LAN Strategies for Performance and High Availability on HP-UX

Rick Petlin
Hewlett-Packard
rick_petlin@hp.com

Purpose

- Provide an overview of technologies and implementations.
- Review factors that affect LAN network performance and availability.
- Examine network strategies and technologies that can improve HP-UX LAN performance and availability.

LAN Strategies for Performance 2

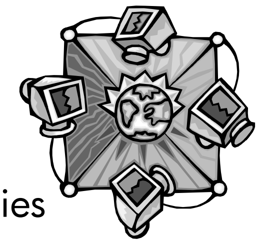
HP WORLD 2002 Conference & Expo

Bandwidth-hungry network applications are fueling the demand for more and more network performance, while mission-critical applications call for network designs that offer higher levels of availability. Enterprises are trying to keep ahead of this ever increasing demand for bandwidth and provide a reliable, fault-tolerant networking infrastructure. Solutions implemented today need to be scalable and protect future network investments.

System administrators now have a wide array of feature-rich HP-UX LAN products to select from to meet these evolving requirements. Key to addressing these demands is selecting the proper network configuration that meets the design requirements. Initial and recurring costs also need to be assessed when evaluating and selecting such a solution.

Often system networking requirements are dictated by current infrastructures, legacy networks, plans for future growth and needs for higher availability. There are many options to address these needs and with diverse options often comes varying degrees complexities.

Agenda



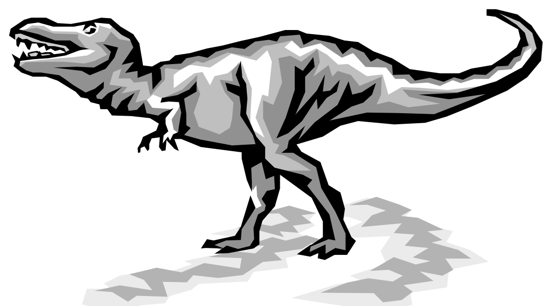
- Ethernet and Legacy LAN Technologies
- LAN link throughput expectations
- Fast-Ethernet and Giga-bit Ethernet
- Jumbo Frames for Performance
- Trunking of LAN Links for Performance and HA
- Virtual LANs
- HyperFabric
- Futures…

There are various HP-UX Ethernet and non-Ethernet technologies. Several will be covered, including Gigabit Ethernet, Ethernet Trunking (APA) that combines multiple link into one logical, LAN failover using LAN Monitor, and VLANs that allow one NIC to appear as many separate LAN interfaces. HP's cluster interconnect product, HyperFabric, will also be examined as an ultra-high speed LAN solution.

Best practices for selecting and implementing these technologies will be discussed, as will avoiding potential pitfalls.

Darwinism of LAN Technologies

- Survival of the fittest
  - Ethernet
  - Token-Ring
  - FDDI
  - ATM
  - 100VG



• **Ethernet** is by far the most ubiquitous of installed LAN technologies. Even with some of its limitations, Ethernet has survived and grown to now account for over 85% of all network connections. Invented by Dr. Robert Metcalf some 25+ years ago, Ethernet has evolved from 10 mega-bit base-band bus technology. Ethernet products are widely available in 100 mega-bit and 1 giga-bit high-speed fiber and copper based full-duplex switched technologies, and soon 10 giga-bit Ethernet product will be widely available.

• **Token-ring** (802.5) has been available in 4, 16 and in limited cases 100Mb. Token-ring token deterministic token passing ring technology had some advantages over Ethernet's early bus implementation. Token-ring had potential performance advantages with typically larger MTU (maximum transmission unit) of ~4K bytes, whereas Ethernet had been limited (in most cases) to 1500 bytes. Though token-ring is generally implemented as a multi-node ring of stations, the wiring is installed in a star from a central concentrator. Evolution of Ethernet speed, Ethernet switches, and reduction in Ethernet cost had reduced any real/perceived advantages token-ring may have once had over Ethernet.

• **FDDI**'s story is much the same as token-ring's. FDDI had speed advantages, 100Mb in days of 10Mb Ethernet and 4/16Mb token-ring. It also had a larger MTU advantage over Ethernet as well as much longer distance available because of its fiber implementations (copper implementation had limited availability.) An advantage FDDI had was the fault-tolerant option of dual rings that allowed station NICs to be potentially connected to two different FDDI concentrators for high-availability. This HA feature made it a favorite backbone and server connectivity technology. Ethernet's evolution had has eroded most FDDI advantages, with cost being one of the biggest factors.

• **ATM** is a far different technology, but has a similar story competing with Ethernet as a LAN technology. (ATM is extensively used as a WAN technology in telco environments.) 8-10 years ago ATM was said by many that it would all but replace Ethernet. Complexities and cost have limited its deployment to niche areas. Larger MTUs possible, but often LANE (LAN emulation) w/1500 byte MTUs was used so ATM node could talk to *"legacy Ethernet"* nodes. ATM edge devices, that connect ATM LANE to Ethernet, run rather sophisticated software stacks and are costly due to performance requirements.

• **100VG-AnyLAN…**

Link Speed ≠ Network Throughput

Max throughput expectations:

- XXXX Base-X is not promise of XXXX Mbit/sec
  - System CPU speeds, I/O bus architecture and DMA rates impact throughput
  - Application/Transport driving the connection Competing network activity from other nodes
  - Switched versus shared connections and topologies
  - Trunking (Auto Port Aggregation) load balancing
- Wide throughput variance in specific tests

• Server configuration, CPU performance, memory, application on many other factors come into play when determining actually server transmit or receive capabilities. Resulting performance may and most likely will vary from the link speed.

• The tools to measure throughput are typically UDP/TCP/IP based tools and vary greatly in their accuracy. You need to understand exactly what the test tool is doing. One also needs to know the limitations of each tool in measuring throughput is important.

• Quick and dirty, but course measurements of throughput:

> • **ftp** – file system and buffercache usage will influence results as will socket buffer sizing.
>
> • **rcp** – file system and buffercache usage will influence results. Socket buffer settings need to be tuned for the test.
>
> • **nfs copy** – influenced by local file system, buffercache, biod process load, mount type, mount protocol, and other processes using NFS….and that's just the client side. There are a similar list of issue on the server.

• More precise tools:

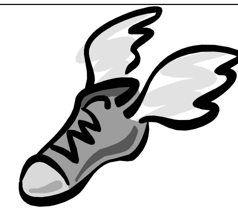
> • **ttcp** – "Test TCP" a simple sockets-based test tool that allows you to specify TCP or UDP transfers to various port number. It does not use disc buffercache so is a better test of pure network throughput, has tunable socket buffer settings and data size settings. It does the throughput calculations for you.
>
> • **netperf** – a benchmarking tool that can be used to measure the performance of many different types of networking. It provides tests for both unidirectional throughput, and end-to-end latency. Primary performance tool of HP-UX network labs.

• For more info on troubleshooting suspected network and application performance problems, refer to **Pat Kilfoyle's *"HP-UX Network Performance Tuning and Application Troubleshooting Tools"*** InterWorks 2002 presentation.

System Architecture

- I/O busses used in HP-UX systems
  - NIO/HPPB
  - EISA
  - HSC
  - PCI 1x, 2x, 4x
  - PCIx

Giga-bit Ethernet



- Two primary implementations
  - 1000Base-SX, fiber based
  - 1000Base-T, using common UTP cable
- Why pick one over the other?
- Cabling specifications

• The IEEE 802.3 working group that was responsible for the 802.3z and 802.3ab standards that specify the gigabit Ethernet implementations. 802.3z specifies the fiber-optic implementations, 1000Base-SX and 1000Base-LX, as well as the relatively unused copper 1000Base-CX.

  • 1000Base-SX is nearly identical at the link layer as Fibre Channel, using the same 8b/10b encoding method.

  • 1000Base-T 802.3ab spec was later ratified to support gigabit Ethernet over less expensive Cat-5 (category 5 and 5e as specified by EIA/TIE 568-A.) Topology rules for 1000Base-t are nearly identical as 100Base-T. (GbE does uses all 4–pairs in Cat-5 cable.)

• Selection based on need where fiber would have the distance advantage. Fiber can also be safely ran between buildings and often required by landlord for network connections between floors.

• There is virtually no difference in HP-UX GbE link performance between fiber and copper GbE NIC products. (Newer igelan driver based GbE fiber and copper links do offer better performance then previous generation gelan driver based NICS.)

• Option of installing 1000Base-T NICs now even if infrastructure limited to 100Mb.

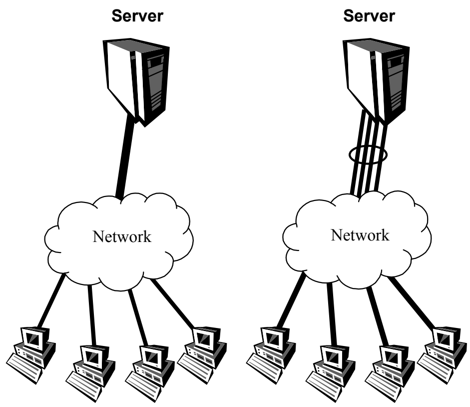
• GbE implementations are (virtually) always full-duplex switched or point-to-point. Spec allowed hub design but unused. Carrier extension needed to maintain 100m length if hub/half-duplex design was ever implemented.

IEEE Designation	Description	Cabling	Maximum Distance
1000Base-T IEEE 802.3ab	Long haul copper	UTP CAT5 or better	100 meters
1000Base-SX IEEE 802.3z	Short haul fiber	62.5u multi-mode or 50u multi-mode	275 meters* 550 meters*
1000Base-LX IEEE802.3z	Long haul fiber Typically not end nodes	9u single-mode	5000 meters
1000Base-CX IEEE802.3z	Short haul copper (virtually unused)	Shielded copper	25m

Jumbo Frames on GbE

- Jumbo Frames up the Ethernet MTU from 1500 to 9000 Bytes
- Reduced CPU overhead
- Increase NIC throughput
- Point-to-point and switched configuration
- All devices in network need to support Jumbo Frames.

Boosting Network and Server Access



- Do I upsize to the next faster link? 10Mb->100Mb->1Gb
- Or, use multiple slower links?
- These are the same questions for end nodes as for the network infrastructure
- Design requirements

Network connectivity bandwidth is often as important as system speed in today's computing environment. The increase demands for added throughput have both server and network administrators deploying various methods to get the most out of their networks and servers.

Boosting Network and Server Access often mean installing the next high speed available link technology. In the case of Ethernet we have seen the evolution from 10 mega-bit to 100 mega-bit and GbE is widely available and 10GbE is on the horizon. Jumping to the next higher speed to increase may not always be an option or the preferred route.

Boosting access may mean adding multiple slower links to both increase performance and availability.

The diagram show two simple high level network configurations. In the first one the server link was increased to a higher speed, whereas in the second, multiple slower links are added to the system. It should be pointed out that the two same options are available to the network administrator for network infrastructure needs. Most mid-size and larger Ethernet switches have the option of bonding multiple links, between switches, into Ethernet trunks also referred to as **Link Aggregation.**

Why not just a higher speed link?

- Cost effective to trunk multiple lower speed links
- End-system may not be able to utilize the higher speed
  - System may not need a 10X boost in network speed
  - Available copper links are more pervasive for lower speed links
  - 10/100Mbps NICs and switch port are generally very cheap
- Protect investment in existing infrastructure
- Multiple links may provide higher availability and resiliency

Replacing slower speed links, to servers or within the LAN infrastructure, with higher speed links may come at too high of a cost. Increasing link speeds has many potential cost associated with the boosting speed. On the system side not only are you going to have to replace the NIC card, but the associated cost of the NIC can be many times the cost of the existing link speed. Depending on the implementation, new cabling or fiber maybe needed.

Jumping up **10 times** in link performance may not always be a supported option depending on the system and that system's configuration. Just because a new faster link, 10 time the speed of the previous link, is installed *does not* mean that the system can fully utilize the available bandwidth. The link speed is only the clock rate at which a frame of data is sent or received. Server configuration, CPU performance, memory, application on many other factors come into play when determining actually server transmit or receive capabilities. Your mileage and resulting performance may and most likely will vary from the link speed.

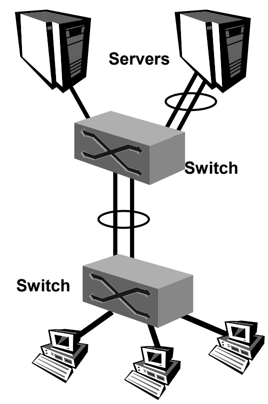
On the infrastructure side of the picture, new and more expensive blades (interfaces) for the switch maybe needed. Also, both the density and availability of the higher speed ports (i.e. giga-bit) is generally much less for the higher speed ports.

As we noted on the previous slide, availability of the system could be the goal and not just speed. Link Aggregation can potentially provide both a higher level of network performance while increasing reliability for servers and network infrastructure.

Ethernet Link Aggregation is scalable and had been used with 10Mb, 100Mb, 1Gb, and in the near future, 10Giga-bit links.

Load Balancing the Link Aggregate

- Layer 2 implementation requires distribution of complete frames
- Load distribution algorithm generally based on MAC addresses
- Other distribution algorithms useful depending on configurations
- Distribution *attempts* to balance load
- Must not mis-order frames
- Must not send duplicate frames



LAN Strategies for Performance

11

HP WORLD 2002 Conference & Expo

Frame distribution and load balancing are two of the most important concepts in link aggregation to grasp.

Only whole frames are distributed. A frame is ***not*** cut up in to pieces and transmitted down all links for re-assembly at the other end.

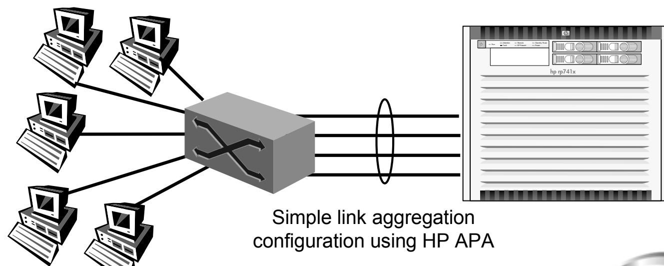
It is also important to maintain the order of frames in a **conversation** between end nodes. A conversation is the set of MAC frames exchanged between a pair of end stations, where all of the MAC frames form a part of an ordered sequence, and where there exists a requirement for ordering to be maintained among the set of MAC frames exchanged. A conversation may be uni-directional or bi-directional. There may be more than one conversation in progress between a given pair of end stations at any one time; similarly, a given end station may take part in conversations with more than one end station at any one time.

The **Load Balancing Algorithm** determines which link of a switch a frame was send out. A fairly typical load balancing algorithm would use the source and destination address pair of the frame to be transmitted. An X-OR operation is performed on the last two bits of the source MAC address and the destination MAC address. This operation yields one of four results: (0 0), (0 1), (1 0), or (1 1). Each of these values points to a link in the link aggregate.

• **The load balancing algorithm selection may vary per aggregate or even per direction of traffic flow. Selecting the proper load balancing algorithm for a given configuration is key in obtaining the best frame distribution and thus performance.**

Auto Port Aggregation: HP-UX's Link Aggregate Implementation

- Aggregates multiple physical LAN links into one logical link
- APA includes both load sharing link aggregates as well as fail-over aggregates.



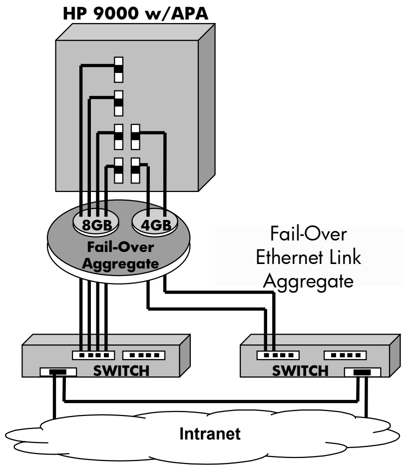
Simple link aggregation configuration using HP APA

Hewlett-Packard's Auto Port Aggregation (APA) increases a server's efficiency by grouping or "aggregating" multiple ports into a single logical link aggregate or fail-over aggregate having a single IP address. Up to five fail-over aggregates per computer are permitted on HP-UX 10.20 and up to a total of **50** aggregates are supported on HP-UX 11.0.

APA on HP-UX 11.0 and 11i (11.11) supports Fast and Gigabit Ethernet with automatic fault detection and recovery as well as load balancing of network traffic, while APA on HP-UX 10.20 only supports the automatic link failure detection and recovery feature for Fast and Gigabit Ethernet.

FDDI and Token Ring links can also be configured as fail-over Link Aggregates on both HP-UX 10.20, 11.0 and 11i, but with no load balancing of network traffic. This enables network managers to use APA throughout their environment to provide higher levels of system and network availability**.**

APA LAN Monitor for 11.0/11i



- LAN Monitor introduced June 2000
- Simple, low cost single system HA link fail-over solution w/o MC/Service Guard complexity and expense
- LAN Monitor Aggregates primary and standby links can be made of individual links or logical link aggregates*

LAN Monitor mode provides a simple, low cost single system high availability solution for HP-UX. It is link fail-over without the complexity and expense of MC/ServiceGuard.

Links in a HP-UX 11.0/11i LAN Monitor fail-over aggregate can be made up of individual 100Mb Ethernet, 1Gb Ethernet, Token-Ring, FDDI or multi-link aggregates. The links comprising the fail-over aggregate must be made up of the same technology.

The user commands are similar to MC/Service Guard commands.

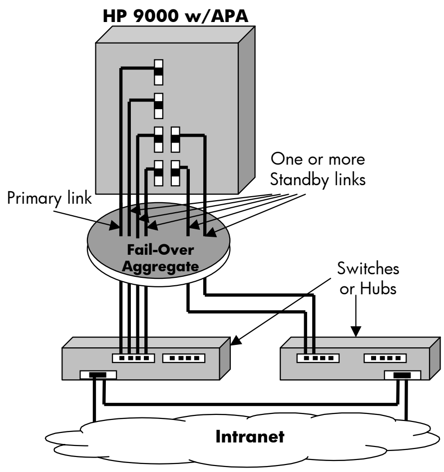
lanqueryconf - Scan existing subnets and determine possible fail-over links. Create /etc/lanmon/lanconfig.ascii file.

lanapplyconf - Apply the contents of the /etc/lanmon/lanconfig.ascii file to the system (I.e. create the fail-over groups).

landeleteconf - Delete the fail-over groups defined in the /etc/lanmon/lanconfig.ascii file from the system.

lancheckconf - Verify the configurations defined in the /etc/lanmon/lanconfig.ascii file.

Auto Port Aggregation for 10.20



- LAN Monitor mode only
- Primary and standby are single links only (not link aggregates)
- Simple, low cost Single System HA network link solution
- Link fail-over w/o MC/Service Guard complexity and expense
- Support for 100Mb, 1Gb, FDDI, Token-Ring
- Supports all hubs and switches

• HP-UX 10.20 supports only the LAN Monitor mode of the HP Auto-Port Aggregation product. The LAN Monitor feature of the HP Auto-Port Aggregation software provides a “Hot-Standby” capability with MC/SG-like configuration tools.

• LAN Monitor created Fail-over groups do not support MC/SG at first release.

• In the event of a link failure, the LAN Monitor software will automatically migrate the data flow from the primary link to one of the standby links in the “Link Aggregate.”

• Links in a HP-UX 10.20 LAN Monitor fail-over aggregate can be made up of *individual* 100Mb Ethernet, 1Gb Ethernet, Token-Ring, or FDDI. (but not multi-link aggregates.) The links comprising the fail-over aggregate must be made up of the same technology.

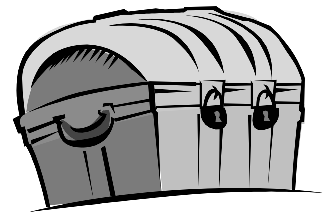
• The LAN Monitor product is available on both HP-UX 10.20 (and 11.0) beginning with the June’00 release (DART 49) and at introduction of 11i (11.11.)

APA Summary Features

- Very good solution for one-to-many or many-to-many connections
- Incremental bandwidth
- Protect investment
- LAN/Monitor is excellent for link failover

Trunking Potential limiting features:

- Frame distribution limits available bandwidth per connection to speed of a single link.
- Additional cabling to install and manage.
- Troubleshooting multiple LAN links on both system and switch.
- Some added management cost to implement.



LAN Strategies for Performance

16

HP WORLD 2002 Conference & Expo

• Though there are multiple frame distributions methods, a single connection will always be limited to the speed of a single link in the aggregate. For example; system “A” and system “B” each have a APA aggregate trunk of four 100Base-T connecting to a single switch.

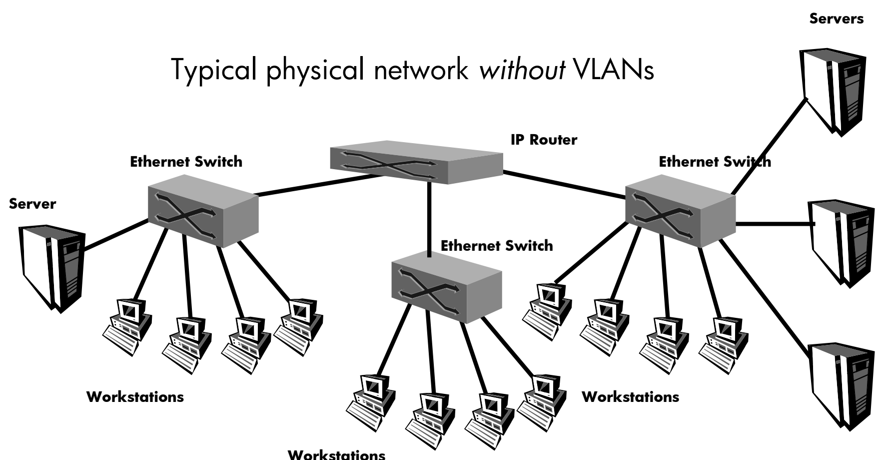
> •The fast a single connection will ever be able to transmit or receive data between these two systems is 100Mb.
>
> • Only when multiple data streams between the two systems are established ***AND*** the system ***AND*** switch (*if capable)* are configure for tcp/udp port based distribution, will the total transferred rate “A” to “B” or “B” to “A” exceed the speed of one link in the aggregate.

• Obvious incremental cost include:

> • Managing the installation and cost of cable.
>
> • Some increase in skills needed to troubleshoot multilink problems
>
> • APA on HP-UX is a licensed product

Introduction to Virtual LANs

Typical physical network *without* VLANs



• Typically network managers use one LAN for each IP subnet in their network and communication between subnets is made possible at the Network layer or "Layer 3", using IP routers. Multiple NICs could also be used to connect servers and/or workstations to the multiple IP subnets.

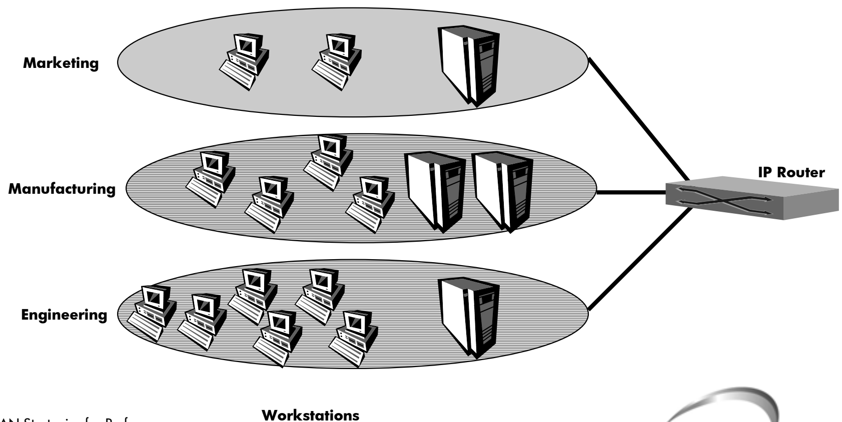
• Often IP subnets are defined along physical boundaries such as buildings, floors, or section of an office. These physical boundaries sometimes force compromises in segmenting what might be a more natural logical segmentation of desirable sub-nets.

• This network configuration may cause a very high packet load to be placed on the router at the center of this network. This could acceptable, but may not be optimal.

• This section is a introduction into how segmenting LANs into Virtual LANs, or VLANs, can be of benefit system and network planner.

Logical Network by Department

- Logical Network by Department



Above is the same set of servers and workstation configured in a network with logical partitioning. The IP subnets, in this example have been aligned by departments. VLANs were enabled and assigned to the switch ports where these nodes were connected.

Logical Network Partitioning

- High level network design
- Implemented with VLAN aware switches
- Explicit and implicate VLAN association
- Each VLAN identified by VLAN ID
- Individual switch ports configured to belong to one or more VLAN

• A high level design could define separate networks along department boundaries. A good network map is of current and planned network implementation is a must.

• Typically core network switches have the ability to support VLANs which allow the switch to be “cut-up” into multiple separate virtual LANs. VLAN aware switch and even hubs can be apart of a network implementing VLANs though those devices would not participate classifying which VLAN network a frame belongs.

• VLAN-aware switches classify inbound frames based on their contents or by the physical port in with the frame arrived. Explicit association includes physical switch port and the frame being “tagged” with a specific identifier indicating which VLAN it was a member. Implicate association would include feature of the frame such as MAC or IP address, protocol (IPv4, IPv6, IPX) or tcp/udp port numbers.

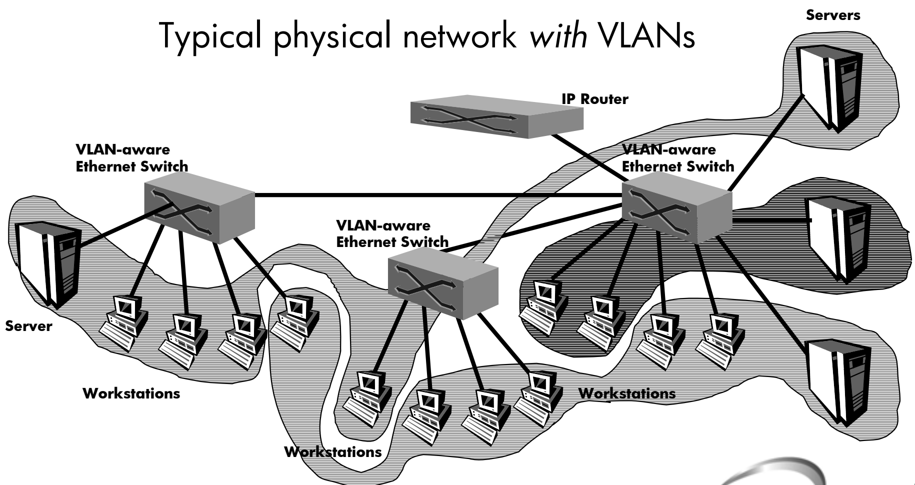
• Each VLAN segment on a switch has a VLAN identification number that groups that set of ports. VLAN ID Number can be sent and received over the link connections to other switch or end-nodes.

• A switch port can have one or more VLAN associations. More about VLAN tags in a few slides.

• Each VLAN on a switch is treated as a separate layer 2 network. Unicast frames frames are sent to a destination port only if port is configure in same VLAN. Broadcast and multicast frame are flooded only to port on the same VLAN where the broad/multicast originated.

Typical VLAN implementation

Typical physical network *with* VLANs



LAN Strategies for Performance 20

• Virtual LAN (VLAN) technology allows us to separate logical network connectivity from physical connectivity. This concept is different from a traditional LAN in that a LAN is limited by its physical connectivity. All users in a LAN belong to a *single* broadcast domain and can communicate with each other at the Data Link layer. Network managers have used LANs to segment a complex network into smaller units for better manageability and improved performance.

• A VLAN may be thought of as a single physical network infrastructure that can be logically divided into discrete LANs that can operate independent of each other. The logical partitioning is done based on various attributes of data frames traversing the network or by simply which physical port a station is plugged into.

• In the above picture, VLANs are configured on the VLAN capable switches. In this sample, the end station ports are assigned to one of the three VLANs. The links between the switches are configured to ***tag*** the traffic leaving the switch to identify which VLAN the data is associated with in the network. This is called ***explicit tagging***. This means that the en

• A lesser used VLAN implementations will use frame/packet content (MAC address, Ethernet type, IP address, tcp/udp port, etc) to ***implicitly*** assign a frame/packet to a VLAN

• Note that the router is still part of the network to route between the networks. Only one link is depicted indicating the router also understands the VLAN tags and that the router is configure in the three subnets. If the router did not support VLAN tagging, three separate LAN links from the router could be connected to each VLAN subnet.

• VLAN switch configurations can associate a given port to one or more VLANs. In this example only the switch-to-switch and switch-to-router links are using tagging.

Benefits of VLANs

- Manageability (console or closet)
- Enhanced Security
- Bandwidth Preservation
- Better use of server resources
- Link consolidation

Manageability: Flexible network segmentation. Moves, adds, and changes to network topology do not require physical changes to network topology. User mobility is much simpler because of the dynamic nature of VLANs.

An end node location can be moved to another sub-net without cabling changes. The switch port is reassigned to another VLAN via management console verses trip to wiring closet. A "software patch panel" to reconfigure a single end-station or a whole sub-net.

If tagging for a given VLAN is already enabled on a connected switch port, the end node can be moved or added to that sub-net with no configuration changes on the switch.

Enhanced Security: Different security domains may be easily constructed to provide various levels of security in the network, since the network design is more flexible than traditional LANs. VLANs create virtual boundaries that can only be crossed through a router. So standard, router-based security measures can be used to restrict access to each VLAN as required

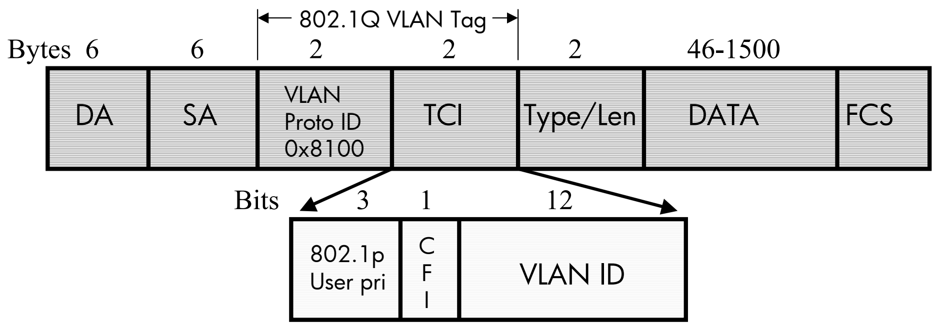
Bandwidth Preservation: A well-designed VLAN implementation will help restrict broadcast and multicast traffic to only those stations that are listening to and responding to that traffic. The network and computing resources of non-participating stations are unaffected, improving performance.

Better use of server resources: With a VLAN-enabled adapter, a server (end-node or router) can be a member of multiple VLANs. This reduces the need to route traffic to and from the server through a router.

Link consolidation: Reduction in number of NICs and I/O slots needed in a given server.

VLAN Tagging

- Diagram of frame with 802.1 Q/p tag



• Most VLAN network implementations use explicit tagging to ID VLAN association.

• The VLAN tag above identifies which VLAN a data frame belongs. Different VLAN ID numbers to and enables traffic from more than one VLAN to use the same switch or LAN card port.

• VLAN ID is a 12 bit value, 0 to 4095

• Distinction between VLAN ID and number of supported VLAN in a given device. A switch might support 16 VLANs but VLAN ID of 4000 would be valid.

• When a VLAN-aware switch receives data from an end-station, the switch determines where the data is to go and whether the VLAN ID should be retained. If the data is to go to a device that can recognize the VLAN tag, the VLAN tag is retained. If the data is to going to a device that has no knowledge of VLANs (VLAN-unaware), the switch typically sends the data without the VLAN tag.

• Network switches and end stations that know about VLANs are said to be VLAN-aware. Network switches and end stations that can interpret VLAN tags are said to be VLAN-tag-aware.

• The extra 4 bytes added to the frame, to the tag, are generally not a problem for devices that have no knowledge of VLANs.

• HP-UX VLAN-tag-aware end stations will add appropriate VLAN tags to standard Ethernet frames, a process called explicit tagging. This is only needed if the end station needs to be a connected to multiple VLANs. If the end station only needs to be connected to one VLAN, it could use the untagged LAN instance and that switch port could then be assigned to that one specific VLAN on the switch.

HP-UX VLAN Features

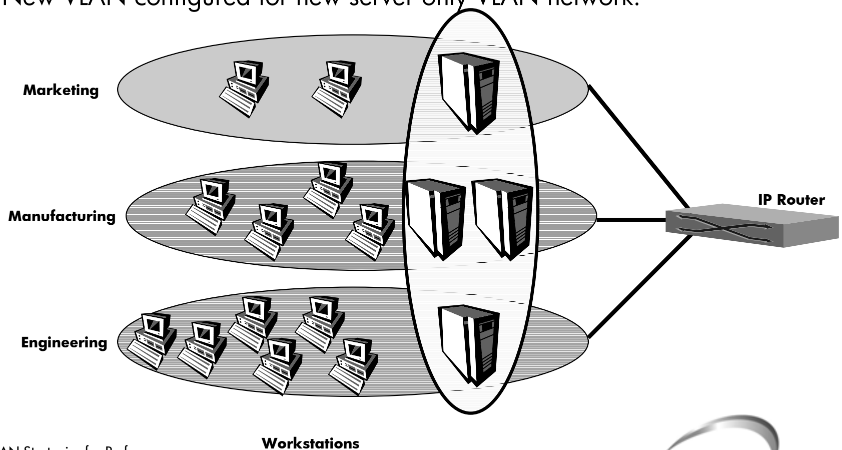
- Host-base 802.1Q/p tagging supported on 11i March 2002 w/patches.
- Supported on HP's HSC & PCI Fast Ethernet and Giga-bit Ethernet NIC cards
- Up to 1024 VLANs per NIC port
- MC/Serviceguard Support
- Tagged Virtual LANs configured via SAM or directly via lanadmin

- HP-UX 11i (11i release 1.0 or 11.11) Networking patches (links, transport, streams…) are need to enable VLAN tagging. This is a feature enhancement to the core HP-UX networking feature and not a separate product. VLAN supporting networking patches available in the June '02 11i HP-UX Quality Patch release. As always, check for patch updates and compatibilities.
- All add-on 10/100 Fast Ethernet (btlan driver) and Giga-bit Ethernet NICs (gelan and new igelan diver based cards,) built-in or add-on NICs can support VLANs. This does exclude any non-Ethernet NICs as-well-as EISA or HP-PB Fast Ethernet NICs.
- HP-UX 11i supports 1024 VLANS (should be enough for most.) The VLAN ID is a 12 bit value (0-4095) and any of the 1024 HP-UV VLANs can be in that 4k range of VLAN Ids. With so many VLANs available, reduction in the number of NICs is possible in some configurations.
- VLANs and MC/Serviceguard support support with SG release 11.14. Currently failover with LanMonitor is ***not supported***. VLAN Trunk aggregates are also currently ***not supported***.
- Existing tools and commands are user to configure new VLAN. "lanadmin –V" or SAM is used to create a new virtual LAN instance. Once the new LAN instance is configured an IP address can be assigned, just as if is was a separate NIC card. The file /etc/rc.config.d/vlanconf is used to preserve/store VLAN information across reboots.

```
#lanscan
Hardware Station        Crd  Hdw   Net-Interface    NM   MAC        HP-DLPI DLPI
Path     Address        In#  State NamePPA          ID   Type       Support Mjr#
8/0/20/0 0x001083052286 0    UP    lan0 snap0       1    ETHER        Yes   119
#lanadmin -V create vlanid 150 vppa 5001 0
#lanscan
Hardware Station        Crd  Hdw   Net-Interface    NM   MAC        HP-DLPI DLPI
Path     Address        In#  State NamePPA          ID   Type       Support Mjr#
8/0/20/0 0x001083052286 0    UP    lan0 snap0       1    ETHER        Yes   119
VLAN5001 0x001083052286 5001 UP    lan5001 snap5001 3    ETHER        Yes   119
```

Adding New VLAN Subnet

New VLAN configured for new server only VLAN network.



After segmenting the IP network into departments from our original example, a new server only IP subnet VLAN is configured for just server-to-server traffic. This was done to reduce the packet load on the router. The typical few steps needed to implement this configuration would be something such as

1) Plan implementation
2) Select new VLAN ID #
3) Configure VLAN on switches
4) Added new VLAN to switch-to-switch port
5) Add new VLAN to server switch ports as a tagged VLAN, leaving original server’s connected VLAN/sub-net as untagged.
6) Configure new virtual LAN instance on server.
7) Configure new IP VLAN via ifconfig or SAM on server.
8) Repeat for all system.
9) Test connectivity.

VLAN Summary

- Allows flexible network configuration
- Potential improvements in overall network and system throughput.
- No cost enhancement to 11i
- Future evolution of HP VLANs

HP is investigating in the following areas for improvements to the HP-UX VLAN product:

• Integration of VLAN and auto port aggregation (APA). APA is HP's Link Aggregation software. Efforts are underway to enable customers to create VLANs on link aggregates.

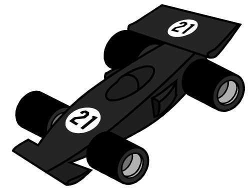
• GVRP and Automatic Configuration. GVRP is an IEEE protocol that allows a switch or an end station to advertise its VLAN membership to its link partner. Using this mechanism, we could develop a mechanism for dynamically assigning VLAN membership to end stations, so that you don't need to manually assign VLAN IDs to each NIC on an end station.

• Stack support for 802.1p/Cos/QoS (multi-queues). HP is investigating methods for implementing an end-to-end Class of Service or Quality of Service solution by improving on priority mechanisms in the network stack and NICs. An important component of this solution will be the 802.1p mechanism.

• Application-based VLAN. Application-based VLANs provide the most flexible way for configuring VLANs—VLAN-aware applications determine the membership of the traffic they generate. This mechanism opens up a number of interesting possibilities. For example, a set of stations may negotiate a dynamically created VLAN for the purpose of carrying on a short-term audio or videoconference.

• HP-UX VLAN implementation will be a key value addition to many new technologies on the horizon, such as iSCSI, 10-Gigabit Ethernet, IPv6, etc.

What is HyperFabric?



- HyperFabric is a high speed low latency system interconnect .
  - Uses switched fabric technology
  - Designed to provide large cluster solutions

HyperFabric Benefits

- Improved throughput by increasing Bandwidth and reducing Latency
- Hyper Fabric provides complete End to End High Availability by implementing
  - Dynamic routing
  - Active-active High Availability
  - Transparent fail over at link level
  - Makes it ideal platform to run mission critical applications such as ERP, DSS. Eg: SAP
- Increased Scalability

Improved throughput

Bandwidth: capable of transmitting 2 + 2 Gbps full duplex

Perfect for network based back-up consolidation solution

Latency: Switched fabric architecture of HyperFabric drastically reduces latency.

Helps compute intensive, latency sensitive Technical Computing applications

Enhanced Availability

Hyper Fabric provides complete End to End High Availability by implementing

Dynamic routing

Active-active High Availability

Transparent fail over at link level

Makes it ideal platform to run mission critical applications such as ERP, DSS. Eg: SAP

Increased Scalability

Almost linear scaling with Hyper Messaging Protocol (HMP).

Drastically improves throughput of scalable database applications such as Oracle (OPS, RAC.)

Points to consider…

- Requirements should drive design.
- Check product features and compatibility.
- Read product release notes.
- Install with latest product version and check install patches and patch dependencies.
- Get familiar with new features before rolling into production.

LAN Strategies for Performance

28

HP WORLD 2002 Conference & Expo

Too often a overly complex design drives the requirements hardware, software and management resources.

Increase link speed may cause processing load on another system ….

Industry Futures

- 10GB
- TOE
- iSCSI

more information…

- www.docs.hp.com/hpux/netcom/index.html
  - Check HP-UX network performance white papers
- www.hp.com/products1/unixserverconnectivity/adapters/index.html
  - HP-UX network connectivity products
- www.hp.com/go/network_city
  - Switch technologies and case studies
- www.cisco.com/warp/public/473/4.html
  - Cisco's write-up on their various families of switches and routers distribution methods
- http://www.10gea.org/
  - Information about current 1Gigabit Ethernet as well a future technologies inc; 10Gbe, TCO, iSCSI

• Detailed 100mb and 1Gb performance information on various platforms. Whitepapers on VLAN and gigabit technologies can also be found under these first two URLs.

• HP's Network City web site has HP network hub and switch products as well as LAN technology whitepapers and case studies.

• Cisco's web site has a wealth of network technology information. Of particular interest to those planning to implement trunking (HP-APA, FEC/LACP) is the paper found at this link titled: "Understanding EtherChannel Load Balancing and Redundancy on Catalyst Switches"

• Another good cisco document on GbE: http://www.cisco.com/warp/public/cc/techno/media/lan/gig/tech/1000b_sd.htm

• Cabling specifications http://www.tiaonline.org/

• For more information on HP-UX specific network tuning and troubleshooting see Pat Kilfoyle's 2002 InterWorks Presentation titled "HP-UX Network Performance Tuning and Application Troubleshooting Tools."

Please feel free to contact me if you have any questions regarding this material or subject matter. I can be reached at rick_petlin@hp.com.