Network Link Aggregation Practices

with the Dell PowerEdge 1855 Blade Server

The networking architecture integrated on the Dell™ PowerEdge™ 1855 blade server enables significant flexibility and resource consolidation. This article describes the relationship between the optional Dell PowerConnect™ 5316M integrated Gigabit Ethernet switch when configured for link aggregation and the Dell PowerEdge 1855 blade server’s networking architecture in enterprise networking environments.

BY BRUCE HOLMES

Offered as an option on the Dell PowerEdge 1855 blade server, the Dell PowerConnect 5316M integrated Gigabit Ethernet switch module is an independent Layer 2 switching device and a member of the Dell PowerConnect 5300 Gigabit Ethernet switch family. In highly integrated networking architectures, it is common practice to use an industry-standard feature called link aggregation, which is defined by the IEEE 802.3ad standard, to configure the optional integrated Gigabit Ethernet switch on the PowerEdge 1855 blade server so that two or more of its external ports are treated as a single logical port.

Aggregated ports offer two main advantages:

• The available bandwidth to and from the switch can be increased with each additional port added to the aggregated link.
• Adding ports to the aggregated link can provide port failover protection.

The Dell PowerEdge 1855 blade server network architecture is shown in Figure 1. Each server blade has a single dual-port LAN on Motherboard (LOM) controller integrated into its motherboard. Each server blade’s LOM is connected through the chassis midplane to the chassis I/O bays in the rear. LOM 1 is connected to the module installed in the NET 1 I/O slot: in this example, PowerConnect 5316M integrated Gigabit Ethernet switch 1. A matching PowerConnect 5316M integrated Gigabit Ethernet switch is installed in the NET 2 I/O slot, and LOM 2 is connected to this second integrated switch. Each PowerConnect 5316M integrated Gigabit Ethernet switch has 10 internal ports (one for each server blade) and six external ports.

Note: Administrators must populate the NET 1 and NET 2 I/O slots with matching connectivity modules; that is, the slots must both have either optional PowerConnect 5316M integrated Gigabit Ethernet switches or the standard Gigabit Ethernet pass-through modules.

Gigabit Ethernet switch link aggregation

The link aggregation feature in the PowerConnect 5316M integrated Gigabit Ethernet switch helps provide the Dell PowerEdge 1855 blade server with scalable bandwidth and high availability. Link aggregation offers the capability to combine multiple physical Gigabit Ethernet ports on the PowerConnect 5316M integrated switch into a single logical port that can deliver up to the combined bandwidth of the individual ports and help provide high...
availability through port failover protection. For example, the six external Gigabit Ethernet ports of the PowerConnect 5316M integrated switch can be combined to provide a single high-bandwidth logical channel. Link aggregation enables high availability through failover and failback features. If the connection is lost on one of the aggregated ports, traffic can be automatically forwarded to the other ports. Once the connection is reestablished, traffic will automatically fail back.

Note: In the following examples, LOM 1 for each server blade is represented by an even-numbered Ethernet address (the least significant bit of the address is a zero) while LOM 2 is represented by an odd-numbered Ethernet address (the least significant bit of the address is a one). The examples are constructed in this manner because the Ethernet address for LOM 2 is always one number higher than the address for LOM 1. Ethernet addresses are also commonly referred to as Media Access Control (MAC) addresses. Each PowerEdge 1855 blade server is configured with one even MAC address and one odd MAC address.

Link aggregation can be configured as either dynamic or static. Dynamic configuration is supported using the IEEE 802.3ad standard, which is known as Link Aggregate Control Protocol (LACP). Static configuration is used when connecting the PowerConnect 5316M integrated Gigabit Ethernet switch to an external Gigabit Ethernet switch that does not support LACP. The advantage of LACP is that this protocol enables the integrated Gigabit Ethernet switch to confirm that the external switch is also configured for link aggregation. When using static configuration, a cabling mistake or configuration mistake involving the external switch could go undetected and lead to undesirable network behavior. Best practices suggest using dynamic link aggregation instead of static link aggregation.

### Configuring external ports for dynamic link aggregation

The following example shows the Dell PowerConnect 5316M integrated Gigabit Ethernet switch command-line interface (CLI) commands for configuring the six external ports on the Gigabit Ethernet switch for LACP. The first command sets the CLI mode to configure the six external Gigabit Ethernet ports (referred to in the command as `g11-16`, which stands for Gigabit Ethernet ports 11 through 16). The second command aggregates the six ports into a link-aggregation group (referred to in the command as `channel-group`), which will use LACP (referred to in the command as `auto`).

```
console(config)# interface range ethernet g11-16
console(config-if)# channel-group 1 mode auto
```

### Configuring external ports for static link aggregation

The following example shows the PowerConnect 5316M integrated Gigabit Ethernet switch CLI commands for configuring three external ports of the Gigabit Ethernet switch for static aggregation. The first command sets the CLI mode to configure three external Gigabit Ethernet ports (g13-15). The second command aggregates the three ports into a static link-aggregation group. Static link-aggregation groups do not use LACP and are defined in the CLI by setting the channel-group mode to “on.”

```
console(config)# interface range ethernet g13-15
console(config-if)# channel-group 4 mode on
```

### Optimal link-aggregation traffic distribution

The link-aggregation technique enables traffic to be distributed across the aggregated ports based on the last six bits of the source MAC address and the last six bits of the destination MAC address of each Ethernet packet. This calculation is derived using a static mathematical algorithm. In an optimal environment—one that has a large number of random Ethernet source or destination addresses—the algorithm is designed to provide a fairly equal distribution across each port in the aggregated link.

![Optimal aggregated link traffic distribution of 64 streams](image)

The following table shows the Gigabit Ethernet switch’s port distribution of 64 destination addresses—each ending in a unique six bits—from a single source address for each of the possible aggregated ports. (Six bits can represent up to 64 different values.) Each source and destination pair is referred to as an outbound stream from the switch’s point of reference. This example represents an ideal environment because it shows 64 streams whose addresses...
are equally distributed across the 64 possible values. This distribution helps make efficient use of a set of external ports configured in a link-aggregation group—particularly when all 10 server blades are communicating to large numbers of Ethernet addresses. In contrast, when all 10 server blades are communicating with a very small number of Ethernet addresses, the switch may not use all the aggregated ports.

The example in Figure 2 shows that, for two aggregated links, traffic would be distributed with 32 streams on port 11 and 32 streams on port 12. For three aggregated links, the traffic would be distributed with 22 streams on port 11, 21 streams on port 12, and 21 streams on port 13. Note: The 10 internal switch ports are numbered 1 through 10, while the six external switch ports are numbered 11 through 16. Link aggregation is configured on the six external ports, 11 through 16; internal switch ports cannot be configured for link aggregation.

Calculating link-aggregation traffic distribution

The IEEE 802.3ad standard defines the protocol used to communicate link aggregation information between switches, but it does not specify the algorithm to be used to assign traffic flows to the individual links within the link-aggregation group. The PowerEdge 1855 blade server’s PowerConnect 5316M integrated Gigabit Ethernet switch algorithm is presented in this section. Other Gigabit Ethernet switch vendors and other Dell PowerConnect switches may use a different calculation based on other bits in the source or destination addresses.

The calculation for determining the aggregated port to which an Ethernet stream is assigned on the PowerConnect 5316M integrated Gigabit Ethernet switch is based on three items:

- The six least significant bits (LSBs) of the source MAC address (SA)
- The six LSBs of the destination MAC address (DA)
- The number of aggregated ports

The link-aggregation formula for the PowerConnect 5316M integrated Gigabit Ethernet switch is as follows:

\[(6 \text{ LSBs } \text{SA}) \text{ XOR } (6 \text{ LSBs } \text{DA}) \text{ MODULUS (number aggregated ports)} + 1 = \text{relative port number}\]

In the equation, XOR is the exclusive OR function. The XOR function is the bitwise comparison of two numbers: If two bits in the same bit position in the number are not equal, then the result will have a 1 in that bit position. If two bits in the same bit position are the same, then the result will have a 0 in that bit position. The following are some examples of the XOR operator for a single bit:

- 0 XOR 0 = 0
- 0 XOR 1 = 1
- 1 XOR 1 = 0
- 1 XOR 0 = 1

The following are examples for six bits:

- 01 0010 XOR 11 1111 = 10 1010
- 00 1111 XOR 00 0000 = 00 1111
- 11 0011 XOR 00 1100 = 11 1111

In the preceding equation, the MODULUS operator takes the remainder. For example:

- 6 MODULUS 4 = 2
- 2 MODULUS 2 = 0
- 7 MODULUS 3 = 1

Figure 3 shows a specific example of the PowerConnect 5316M integrated Gigabit Ethernet switch link-aggregation port-distribution calculation.

In a scenario in which the connection is lost on one or more of the aggregated ports, traffic can be forwarded based on the number of connected ports. Thus, if the switch is configured for five aggregated ports but one of the links is disconnected, traffic can be distributed as if the switch were configured with four aggregated ports.

The sample calculation shown in Figure 3 applies only to outbound, or transmitted, traffic from the PowerConnect 5316M integrated Gigabit Ethernet switch. The distribution of received traffic is controlled by the external switch to which the PowerConnect 5316M integrated Gigabit Ethernet switch is connected. If the external switch is a Dell PowerConnect 5324 switch or PowerConnect 6024 router.
the same equation applies. If the external switch is a different Dell PowerConnect switch or another vendor’s switch, the received distribution may be based on a different calculation.

Comparing link aggregation to load balancing
For Dell and other vendors of Layer 2 Gigabit Ethernet switches, the calculation for aggregated port traffic distribution is based on static features of the MAC addresses of the traffic. This means the PowerConnect 5316M Gigabit Ethernet switch does not provide load balancing—nor do standards-based Gigabit Ethernet switches from other vendors. That is, as traffic increases beyond a single port’s bandwidth, the traffic does not spill over to the next aggregated port. This type of load balancing is not possible for link-aggregation groups based purely on Layer 2 information because of the static nature of the calculation, and the requirement to deliver Layer 3 packets in order. Thus, traffic for the same pair of source and destination addresses always travels on the same link, even if the link is oversubscribed.

Common PowerEdge 1855 blade server network topologies and link aggregation
The Dell PowerEdge 1855 blade server fits into many network environments while requiring few or no manual configuration settings. The PowerConnect 5316M integrated Gigabit Ethernet switch can communicate with any external Gigabit Ethernet switches (such as Dell PowerConnect switches), hubs, routers, or network interface cards (NICs) that support IEEE 802.3, 802.3u, or 802.3ad standards. Using the link-aggregation feature of the PowerConnect 5316M integrated Gigabit Ethernet switch can help administrators take advantage of enhanced bandwidth and link failover. The most common network environments and considerations when using the link-aggregation feature are discussed in the following sections.

Ten server blades in a general network
As shown in Figure 1, a common network environment for the PowerEdge 1855 blade server includes Gigabit Ethernet switches on two separate networks, with all server blades transmitting and receiving from a large number of random destination addresses on each network. In Figure 1, the public and private networks represent a large number of Ethernet addresses that have an equal distribution of the last six bits of their addresses. This environment can result in optimal traffic distribution across the aggregated ports as described in the “Optimal link-aggregation traffic distribution” section in this article. Each traffic stream has its own profile, so the traffic distribution across the link-aggregation group will vary with time. On average, however, the distribution will be fairly uniform.

Ten server blades with a single default router
Figure 4 shows another common network environment for the Dell PowerEdge 1855 blade server. In this example, all server blades have a single default path to a router, such as the Dell PowerConnect 6024, or to multiple routers running Virtual Router Redundancy Protocol (VRRP) to function as a single virtual router. To simplify this scenario, this example involves only a connection to LOM 1. This type of environment requires administrators to consider how traffic is distributed on the aggregated ports.

In this scenario, all SAs are even (because the traffic originates from LOM 1 of the server blades) and there is a single DA—that of the router or virtual router. Figure 5 shows the results of the port distribution calculation for the Dell PowerEdge 1855 blade server Ethernet switch for different numbers of ports used in a link-aggregation group. Note that in this network environment, traffic is generally not uniformly distributed across all links in the link-aggregation group because the SAs are from such a small sample (that is, the 10 server blades).

In this type of single-destination Ethernet address environment, best practices recommend using an odd number of aggregated ports because traffic could be distributed across all links in the link-aggregation group. Selecting an even number of ports for the link-aggregation group is not optimal because only half of the ports will carry traffic in this environment. Such a situation arises because of the nonrandom nature

<table>
<thead>
<tr>
<th>Number of aggregated ports</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>2</td>
<td>2</td>
<td></td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td></td>
<td>3</td>
<td>0</td>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>

Figure 5. Best-case analysis of a single-destination Ethernet address

Figure 4. Link aggregation using a single default route
of the SAs—all even or all odd, depending on whether LOM 1 or LOM 2 is used.

How to check and modify traffic distribution
The overview provided in this article can help administrators examine their network environment and server blade MAC addresses to calculate how traffic would be distributed when using the PowerConnect 5316M integrated Gigabit Ethernet switch. The use of a few CLI commands allows an administrator to investigate traffic distribution. These commands include:

  console# clear counters
  console# show interfaces counters

It is important to note that the switch can control the distribution of transmitted—but not received—traffic from its ports. Thus, administrators should examine only the output statistics when determining whether the system is providing sufficient link-aggregation traffic distribution. Also, system administrators should ensure that they have a good understanding of the network traffic patterns and a thorough sampling of data before changing network traffic. Then traffic distribution either can be calculated using the algorithm discussed in “Calculating link-aggregation traffic distribution” section in this article or determined by examining the counters. If poor distribution is evident, system administrators can opt to change the number of aggregated ports or move server blades among different PowerEdge 1855 blade server chassis to change the source MAC address mix and therefore the distribution.

Enhanced network capabilities with the PowerEdge 1855 blade server
As explained in this article, properly configuring the PowerConnect 5316M integrated Gigabit Ethernet switch can help administrators deploy the PowerEdge 1855 blade server to best advantage in various networking scenarios. By understanding different deployment options and determining the most effective one for a specific environment, administrators can enhance networking performance and availability of a PowerEdge 1855 blade server equipped optional PowerConnect 5316M integrated Gigabit Ethernet switches.

Bruce Holmes is a senior test engineer in the Dell PowerConnect Group. He has worked at Dell for one year and supports Dell PowerConnect switches in all phases of product development and testing. Prior to Dell, Bruce was a networking software engineer at Wind River Systems, Alcatel, and Texas Instruments. He has a B.S. in Electrical Engineering from The University of Texas at Austin.

FOR MORE INFORMATION
Dell blade servers:
dell.com/blades