Open Source Tooling for PHY Management and Testing

Abstract

One of the techniques Ethernet uses to achieve a high rate of speed is pulse-amplitude modulation. This allows increasing the density of the data being sent by a signal by defining equalization ranges for different combinations of bits. These ranges differ based on speed and are configured in firmware before the link is up.

Ranges are typically defined in firmware and seldom changed. However during hardware interoperability testing they must be evaluated, tested, and potentially modified. Testing requires manually starting and stopping pseudo random binary sequence(PRBS) modes and other indefinite pattern tests. Various vendors provide their own proprietary tools to set equalization values and run pattern tests. No open source tool exists to provide this functionality.

At the Netdev 0x18 - Driver and H/W APIs Workshop the community discussed various methods to implement new common features. The community acknowledged that the kernel must never break user space and only accept new features once they have matured enough to become a standard interface. The purpose of this talk is to discuss whether setting equalization values and enabling pattern tests has matured enough across various vendors to develop a standard open source interface.

Current State

Looking across three vendors we can see that setting equalization values and running pattern tests are achieved using very similar, yet different, commands.

Mellanox

Current Equalization Values

```
[root@localhost ~]# mstlink -d 19:00.0 --show_serdes_tx --advance

Operational Info
----------------
State                           : Active
Physical state                  : ETH_AN_FSM_ENABLE
```

```
Speed                              : 400G
Width                              : 8x
FEC                                : Standard_RS-FEC - (544,514)
Loopback Mode                      : No Loopback
Auto Negotiation                   : ON

Supported Info
--------------
Enabled Link Speed (Ext.)          : 0x000097f2
(400G_8X,200G_4X,100G_2X,100G_4X,50G_1X,50G_2X,40G,25G,10G,1G)
Supported Cable Speed (Ext.)       : 0x000097fe
(400G_8X,200G_4X,100G_2X,100G_4X,50G_1X,50G_2X,40G,25G,10G,5G,2.5G,1G)

Troubleshooting Info
--------------------
Status Opcode                      : 0
Group Opcode                       : N/A
Recommendation                     : No issue was observed.

Tool Information
----------------
Firmware Version                   : 28.33.2018
amBER Version                      : 2.00
MSTFLINT Version                   : mstflint 4.20.1-1.fb1

Serdes Tuning Transmitter Info
------------------------------
Serdes TX parameters               : fir_pre2,fir_pre1,fir_main,fir_post1
Lane 0                             : 2,-8,53,0
Lane 1                             : 2,-8,53,0
Lane 2                             : 2,-8,53,0
Lane 3                             : 2,-8,53,0
Lane 4                             : 2,-8,53,0
Lane 5                             : 2,-8,53,0
Lane 6                             : 2,-8,53,0
Lane 7                             : 2,-8,53,0
```

Setting Equalization Values

```
[root@localhost ~]# mstlink -d 19:00.0 --serdes_tx 2,-9,49,-2 --advance
```

```
Operational Info
----------------
State                              : Active
Physical state                     : ETH_AN_FSM_ENABLE
Speed                              : 400G
Width                              : 8x
FEC                                : Standard_RS-FEC - (544,514)
Loopback Mode                      : No Loopback
Auto Negotiation                   : ON

Supported Info
--------------
Enabled Link Speed (Ext.)          : 0x000097f2
(400G_8X,200G_4X,100G_2X,100G_4X,50G_1X,50G_2X,40G,25G,10G,1G)
Supported Cable Speed (Ext.)       : 0x000097fe
(400G_8X,200G_4X,100G_2X,100G_4X,50G_1X,50G_2X,40G,25G,10G,5G,2.5G,1G)

Troubleshooting Info
--------------------
Status Opcode                      : 0
Group Opcode                       : N/A
Recommendation                     : No issue was observed.

Tool Information
----------------
Firmware Version                   : 28.33.2018
amBER Version                      : 2.00
MSTFLINT Version                   : mstflint 4.20.1-1.fb1

Configuring Port Transmitter Parameters...
```

Starting and Stopping Pattern Testing

```
[root@localhost ~]# mlxlink -d /dev/mst/mt53104_pciconf0 --port 3 --cable
--prbs_select HOST --prbs_mode EN --checker_pattern PRBS13 --invert_checker
--generator_pattern PRBS31 --swap_generator --lane_rate HDR

[root@localhost ~]# mlxlink -d /dev/mst/mt53104_pciconf0 --port 3 --cable
```

```
--prbs_select HOST --prbs_mode DS
```

Broadcom

Current Equalization Values

```
[root@localhost ~] niccli -pci 0000:03:00.0 txfir -get -modtype PAM4 -lane
0

-------------------------------------------------------------------------
----
NIC CLI v232.0.153.0 - Broadcom Inc. (c) 2024 (Bld-106.52.39.138.16.0)
-------------------------------------------------------------------------
----

Transmit Finite Impulse Response (TxFIR):
     Pre1   : -20
     Pre2   : 0
     Main   : 132
     Post1  : -16
     Post2  : 0
     Post3  : 0
     Amp    : 0
```

Setting Equalization Values

```
[root@localhost ~] niccli -pci 0000:03:00.0 txfir -set -modtype PAM4 -lane
1 -pre1 1 -pre2 -2 -main 12 -amp 10 -post1 -10 -post2 15 -post3 10
```

Starting and Stopping Pattern Testing

```
[root@localhost ~] niccli -pci 0000:03:00.0 prbs_test -enable -mode PRBS31
-rxlanemask 255 -txlanemask 255 -duration 10
[root@localhost ~] niccli -pci 0000:03:00.0 prbs_test -enable -disable
```

fbnic

Note: The following is our current proof of concept which has yet to be upstreamed. We are using DebugFS due to the lack of tooling being discussed here.

Current Equalization Values

```
[root@localhost ~]# cd /sys/kernel/debug/fbnic/0000\:01\:00.0/fbn/tx_fir/
[root@localhost tx_fir]# for i in */*; do echo "$i - $(cat $i)"; done
global/main_tap - 34
global/post_tap - 0
global/pre_tap_1 - 3
global/pre_tap_2 - 0
global/pre_tap_3 - 0
lane0/main_tap - 34
lane0/post_tap - 0
lane0/pre_tap_1 - 3
lane0/pre_tap_2 - 0
lane0/pre_tap_3 - 0
lane1/main_tap - 34
lane1/post_tap - 0
lane1/pre_tap_1 - 3
lane1/pre_tap_2 - 0
lane1/pre_tap_3 - 0
```

Setting Equalization Values

```
[root@localhost tx_fir]# echo 35 > global/main_tap
[root@localhost tx_fir]# echo 4 > lane0/pre_tap_1
[root@localhost tx_fir]# echo 2 > lane1/pre_tap_1
[root@localhost tx_fir]# echo set > ctrl
```

Starting and Stopping Pattern Testing

```
[root@localhost testing]# cd /sys/kernel/debug/fbnic/0000\:01\:00.0/testing
[root@localhost testing]# ls -alh
total 0
drwxr-xr-x  2 root root 0 Aug 22 07:28 .
drwxr-xr-x 50 root root 0 Aug 22 07:28 ..
-rw-------  1 root root 0 Aug 22 07:28 rx
-r--------  1 root root 0 Aug 22 07:28 supported_tests
-rw-------  1 root root 0 Aug 22 07:29 tx
-rw-------  1 root root 0 Aug 22 07:28 tx_prbs_8_10
```

```
[root@localhost testing]# cat supported_tests
Disabled
PRBS7
PRBS9
PRBS11
PRBS11.0
PRBS11.1
PRBS11.2
PRBS11.3
PRBS15
PRBS16
PRBS23
PRBS31
PRBS32
PRBS13.0
PRBS13.1
PRBS13.2
PRBS13.3
[root@localhost testing]# cat tx_prbs_8_10
N
[root@localhost testing]# echo "prbs23" > rx
[root@localhost testing]# cat rx
PRBS23
[root@localhost testing]# cat tx
Disabled
[root@BW16281 testing]# echo 0 > rx
[root@localhost testing]# echo y > tx_prbs_8_10
[root@localhost testing]# cat tx_prbs_8_10
Y
[root@localhost testing]# echo "prbs13" > tx
[root@localhost testing]# cat tx
PRBS13
[root@localhost testing]# cat rx
Disabled
```

Proposal

ethtool is the standard tool for controlling network drivers and hardware. This proposal suggests extending ethtool to provide a standard interface for developers and users to get and set equalization values as well as starting and stopping pattern testing.

Current Equalization Values

```
[root@localhost ~] ethtool --get-equalization eth0
                pre_tap_3, pre_tap2, pre_tap1, main_tap, post_tap
lane0              0,         0,        3,         34,       0
lane1              0,         0,        3,         34,       0
[root@localhost ~] ethtool --json --get-equalization eth0
[ {
    “lane0”: {
                 “pre_tap_3”: 0,
                 “pre_tap_2”: 0,
                 “pre_tap_3”: 3,
                 “main_tap”: 34,
                 “post_tap”: 0
             },
    “lane1”: {
                 “pre_tap_3”: 0,
                 “pre_tap_2”: 0,
                 “pre_tap_3”: 3,
                 “main_tap”: 34,
                 “post_tap”: 0
             }
    } ]
```

Settings Equalization Values

Default sets all lanes

```
[root@localhost ~] ethtool --set-equalization eth0 0, 0, 4, 35, 0
```

Setting equalization for a specific lane

```
[root@localhost ~] ethtool --set-equalization eth0 --lane 1 0, 0, 4, 35, 0
```

Starting and Stopping Pattern Testing

Show supported tests:

```
[root@localhost ~] ethtool --get-supported-pattern-tests eth0
Disabled
PRBS7
PRBS9
PRBS11
PRBS11.0
```

```
PRBS11.1
PRBS11.2
PRBS11.3
PRBS15
PRBS16
PRBS23
PRBS31
PRBS32
PRBS13.0
PRBS13.1
PRBS13.2
PRBS13.3
[root@localhost ~] ethtool --json --get-supported-pattern-tests eth0
[
    “Disabled”,
    “PRBS7”,
    “PRBS9”,
    “PRBS11”,
    “PRBS11.0”,
    “PRBS11.1”,
    “PRBS11.2”,
    “PRBS11.3”,
    “PRBS15”,
    “PRBS16”,
    “PRBS23”,
    “PRBS31”,
    “PRBS32”,
    “PRBS13.0”,
    “PRBS13.1”,
    “PRBS13.2”,
    “PRBS13.3”
]
```

Start/stop testing:

```
[root@localhost ~] ethtool --set-pattern-test tx prbs13 tx-prbs-8-10 eth0
```

Get test status

```
[root@localhost ~] ethtool --get-pattern-test eth0
TX           PRBS13.0
TX PRBS 8-10 Enabled
RX           Disabled
```

```
[root@localhost ~] ethtool --json --get-pattern-test eth0
[ {
    “TX”: “PRBS13.0”,
    “TX PRBS 8-10”: true,
    “RX”: “Disabled”
  }]
```

References

AN 835: PAM4 Signaling Fundamentals, Intel, www.intel.com/content/www/us/en/docs/programmable/683852/current/nrz-fundamentals.html?fbclid=IwZXh0bgNhZW0CMTEAAR1vH3TWqX6SxZh4CbapLLf0kLqU7Yfiwe6lbgL7gGLBcK3h2b0jXfy_Geo_aem_f5lhJGotMDhX2rFHms-OYw. Accessed 29 Jan. 2025.

“Mstlink .” *NVIDIA Firmware Tools (MFT) Documentation v4.26.1 LTS*, 4.20.1-1.fb1, Mellanox, https://docs.nvidia.com/networking/display/mftv4261lts/mstlink+utility. Accessed 29 Jan. 2025.

“Niccli.” *Support Documents and Downloads*, 232.0.153.0, Boardcom, 25 Jan. 2025, https://www.broadcom.com/support/download-search?legacy=1&pf=Ethernet+Controllers&pg=Ethernet+Connectivity%2C+Switching%2C+and+PHYs.

Author Biography

Trager, Lee is an experienced open source developer and enthusiast. Contributions range from the Linux kernel to user space tools used in Ubuntu, and Fedora based operating systems. Currently Lee works at Meta on the driver and firmware for fbnic, Meta’s datacenter NIC.