Enhancing Network Availability and Performance on the Dell PowerEdge 1855 Blade Server Using Network Teaming

Network interface card teaming and LAN on Motherboard teaming can provide organizations with a cost-effective method to quickly and easily enhance network reliability and throughput. This article discusses network teaming on the Dell™ PowerEdge™ 1855 blade server and expected functionality using different network configurations.

The modular Dell PowerEdge 1855 blade server integrates up to 10 server blades into a highly dense, highly integrated 7U enclosure, including two Gigabit Ethernet switch modules or Gigabit Ethernet pass-through modules and an embedded remote management module. Each integrated switch module is an independent Layer 2 switching device, with 10 internal ports connected to the integrated LAN on Motherboards (LOMs) on the server blades and six 10/100/1000 Mbps external uplink ports (a 10:6 ratio). Each integrated pass-through module is an independent 10-port device directly connecting the individual server blade’s integrated LOM to a dedicated external RJ-45 port (a 1:1 ratio). Unlike a switch module, a pass-through module can connect only to 1000 Mbps ports on external switches; pass-through modules do not offer support for 10 Mbps or 100 Mbps connections.

Each server blade in the Dell PowerEdge 1855 blade server has two embedded LOMs based on a single dual-port Intel® 82546GB Gigabit Ethernet Controller. The two LOMs reside on a 64-bit, 100 MHz Peripheral Component Interconnect Extended (PCI-X) bus. The LOMs are hardwired to the internal ports of the integrated switches or pass-through modules over the midplane, a passive board with connectors and electrical traces that connects the server blades in the front of the chassis with the infrastructure in the rear. The LOMs provide dedicated 1000 Mbps full-duplex connections (see Figure 1). LOM 1 on each server blade connects to an internal port of switch 1 or pass-through module 1, and LOM 2 on each server blade connects to the counterpart port of switch 2 or pass-through module 2. Note: The second switch or pass-through module is optional. However, two installed switches or pass-through modules can enable additional connectivity or network redundancy and fault tolerance provided that the limitations and capabilities of these features are fully understood and implemented correctly.

One important distinction between a blade server and other types of servers is that the connection between the LOM and internal ports of the integrated I/O module (switch or pass-through) is hardwired through the midplane. This design enables the link between the LOM and the integrated switch to be almost always in an up, or connected, state—unless either a LOM or a switch port fails. The link can remain active even in the absence of a network connection between the external uplink ports on the integrated switch and the external network.

1This term does not connote an actual operating speed of 1 Gbps. For high-speed transmission, connection to a Gigabit Ethernet server and network infrastructure is required.
The connection between the LOM and internal ports of the integrated I/O module is critical in a network teaming scenario where the integrated switch is being used. This is because, to trigger a failover event, the teaming software is looking for only the loss of the link between the LOM and the first switch. (Failures to cables or switch ports outside the enclosure do not trigger failover events.)

The scenario is different if a pass-through module is used. With a pass-through module, the link is in a connected state only if a network connection exists between the external ports on the pass-through module and the switch port outside the enclosure—as is true for a stand-alone server. Thus, for the pass-through module, the teaming software triggers a failover event if the LOM, pass-through port, cable, or external switch port fails.

**NIC and LOM teaming benefits**

Network interface card (NIC) teaming and LOM teaming can help ensure high availability and enhance network performance. Teaming combines two or more physical NICs or LOMs on a single server into a single logical device, or virtual adapter, to which one IP address can be assigned. The teaming approach requires teaming software—also known as an intermediate driver—to gather physical adapters into a team that behaves as a single virtual adapter. In the case of the Dell PowerEdge 1855 blade server, the intermediate driver is Intel Advanced Networking Services (iANS). Intermediate drivers serve as a wrapper around one or more base drivers, providing an interface between the base driver and the network protocol stack. In this way, the intermediate driver gains control over which packets are sent to which physical interface on the NICs or LOMs as well as other properties essential to teaming.

A virtual adapter can help provide fault tolerance (depending on the type of failure) and bandwidth aggregation (depending on the teaming mode selected). If one of the physical NICs or LOMs—or the internal switch ports to which they are connected—fails, then the IP address remains accessible because it is bound to the logical device instead of to a single physical NIC or LOM.

One important distinction between a blade server and other types of servers is that the connection between the LOM and internal ports of the integrated I/O module (switch or pass-through) is hardwired through the midplane.

To provide fault tolerance, when a team is created iANS designates one physical adapter in the team as the primary adapter and the remaining adapters as secondary. If the primary adapter fails, a secondary adapter assumes the primary adapter’s duties. There are two types of primary adapter:

- **Default primary adapter**: If the administrator does not specify a preferred primary adapter in the team, iANS chooses an adapter of the highest capability (based on model and speed) to act as the default primary adapter. If a failover occurs, a secondary adapter becomes the new primary. Once the problem with the original primary is resolved, traffic does not automatically restore to the original default primary adapter. The restored adapter will, however, rejoin the team as a secondary adapter.

- **Preferred primary adapter**: The administrator can specify a preferred adapter using the Intel PROSet tool. Under normal conditions, the primary adapter handles all traffic. The secondary adapter receives fallback traffic if the primary fails. If the preferred primary adapter fails but is later restored to an active status, iANS automatically switches control back to the preferred primary adapter.

Teaming features include failover protection, traffic balancing among team members, and bandwidth increases through aggregation. **Note**: Administrators can configure any team by using the Intel PROSet tool, which provides configuration capability for fault tolerance, load balancing, and link aggregation.

**Fault tolerance.** Fault tolerance provides NIC or LOM redundancy by designating a primary adapter within the virtual adapter and utilizing the remaining adapters as backups. This feature is designed to ensure server availability to the network. When a primary adapter loses its link, the intermediate driver fails traffic over to the secondary adapter. When the preferred primary adapter’s link is restored, the intermediate driver fails traffic back to the primary adapter.
The intermediate driver uses link-based tolerance and probe packets to detect network connection failures, depending on the teaming mode selected:

- **Link-based tolerance**: Using this mechanism, the intermediate driver checks the link status of the local network interfaces belonging to the team members. Link-based tolerance provides failover and failback for physical adapter link failures only.
- **Probe packets**: Probing is another mechanism used to maintain the status of the adapters in a fault-tolerant team. Probe packets are sent to establish known, minimum traffic between adapters in a team. At each probe interval, each adapter in the team sends a probe packet to other adapters in the team. Probing provides failover and failback for physical adapter link failures as well as external network failures in the single network path of the probes between the team members.

**Note**: When only two members exist in the team, the intermediate driver uses link-based tolerance and receive-traffic activity to detect network connection failures. Probing is not used for failover or failback in teams that have only two members.

**Load balancing.** Adaptive Load Balancing (ALB) mode in iANS provides transmission load balancing by dividing outgoing traffic among all the NICs or LOMs, and can shift traffic away from any NIC or LOM that goes out of service. Receive Load Balancing (RLB) mode in iANS balances receive traffic.

**Link aggregation.** This feature combines several physical channels into one logical channel. Link aggregation is similar to ALB, and is available in two modes: Static Link Aggregation mode supports Cisco Fast EtherChannel (FEC) and Cisco Gigabit EtherChannel (GEC), while IEEE 802.3ad mode supports the IEEE 802.3ad standard.

**Note**: Link aggregation is not supported on PowerEdge 1855 blade servers configured with integrated switches because of the server hardware design implementation. For more details, refer to the “Static Link Aggregation” and “IEEE 802.3ad” sections in this article. Other blade server systems that have hardware and network configurations similar to the PowerEdge 1855 blade server tend to function in a similar manner.

**iANS teaming software architecture**

iANS supports the following teaming modes:

- Adapter Fault Tolerance (AFT)
- Switch Fault Tolerance (SFT)
- Adaptive Load Balancing and Receive Load Balancing
- Static Link Aggregation (SLA)
- IEEE 802.3ad

Figure 2 is the teaming support matrix for PowerEdge 1855 blade servers.

**Adapter Fault Tolerance**

AFT enables automatic recovery from a link failure caused by a failure in a physical NIC or LOM, and internal ports on the integrated switch, by redistributing the traffic load across a backup adapter. Failures are detected automatically, and traffic rerouting takes place as soon as the failure is detected. The goal of AFT is to ensure that load redistribution takes place fast enough to prevent user protocol sessions from being disconnected. AFT supports two to eight adapters (any combination of NICs and LOMs) per team. Only one active team member transmits and receives traffic. If this primary connection fails, a secondary, or backup, adapter takes over. After a failover, if the connection to the primary adapter is restored, control passes automatically back to the primary adapter. AFT is the default mode when a team is created using iANS; however, this mode does not provide load balancing.

The Intel intermediate driver uses the immediate physical loss of a link—either between the NIC or LOM and the integrated switch or between the NIC or LOM, the pass-through module inside the server enclosure, the cable, and the port of the external switch to which it is connected—to trigger a failover when there are only two adapters in the team. The “X’s” in Figure 3a and Figure 3b indicate the failover and failback for link failures. Notice that, in Figure 3a, a failover between the LOM and the external port link does not occur for switches. This is because, as previously discussed, the internal

---

**Figure 2: Teaming support matrix for the Dell PowerEdge 1855 blade server**

<table>
<thead>
<tr>
<th>Team mode</th>
<th>Integrated switches</th>
<th>Integrated pass-throughs</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFT</td>
<td>Yes*</td>
<td>Yes</td>
</tr>
<tr>
<td>SFT</td>
<td>Yes*</td>
<td>Yes</td>
</tr>
<tr>
<td>ALB and RLB</td>
<td>Yes*</td>
<td>Yes</td>
</tr>
<tr>
<td>SLA**</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>IEEE 802.3ad**</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*This mode cannot detect connectivity failures outside the chassis.

**SLA and IEEE 802.3ad are not supported on Dell PowerEdge 1855 blade servers when using integrated switches because the two NICs or LOMs on the blade are hardened to different integrated switches. However, these modes are supported when using the pass-through modules.
ports of the integrated switches are hardwired to the LOMs through the midplane of the enclosure; therefore, the intermediate driver does not detect a loss of link.

**Switch Fault Tolerance**

SFT functionality is similar to AFT except that SFT supports only two NICs or LOMs in a team connected to two different switches. In SFT mode, one adapter is the primary adapter and one adapter is the secondary adapter. During normal operation, the secondary adapter is in standby mode. In its standby state, the adapter is inactive and waiting for failover to occur, and the adapter does not transmit or receive network traffic. If the primary adapter loses connectivity, the secondary adapter automatically takes over.

The local network interfaces in Figure 3a show the links between the LOMs and the internal ports of the integrated switches, and Figure 3b shows the links between the LOMs and the internal ports of the pass-through modules.

If a team has more than two members, enabling probe packets can help detect network connection failures, along with the link-based method of fault tolerance. When probes are enabled, the primary adapter will send the probe packets to secondary adapters and vice versa at administrator-defined time intervals. When the primary adapter’s probe packets are not received by the secondary adapters, the intermediate driver attempts retry mechanisms before eventually failing over. Probe packets can help detect network failures in the single network path of probe packets between the primary and secondary adapters.

**Additional network redundancy.** Because the AFT or SFT teams cannot probe the link status of the integrated switch’s external uplink ports, or link failures elsewhere in the network, additional protection may be required. Administrators can achieve increased network redundancy by adding redundant links between the integrated switches and external switches or by aggregating at least two of the uplink ports using the Link Aggregate Control Protocol (LACP). (Note: The external switches must support and be configured for 802.3ad link aggregation.) When properly configured with link aggregation and adapter fault-tolerance teams, the PowerEdge 1855 blade server is designed to maintain network connectivity if any of the LOMs, internal switches, or switch uplink ports fail.

The PowerEdge 1855 blade server can be configured to provide a highly redundant network environment when plugged into a single external switch. However, administrators may want to provide a further level of redundancy to protect against the failure of an external switch. Figure 4 shows the necessary network connections for the PowerEdge 1855 blade server in such an environment. Administrators can implement this redundancy by creating two dual-port link aggregation groups between each external switch and integrated switches 1 and 2. In this configuration, even if an external switch fails, individual server blades maintain network connectivity with the help of Spanning Tree Protocol (STP) in the switches. Configuration of STP is very important for this network configuration to work reliably.

**Adaptive Load Balancing and Receive Load Balancing**

ALB is a method for dynamic distribution of data traffic load among multiple physical channels. The purpose of ALB is to improve overall bandwidth and end-node performance. The ALB approach provides multiple links from the server to the switch, and the intermediate driver running on the server performs the load-balancing function.
NEW-GENERATION SERVER TECHNOLOGY

The ALB architecture uses knowledge of Layer 3 information, such as IP address, to achieve optimum distribution of the server transmission load.

ALB is implemented by assigning one of the physical channels as “primary” and all other physical channels as “secondary.” Packets leaving the server can use any one of the physical channels, but incoming packets can use only the primary channel. When enabled, RLB balances the IP receive traffic. The intermediate driver analyzes the send and transmit loading on each physical adapter in the team and balances the rate across all adapters based on the destination address. Adapter teams configured for ALB also provide the benefits of fault tolerance, as described in the “Adapter Fault Tolerance” section in this article.

Static Link Aggregation
Link aggregation is similar to ALB in that it combines several physical channels into one logical channel. The Intel intermediate driver supports link aggregation for FEC and GEC.

FEC is a trunking technology developed by Cisco to aggregate bandwidth between switches working in Fast Ethernet. Using FEC, administrators can group multiple switch ports together to provide additional bandwidth. Switch software treats the grouped ports as a single logical port. Administrators can connect an end node, such as a high-speed server, to the switch using FEC.

The PowerEdge 1855 blade server can be configured to provide a highly redundant network environment when plugged into a single external switch.

FEC link aggregation provides load balancing in a similar manner to ALB, including the use of the same algorithm in the transmit flow.

The transmission speed for FEC does not exceed the adapter base speed to any single address; teams must match the capability of the switch. Adapter teams configured for SLA also provide the benefits of fault tolerance and load balancing. When using SLA mode, administrators are not required to set a preferred primary adapter. All adapters in an SLA team must run at the same speed and must be connected to an SLA-capable switch. If the speed capability of adapters in an SLA team is different from the speed of the switch, the speed of the team is dependent on the switch. SLA teaming requires that the switch be set up for SLA teaming and that STP be turned off.

GEC link aggregation is essentially the same as FEC link aggregation, except that GEC supports 1000 Mbps speeds versus 100 Mbps for FEC.

IEEE 802.3ad
IEEE 802.3ad is the IEEE standard for the technology incorporated in the Cisco FEC standard. Intel’s intermediate driver support for IEEE 802.3ad is similar to its FEC and GEC support. Administrators can configure a maximum of two IEEE 802.3ad dynamic teams per server, and the configuration must use 802.3ad-capable switches in dynamic mode. Adapter teams configured for IEEE 802.3ad also provide the benefits of fault tolerance and load balancing. The 802.3ad teaming mode allows all network communication protocols to be load balanced.

Dynamic mode supports multiple aggregators, and adapters should operate at the same speed. Only one team can be active at a time.

The advantages of modular computing
The Dell PowerEdge 1855 blade server can help organizations deliver on the promise of modular computing—namely lower acquisition cost, lower total cost of ownership, rack density, and power efficiency—all without trading off enterprise-class features. Moreover, by understanding the unique aspects of the PowerEdge 1855 blade server’s architecture, as described in this article, administrators can create a highly reliable network infrastructure.

Mike J. Roberts is responsible for blade server product planning at Dell. Mike has a B.A. from Marquette University and an M.B.A. from The University of Texas at Austin.

Doug Wallingford is a sustaining engineer in the Enterprise International Product Support department at Dell. Doug has 14 years of experience in network field support, seven years as a Xerox field engineer, and seven years at Dell. He has numerous Microsoft® and Novell® certifications—including Microsoft Certified Systems Engineer (MCSE), Microsoft Certified Systems Administrator (MCSA), and Certified Novell Administrator (CNA).

Balaji Mittapalli is a development engineer in the Server Networking and Communications department at Dell. Balaji has an M.S. in Electrical Engineering from The University of Texas at San Antonio and a B.S. in Electronics and Communication Engineering from S.V. University in India. Balaji’s areas of interest are wired and wireless networks.