DELL EMC 25 GIGABIT ETHERNET AUTO-NEGOTIATION TECH NOTE

S5048F-ON and S5148F-ON

ABSTRACT

This paper describes some of the scenarios that may be experienced in the field while connecting 25GbE switches to 25GbE Servers. The primary focus is to highlight the settings on the switch depending on which cable or optic are installed, and some configuration examples with 14G servers.

December, 2017

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This document will provide a brief introduction to some issues potentially encountered when connecting different 25GbE devices together, due to the manner in which 25 Gigabit Ethernet standards evolved.

It will cover the switch default settings for auto-negotiation and FEC (Forward Error Correction) depending on the inserted media and provide configuration details. It will also list the various NICs that were tested, as well as a detailed test report for one NIC, with server and device settings.

AUDIENCE

This Tech Note is intended for Pre- and Post-Sales Systems Engineers, Support Engineers, Consultants, and Network Administrators engaged in the acquisition, set up, and troubleshooting of 25GbE switches and attached devices (servers, storage, etc.).

INTRODUCTION

To keep pace with the rapid evolution of technology in the data center, the networking industry continues to deliver products with higher and higher bandwidth capabilities at more cost effective price points. With the finalization of the IEEE 802.3bj standard and 100Gbps over 4x25G PHY introduction, it opened a path to 25GbE over a single lane. The 25 Gigabit Ethernet Consortium formed with the intent to quickly drive a standard for 25 Gigabit Ethernet over a single lane, based off of the IEEE 802.3bj standard. Shortly after the 25 Gigabit Ethernet Consortium formed, the IEEE 802.3 Task Force for single lane 25Gbps Ethernet was also formed.

These two bodies continued developing their version of a standard in parallel. Since these standards both leveraged the IEEE 802.3bj standard, they were similar, but there were also some slight differences. Depending on when a given 25 Gigabit product was introduced to the market, it may have been pre-standard (both Consortium and IEEE), it may have been compliant with the 25 Gigabit Ethernet Consortium standard, or it may be compliant with IEEE 802.3by (25 Gigabit Ethernet) standard, or a combination of each.

Because each device that is being connected together may have come out at a different point in time with respect to the 25 Gigabit "standards," each may have slightly different capabilities and/or default settings. These different capabilities and/or settings may cause the auto-negotiation capabilities in the devices to not function as seamlessly as would normally be expected. This paper strives to shed some light on these differences, their associated settings, and some examples of what would need to be configured or changed to achieve a successful link status.

AUTO-NEGOTIATION & FORWARD ERROR CORRECTION (FEC)

Auto-negotiation is a physical layer technique used in Ethernet networks to allow two connecting devices to negotiate the appropriate transmission parameters and settle on the highest, mutually supported set of characteristics. The concept was introduced as part of the Fast Ethernet standard, and has progressed as each new Ethernet technology came to market, incorporating new technology abilities as required.

FEC was introduced into the Ethernet standards in order to continue to drive higher transmission rates, while maintaining acceptable Bit Error Rate (BER). FEC encoding protects IP packets against errors caused by noise and other impairments, and is able to detect and correct those errors. However, the more capable a particular FEC encoding scheme is at detecting and fixing errors, the more latency it adds into the network. So the lightest-weight variant of FEC (or none) that ensures the acceptable BER is what is required by the standard.

The IEEE 802.3by standard modified, among other things, the ability for auto-negotiation to request FEC (BASE-R FEC (Firecode), RS-FEC (Reed Solomon), or both. If neither PHY devices requests FEC, then FEC is not enabled. If either PHY device requests RS-FEC, then RS-FEC is enabled. If either PHY devices requests BASE-R FEC only, then BASE-R FEC is enabled. For 25GBASE-CR-S PHY device, which does not support RS-FEC, BASE-R FEC is enabled if either PHY device requests RS-FEC or BASE-R FEC. In IEEE 802.3by, Clause 108 (CL-108 or CL108) specifies Reed Solomon or RS-FEC, whereas Clause 74 (CL-74 or CL74) refers to Firecode or BASE-R FEC.

SWITCH DEFAULT SETTINGS

The table below lists the default auto-negotiation and FEC settings on the S5048F-ON (OS9.12.1) and S5148F-ON (OS10.3.2) switches when the listed cable or optic is installed.

Optic/Cable	SKU	Dell P/N	AutoNeg (Default)	FEC (Default)	FEC Clause (CL) type Enabled
SFP28 25G DAC 1m	470-ACES	2JVDD	Enabled	Not requested	
SFP28 25G DAC 2m	470-ACET	DOR73	Enabled	Not requested	
SFP28 25G DAC 3m	470-ACEV	VXFJY	Enabled	Not requested	
SFP28 25G DAC 5m	470-ACEW	9X8JP	Enabled	Requested	CL108
SFP28 25G AOC 7m	470-ACIM	3YWG7	Disabled	Enabled	CL108
SFP28 25G AOC 10m	470-ACIN	5CMT2	Disabled	Enabled	CL108
SFP28 25G AOC 15m	470-ACIL	RCVP5	Disabled	Enabled	CL108
SFP28 25G AOC 20m	470-ACIG	X5DH4	Disabled	Enabled	CL108
100G to 4*25G DAC 1m	470-ABOR	26FN3	Enabled	Not requested	
100G to 4*25G DAC 2m	470-ABOS	YFNDD	Enabled	Not requested	
100G to 4*25G DAC 3m	470-ABOT	7R9N9	Enabled	Requested	CL108
SFP28 SR	407-BBWO	P7D7R	Disabled	Enabled	CL108
SFP28 SR-NOF	407-BBXX	W4GPP	Disabled	Enabled	CL108
SFP28 LR	407-BBXY	0YR96	Disabled	Disabled	

SWITCH CONFIGURATION

Auto-negotiation gets enabled/disabled by default in the switch according to the media type interface inserted in the SFP28 port (refer to above table).

In the S5048F-ON and S5148F-ON, IEEE 802.3by is the default auto-negotiation standard. It can settle down to IEEE 802.3by standard or the 25 Gigabit Ethernet Consortium standard according to peer end auto-negotiation configurations. The following are the CLI commands required to enable/disable auto-negotiation both for OS9 and OS10.

CLI to enable Auto-Neg in 25G interface:

‘intf-type cr1 autoneg’ (OS9) / ‘negotiation on’ (OS10)

CLI to disable Auto-Neg in 25G interface:

‘no intf-type cr1 autoneg’ (OS9) / ‘negotiation off’ (OS10)

FEC type gets enabled/disabled (requested or not requested in the case when auto-negotiation is enabled) by default according to media types. When auto-negotiation is enabled and FEC type is requested by default:

FEC Clause 74 gets enabled when auto-negotiation settles down to the 25 Gigabit Ethernet Consortium standard

FEC Clause 108 gets enabled when auto-negotiation settles down to the IEEE 802.3by standard

The following are CLI commands to enable/disable FEC.

CLI to enable FEC in 25G interface:

'fec enable CL74' (OS9) / 'fec CL74-FC' (OS10)

'fec enable CL108' (OS9) / 'fec CL108-RS' (OS10)

CLI to disable FEC in 25G interface:

'no fec enable' (OS9) / 'fec off' (OS10)

The following table lists the FEC classifications and settings in the switch and server NIC

In Switch	In Server NIC	Dell Part Number
CL-74	CL-74 in Switch corresponds to BaseR-FEC or Firecode FEC	Supported by Ethernet Consortium standard
CL-108	CL-108 in Switch corresponds to RS-FEC or Reed Solomon FEC	Supported by IEEE 802.3by standard*

*Note: Some NIC configuration settings may show CL91, which is IEEE 802.3bj Reed Solomon FEC (RS-FEC). CL-108 should be the switch-side FEC setting to match this.

LIST OF ADAPTERS TESTED

The table below lists the various 25GbE network adapters that were tested with the S5048F-ON and S5148F-ON prior to RTS. Each of these were tested in a PowerEdge® R940 server with the various cables and optics listed in the table in the "DEFAULT SWITCH SETTINGS" section. The table below lists the vendor, the form factor (PCIe or rNDC), the Dell Part Number, and the Firmware running on the NIC during test.

Vendor	Model	Form Factor	Speed/Ports	Dell Part Number	Dell Server	NIC Firmware Version
Broadcom	57414	PCIe	2 x 25GbE SFP28	CX94X/24GFD	R940	20.06.05.08
Broadcom	57414	rNDC	2 x 25GbE SFP28	9XY73	R940	20.06.05.10
Mellanox	ConnectX-4 LX	rNDC	2 x 25GbE SFP28	R887V	R940	14.14.60
Mellanox	ConnectX-4 LX	PCIe	2 x 25GbE SFP28	20NJD/MRT0D	R940	14.14.60
QLogic	QL41262	PCIe	2 x 25GbE SFP28	V5TMM/R491V	R940	14.04.00

SAMPLE RESULTS (SWITCH-TO-SERVER CONNECTION)

The table below shows the steps required to get a Broadcom 57414 adapter in an R940 server to successfully establish link with a default configured S5048F-ON. The comments field denotes the required action (where there is no comment, no additional action was required for link to come up).

Broadcom NIC: 57414

Firmware Version: 20.06.05.10

Auto-Negotiation Standard: Enabled

FEC Clause Enabled: CL108

OS Tested: Windows Server 2012 R2 Standard

Optic/Cable	AutoNeg (Default)	FEC (Default)	S5048F Link Status	S5048F FEC Status	Comments
SFP28 25G DAC 1m, 2m, 3m	Enabled	Not requested	Up	None	
SFP28 25G DAC 5m	Enabled	Requested	Up	CL108	
SFP28 25G AOC 7m, 10m, 15m, 20m	Disabled	Enabled	Down	None	Link is down because of Auto-Neg mismatch. Disable Auto-Neg in the server to bring up link with FEC enabled
100G to 4*25G DAC 1m, 2m	Enabled	Not requested	Up	None	
100G to 4*25G DAC 3m	Enabled	Requested	Up	CL108	
SFP28 SR/SR-NOF	Disabled	Enabled	Down	None	Link is down because of Auto-Neg mismatch. Disable Auto-Neg in the server to bring up link with FEC enabled
SFP28 LR	Disabled	Disabled	Down	None	Link is down because of Auto-Neg mismatch. Disable Auto-Neg and disable FEC in the server to bring up the link.

The screenshot below captures the relevant settings in the Broadcom 57414 NIC in an PowerEdge R940 server