Competitive SAN Local Switching Solution: Limiting Scalability, Not Improving It

Use Case for Local Switching

Local switching is switching in which traffic does not traverse the backplane of a switch but rather stays on a line card. This concept is positioned as a method for increasing the available bandwidth of a line card or reducing traffic on the backplane of a device. While the concept of local switching is sound and local switching has been successfully deployed on many types of devices, not all local switching implementations are equally effective.

Competitive Architecture

The architecture of the SAN solution that is competitive with the Cisco® SAN solution is a three-tier stack of common application-specific integrated circuits (ASICs); see Figure 1. This architecture has the advantages of being simple and quickly built. However, it also has serious limitations due to routing decisions that occur independently at each ASIC.

Figure 1.   Competitive Director Architecture
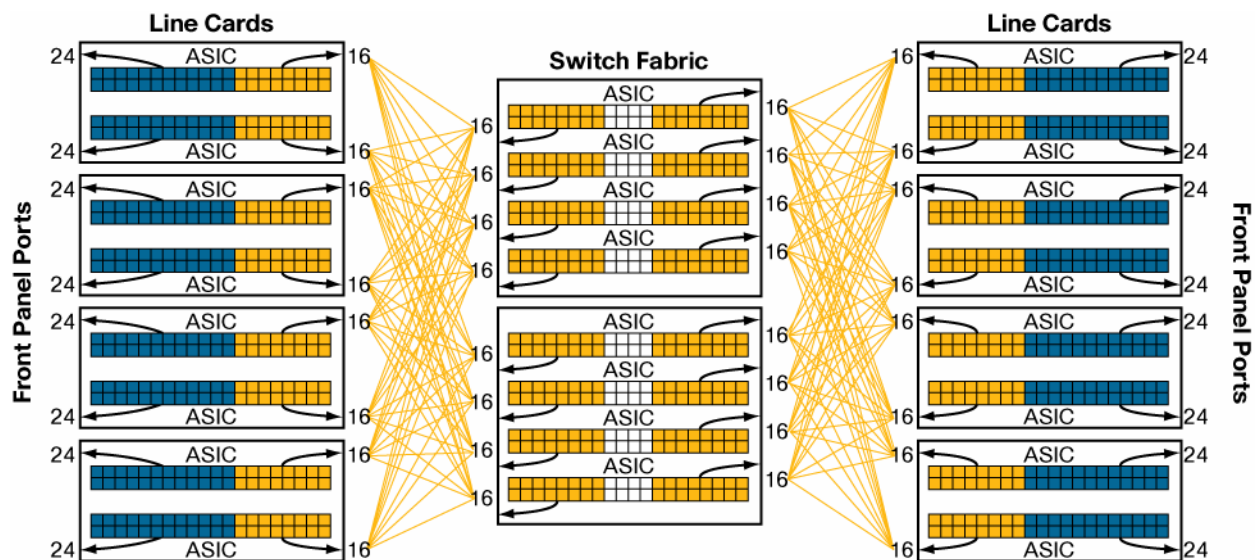
Each ASIC in this architecture makes autonomous routing decisions based on individual instances of the Fabric Shortest Path First (FSPF) routing protocol. Since each ASIC is independent, one ASIC cannot adjust traffic paths to accommodate load on another ASIC on the same line card, the same switching fabric, or another line card. This limitation leads to performance problems when one autonomous routing decision affects another autonomous routing decision.

Architecture and Local Switching

To fully understand local switching, you must understand the confines of the local switch. Local switching can be defined as the switching of traffic in a local domain. This domain may be an entire line card or, as in the case of competitive SAN solutions, only half of the ports of a line card (Figure 2).

Figure 2. Competitive 48-Port Line Card Architecture
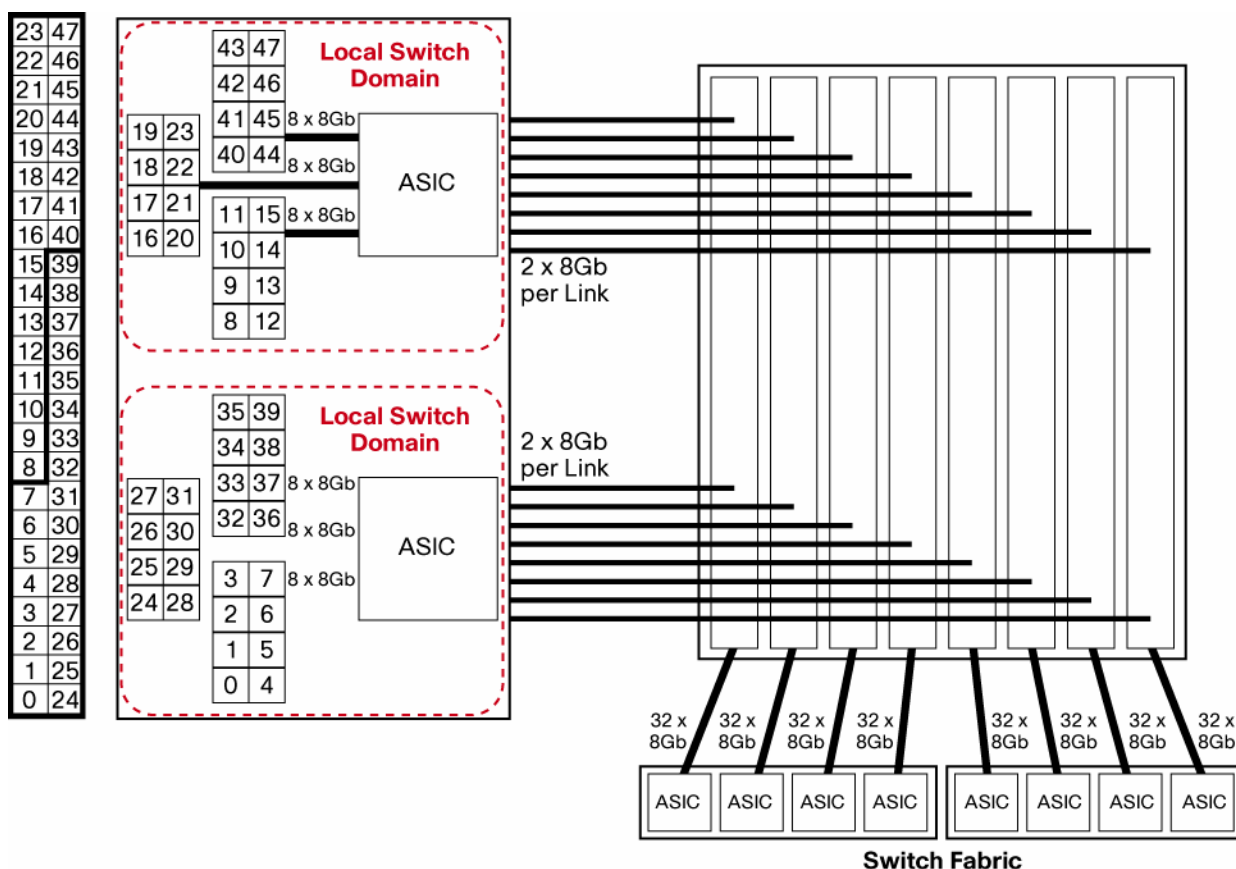
In Figure 2, the first local switch domains consist of line card ports 0 to 7 and 24 to 39, and the second local switching domain consists of line card ports 8 to 23 and 40 to 47. This numbering scheme is neither contiguous nor intuitive and may lead to deployment errors when you try to implement local switching.

Performance Problems When Mixing Local and Non-Local Switching

Numerous tests have been conducted using different combinations of storage traffic that mix locally switched and non-locally switched traffic. The results of these tests indicate that, in some cases, locally switched traffic is affected by non-locally switched traffic, and in other cases, non-locally switched traffic is affected by locally switched traffic. Results are further complicated by routing policies that are configurable on the competitive switch. Figures 3 and 4 show some examples of traffic throughput inequities when locally switched traffic and non-locally switched traffic are mixed.

Figure 3.   Local Switch Test 1
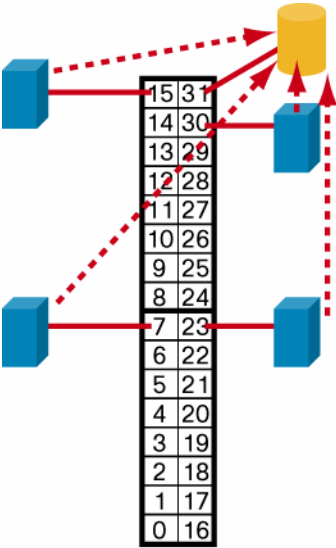
Traffic Flows:

Port 15 (I) to Port 31 (T)—Locally Switched
Port 30 (I) to Port 31 (T)—Locally Switched
Port 7 (I) to Port 31 (T)—Non-Locally Switched
Port 23 (I) to Port 31 (T)—Non-Locally Switched

All Initiators Should Get 210 MB of Throughput

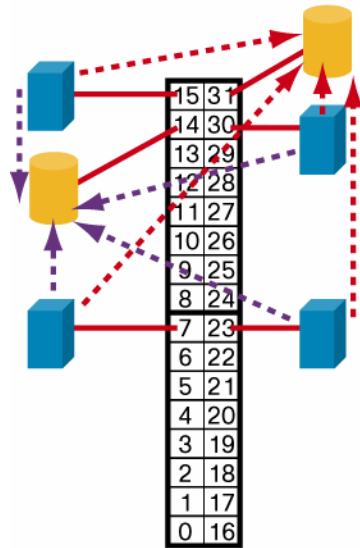| Flow	Throughput	Latency
Port 30 to Port 31	110 MB	154 Microseconds
Port 15 to Port 31	110 MB	154 Microseconds
Port 23 to Port 31	310 MB	383 Microseconds
Port 7 to Port 31	310 MB	383 Microseconds
Total to Port 31	840 MB	

Exchange-Based Routing / Locally Switched Traffic Affected

Flow	Throughput	Latency
Port 30 to Port 31	240 MB	69 Microseconds
Port 15 to Port 31	240 MB	69 Microseconds
Port 23 to Port 31	180 MB	230 Microseconds
Port 7 to Port 31	180 MB	230 Microseconds
Total to Port 31	840 MB	

Port-Based Routing / Non-Locally Switched Traffic Affected

Figure 4.   Local Switch Test 2

Traffic Flows:

Port 15 (I) to Port 31 (T)—Locally Switched
Port 30 (I) to Port 31 (T)—Locally Switched
Port 7 (I) to Port 31 (T)—Non-Locally Switched
Port 23 (I) to Port 31 (T)—Non-Locally Switched

Port 15 (I) to Port 14 (T)—Locally Switched
Port 30 (I) to Port 14 (T)—Locally Switched
Port 7 (I) to Port 14 (T)—Non-Locally Switched
Port 23 (I) to Port 14 (T)—Non-Locally Switched

All Initiators Should Get 210 MB of Throughput

	Flow	Throughput	Latency
Exchange-Based Routing	Port 30 to Port 31	76 MB	135 Microseconds
	Port 15 to Port 31	76 MB	135 Microseconds
	Port 23 to Port 31	344 MB	268 Microseconds
	Port 7 to Port 31	344 MB	268 Microseconds
	Total to Port 31	840 MB	
Locally Switched Traffic Affected	Port 30 to Port 14	76 MB	86 Microseconds
	Port 15 to Port 14	76 MB	86 Microseconds
	Port 23 to Port 14	344 MB	114 Microseconds
	Port 7 to Port 14	344 MB	114 Microseconds
	Total to Port 14	840 MB	

	Flow	Throughput	Latency
Port-Based Routing	Port 30 to Port 31	168 MB	48 Microseconds
	Port 15 to Port 31	168 MB	48 Microseconds
	Port 23 to Port 31	252 MB	128 Microseconds
	Port 7 to Port 31	252 MB	128 Microseconds
	Total to Port 31	840 MB	
Locally Switched Traffic Affected	Port 30 to Port 14	168 MB	48 Microseconds
	Port 15 to Port 14	168 MB	48 Microseconds
	Port 23 to Port 14	252 MB	128 Microseconds
	Port 7 to Port 14	252 MB	128 Microseconds
	Total to Port 14	840 MB	

As can be seen from the test results, mixing locally switched traffic and non-locally switched traffic causes inconsistent traffic flows and affects the performance of some devices. Because the effect of the performance degradation is inconsistent, a SAN architect cannot design around this limitation and must instead require that the user never mix local switched and non-local switched traffic on competitive SAN directors.

Effect of Port Density

One potential benefit of local switching is greater port density with higher per-port speeds. This benefit is achieved by having some traffic locally switched and other traffic non-locally switched. As the test results previously described showed, this scenario is not possible with competitive SAN directors. Some traffic is given preferential bandwidth,

while other traffic is negatively affected. Because of this behavior, traffic should not be mixed between locally switched and non-locally switched. Because of the performance problems, this behavior also means that use of local switching limits the number of ports deployable per line card.

Typical SAN deployments use fan-out ratios to determine the number of servers (initiators) per storage array (targets) ports. Storage vendors determine these fan-out ratios based on server type, application performance profile, and storage array port performance. These fan-out ratios vary, but typical ratios range from 8:1 (eight servers to one storage array port) to 20:1 or more. Using these values, Table 1 shows how the inability to mix locally switched and non-locally switched traffic leaves some ports unusable.

Table 1.   Local Switch Port Deployments

Fan-Out Ratio	Total Host Ports and Array Ports	32-Port Line Card Usable Ports		32-Port Line Card Unusable Ports		48-Port Line Card Usable Ports		48-Port Line Card Unusable Ports	
		Per Port Group	Per Line Card	Per Port Group	Per Line Card	Per Port Group	Per Line Card	Per Port Group	Per Line Card
8:1	9	9	18	7	14	18	36	6	12
10:1	11	11	22	5	10	22	44	2	4
12:1	13	13	26	3	6	13	26	11	22
14:1	15	15	30	1	2	15	30	9	18
16:1	17	Cannot Deploy				17	34	7	14
18:1	19					19	38	5	10
20:1	21					21	42	3	6

As can be seen in Table 1, the number of ports that cannot be deployed in local switched deployments can vary from as few as 2 ports (6 percent) of a 32-port line card to as many as 22 ports (46 percent) of a 48-port line card, depending on the initiator-to-target fan-out ratio. This limitation additionally affects the customer because this solution does not allow for network growth without extensive recabling of the entire competitive switch, ultimately limiting the number of deployable ports on the competitive SAN solution and requiring more chassis to be deployed.

Local Switching and Virtual Machine Deployments

Today's data center is increasingly using virtualization to reduce the number of servers being deployed and to increase the utilization of the deployed servers. This virtualization is typically deployed in clustered solutions, grouping 8, 16, 32, or more servers in a cluster. Application performance for the servers must be consistent and must not vary based on the server on which the virtual machine is running.

One requirement of the server clusters is that all servers have the same access to the same storage, meaning that a SAN administrator must ensure that all server ports and storage ports are in the same local switch domain. This requirement cannot be met on the competitive SAN solution due to the small size of the local switch domain (a maximum of 24 ports). In addition, as performance tests have shown, mixing traffic between local switch domains causes performance inconsistencies that are in direct conflict with the requirement for equal performance for all servers in a cluster.

Local Switching and Cluster High Availability

Another consideration when looking at local switching is the fault domain of the solution. When initiators and targets are dispersed over multiple port groups and line cards, the failure of an ASIC or line card will affect only a small number of the devices in a cluster. When local switching is deployed, the failure of a single ASIC or line card will cause failure in an entire cluster of servers and storage devices.

Conclusion

Although local switching is a valid method for potentially increasing performance, the usability of such of feature is ultimately determined by a vendor's implementation. The competitive SAN local switch implementation imposes too many restrictions to be used successfully:

- The number of ports that can be deployed when using local switching is limited.
- Performance degrades when locally switched and non-locally switched traffic are mixed.
- The performance effect is unpredictable and cannot be designed around.
- Deployments of virtual servers and clusters are limited.
- High availability is compromised, and a single point of failure is created.