Clarification Regarding Load-Balancing and Trunking

on HP ProCurve Routing Switches

The increasing demand on network bandwidth to transfer data between clients and servers or between two networking devices led to the idea of combining multiple physical links into one logical link. This point-to-point connection between devices is called a trunk and can be viewed as one logical link or "big pipe" having the same characteristics as the single links that comprise it. Besides the increased bandwidth, the trunk also offers redundancy in case of individual link failures. In figure 1, below, the four-port trunk-group consists of four single, Gigabit links forming one single, logical link (trunk). The theoretical maximum throughput for this trunk group is eight Gigabits/sec (full-duplex mode only).

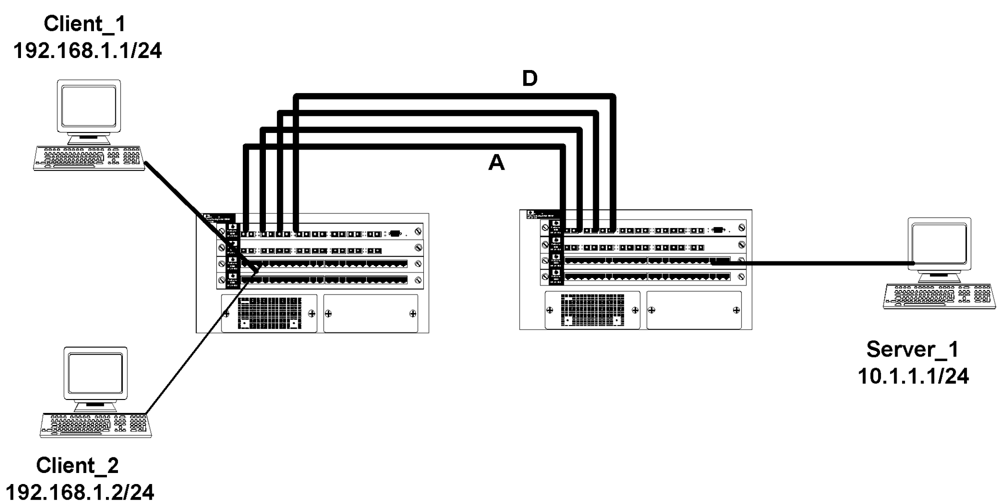

Figure 1 HP Routing Switch Trunk

Note Regarding Load-Balancing: There has been a general misunderstanding of how trunking (especially load-balancing) is implemented on the routing switches. Load balancing is based on a "per communication" flow. Looking at figure 1, if Client_1 transfers four packets of data to Server_1, it uses physical link A to send the data between the two routing switches. All four packets take the same path. It is a common misunderstanding that assumes the first packet takes link A, the second packet link B, the third packet link C, and so on. If this were true, ensuring that the packets remained in sequence would create a heavy load on the routing switch CPU and would require all vendors to implement the same algorithm to ensure packet sequencing. However, in practice, most vendors have implemented a solution based on communication flow between devices and not on packet distribution across trunked links. To actually balance traffic across multiple physical links, some vendors have implemented proprietary solutions that use either a "round-robin" scheme or another mechanism to balance traffic. However, this type of solution requires equipment from the same vendor on each end of the trunk.

When trunking is configured on the routing switches, two options are available: "trunk switch" and "trunk server" . "Trunk server" is typically used when the trunk is terminated by a server and "trunk switch" is typically used when the trunk is terminated by a switch or routing switch. The behavior of the two differs in the following way:

Trunk-Server Method

Traffic load-balancing is based on the destination MAC address for all protocols. When the routing switch forwards traffic across the trunk, it uses the destination MAC address to load-balance traffic. This option is designed for servers with a unique MAC address on each NIC in the trunk. If the server uses a single MAC address for the trunk, the trunk-server option is ineffective for load balancing because all traffic from the routing switch to the server goes over only one of the available links.

Trunk-Switch Method

Traffic load-balancing is based on different criteria for different protocols, as follows:

Protocol Type	Criteria
IP	Destination IP Address
IPX	Destination IPX Address
AppleTalk	Destination AppleTalk Address
Other	Destination MAC Address

Regardless of the trunk type, the failover functionality will provide redundancy for the connection.

The following example describes the traffic flow between the clients and servers using the Trunk-Switch method between the switches. This example covers only IP traffic.

Example:

In figure 2, Client_1, Client_2, and Client_3 are transferring large files to Server_1, and Client_4 is requesting a file via ftp from Server_2.

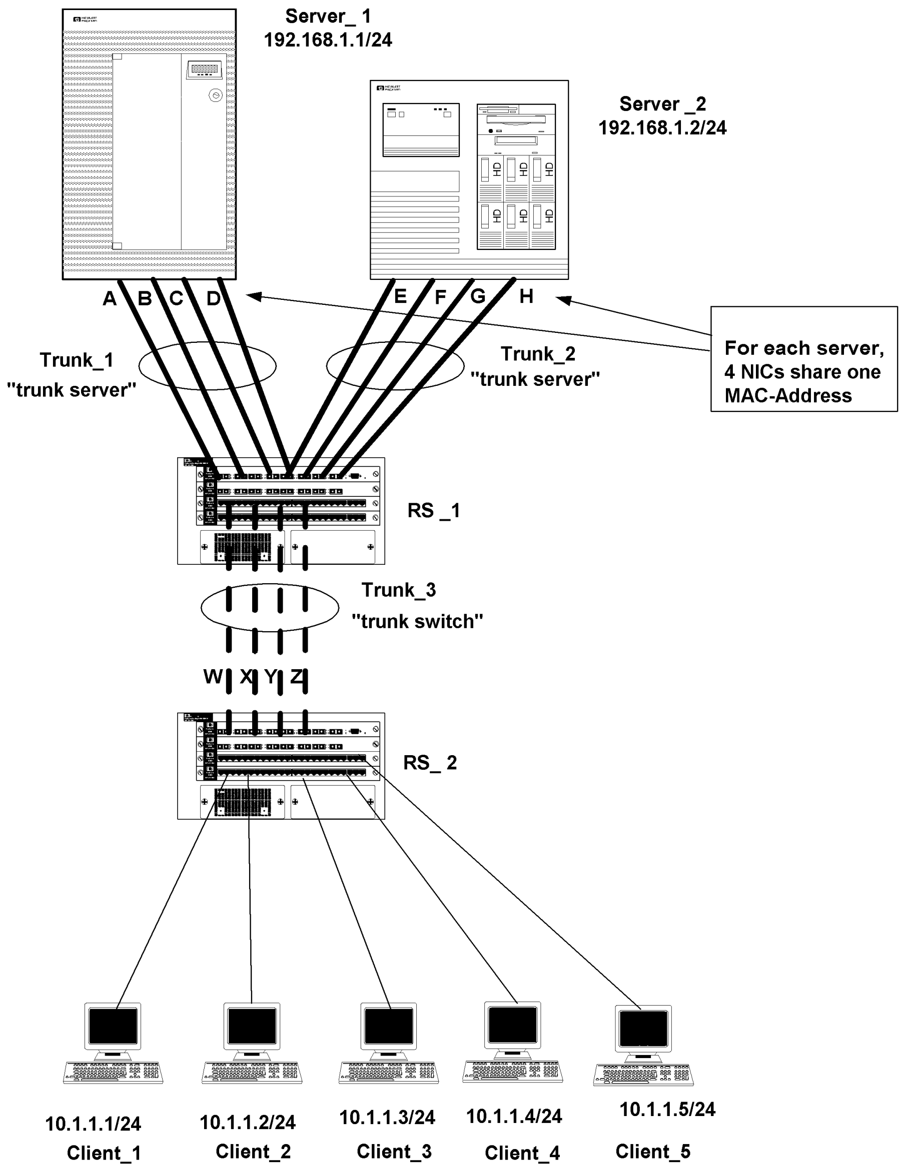

Figure 2 IP Traffic Flow Using Both the Trunk-Switch and the Trunk-Server Methods

Description of traffic-flow from clients to servers:

Traffic from Client_1, Client_2, and Client_3 uses link W of Trunk_3 and link A on Trunk_1.

Explanation: Trunk_3 is configured for "trunk switch" which means, load balancing is based on destination IP address. Since all three clients communicate with the same server IP address, only one link is used. Traffic sent from RS_1 to Server_1 (over Trunk_1, "trunk server") uses physical link A only, since the trunk has only one MAC address.

Traffic from Client_4 uses link W on Trunk_3 and link E on Trunk_2.

Explanation: RS_2 also uses link W on Trunk_3 because, by default, CAM entries are made for the destination network, not the host. Since Server_1 and Server_2 are on the same IP network, the same physical link is used. This behavior can be changed by using the global **ip load-sharing by-host** command (not previously documented). Version 7.1.X of the routing switch software allows more flexibility in defining how networks and hosts are balanced across a trunk. For a complete description of this feature consult the routing switch documentation for version 7.1.X.

Description of traffic flow from servers to clients:

Traffic from Server_1 to Client_1, Client_2, and Client_3 should use different physical links from Server_1 to RS_1 (depending on server implementation). Traffic from RS_1 to RS_2 uses link X on Trunk_3. By default, all traffic takes the same path since the destination clients are on the same network. Again, this behavior can be changed using the **ip load-sharing by-host** command.

Traffic from Server_2 to Client_4 follows the same pattern. Load distribution on Trunk_2 depends on the trunking implementation of the server. Since the destination host is on the same network as Client_1, Client_2, and Client_3, traffic from RS_1 to RS_2 also uses link X if the global command **ip load-sharing by-host** has not been issued .

Again, regardless of which physical link is used, redundancy will be provided. If link A on Trunk_1 fails, all clients will then use link B.

HP Part Number: 5969-2372
Edition 1, January 2001