Node number
  - SCSSYSTEMID = same 16 bit value
  - AA-00-04-00-nn-mm
  - nn-mm = byte reversed hexadecimal 16 bit address

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.

TIP: 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).

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

- 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)

- 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

- OpenVMS V7.3-2 onwards
- Can disable DTSS by defining the NET$DISABLE_DTSS logical name in SYLOGICALS.COM.
- DTSS server can receive time from NTP (example provided - see SYS$EXAMPLES:)
- 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)

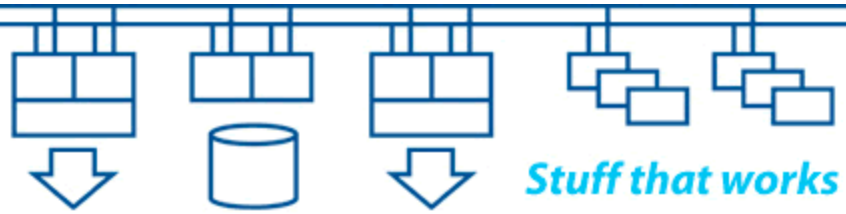
DECnet over IP (1 of 2)

- 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 "*"



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

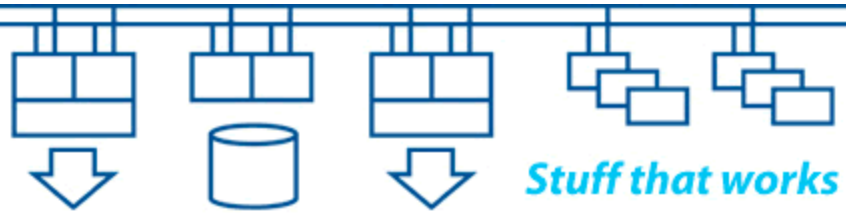
Traffic flow and performance

- 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

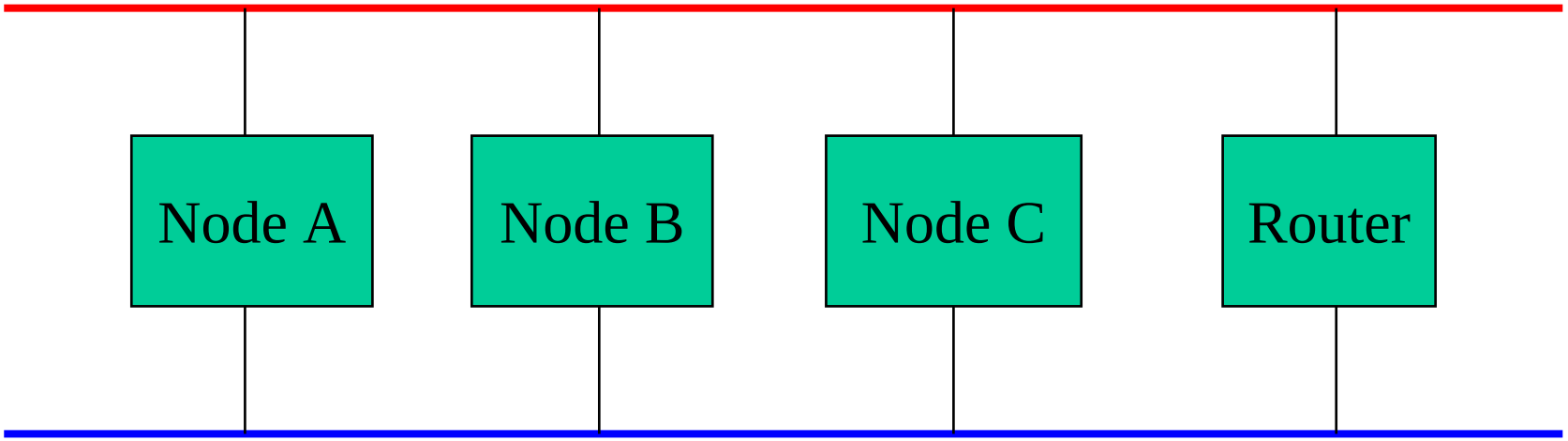


Multiple LAN interface behaviours

- 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



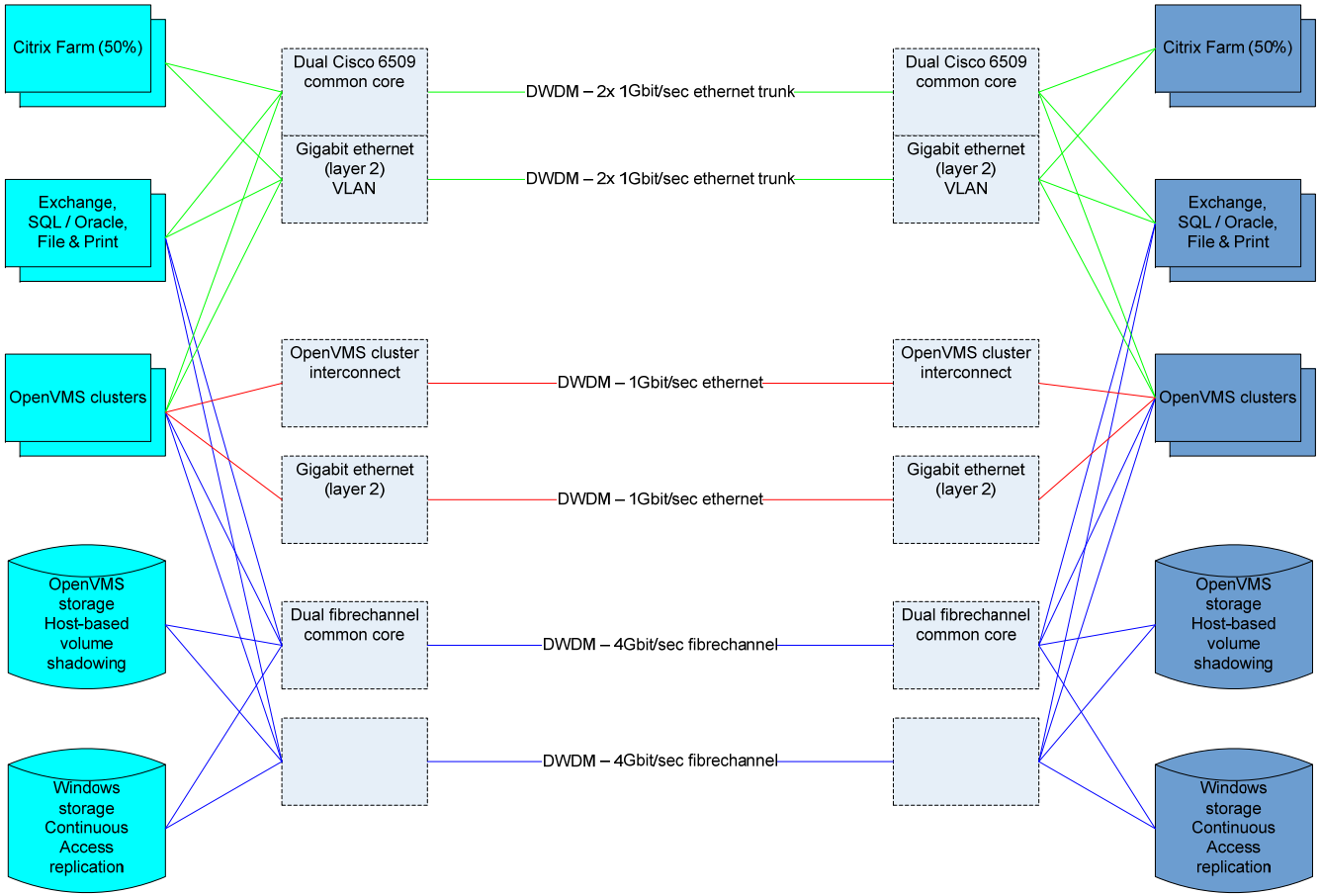
- TCP/IP – multiple NICs per subnet, dynamic routing
- DECnet Phase IV– L1 routers or end-node failover
- DECnet-Plus –Multi-homed ES or IS, load balancing

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

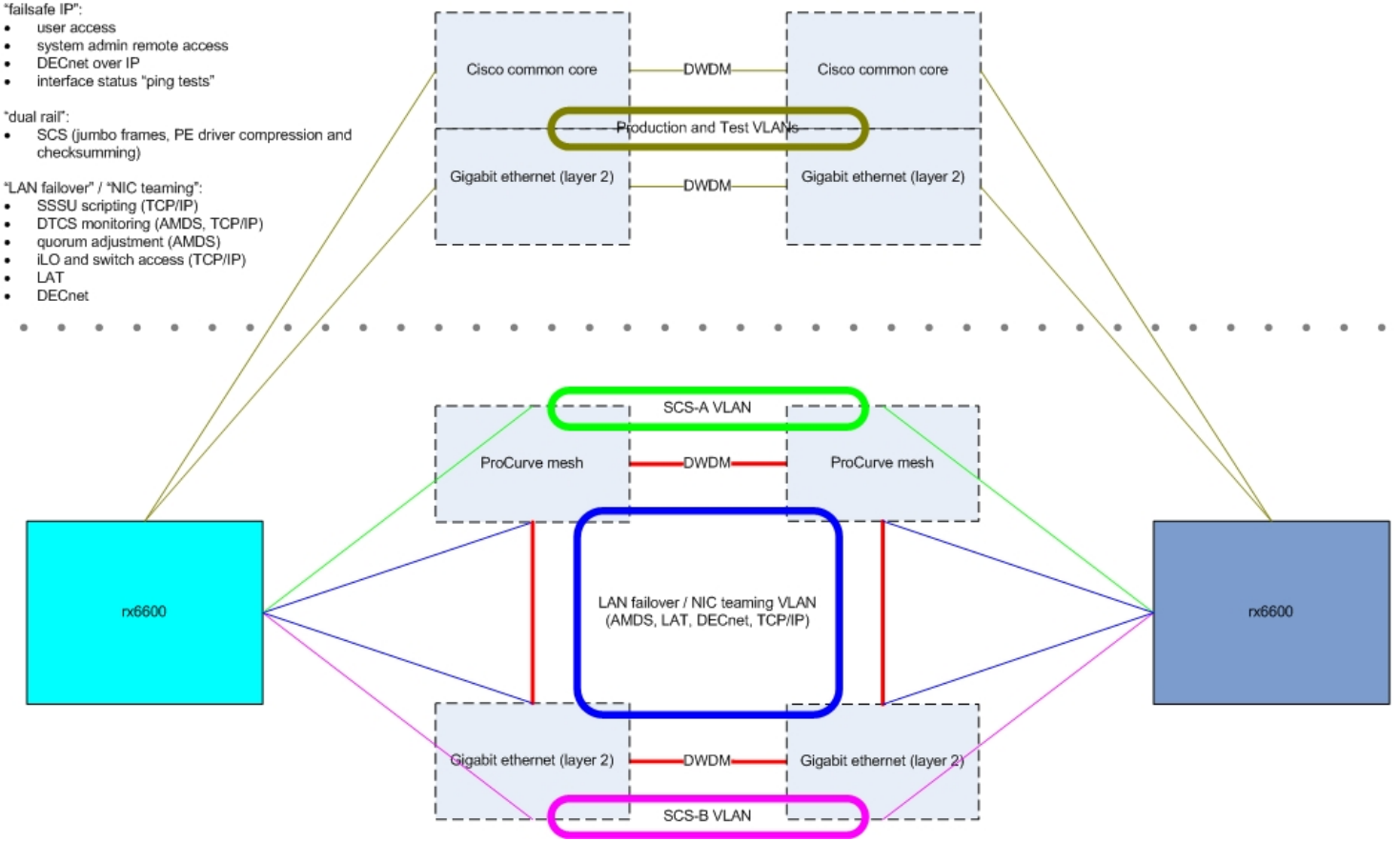


Common network infrastructure

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

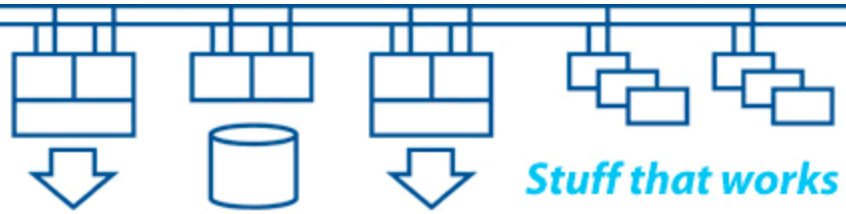
"dual rail":
- SCS (jumbo frames, PE driver compression and checksumming)

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

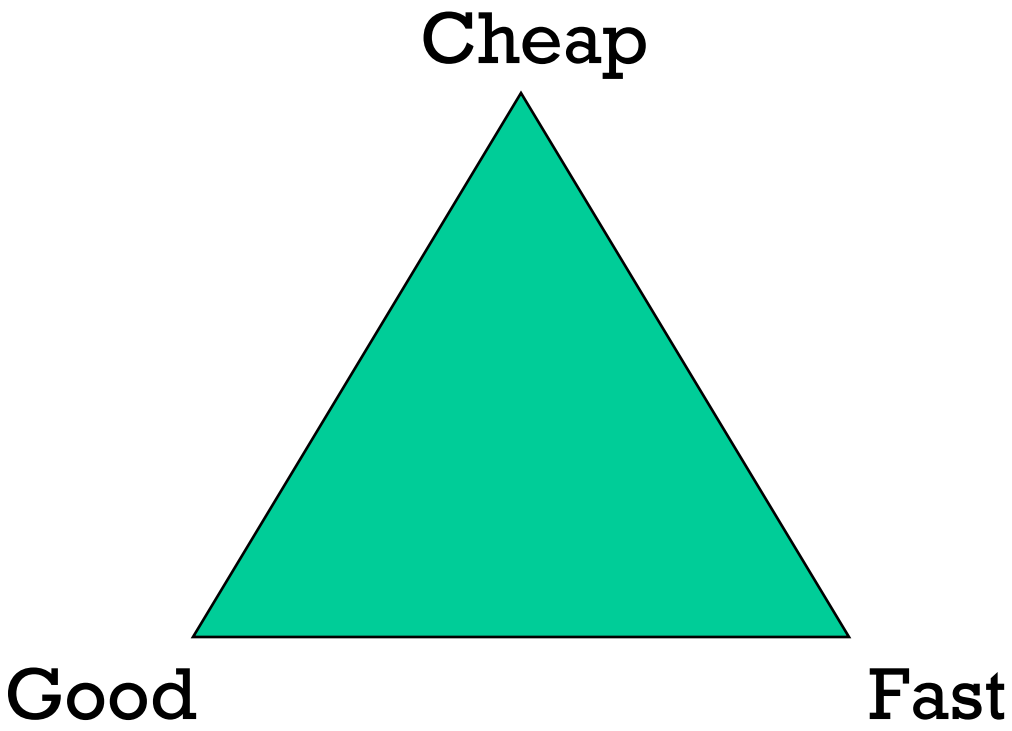


Converged ethernet and RNIC

- "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.

Discussion!