Consistent Network Device Naming in Linux

Narendra K (Narendra_K@Dell.com)
Linux Engineering

Document Version 1.0
25th Jan, 2012

Table of Contents

1. Introduction

Modern x86 based Servers support an increasing number of Network interface ports on the mother board in addition to add-in network adapters. Linux based OSes name these interfaces as ethN. The naming of network interfaces is currently non-deterministic and not governed by any standard in terms of their relationship to the way the ports are wired on the system. Common user expectation like 'eth0' representing the first network port on the mother board as labeled on the server chassis cannot be fulfilled in many cases.

Ensuring that the Ethernet interface names follow the order of the devices as intended by the system designer might not be sufficient. The "ethN" names currently in use do not suggest the Ethernet interface's physical location, whether it is on the systems' mother board or if it is on an add-in card; and if it is on an add-in card with multiple ports, which port on the card is it on.

A naming mechanism that can impart meaning to the network interface's name based on the physical location of a network port in concordance to the intended system design is necessary. To achieve that, the system firmware has the ability to communicate the intended order for network devices on the mother board to the Operating System via standard mechanisms such as SMBIOS and ACPI.

The new naming scheme uses 'biosdevname' udev helper utility , developed by Dell and released under GPL, suggests new names based on the location of the network adapters on the system as suggested by system BIOS.

2. The new naming scheme

Lan-On-Motherboard interfaces
>> **em\<port number>_\< virtual function instance / NPAR Index>**
>>> (ethernet-on-motherboard <1,2 ..>)

PCI add-in interfaces
>> **p\<slot number>p\<port number>_\<virtual function instance / NPAR Index>**

2.1. Firmware information that "biosdevname" utilizes to suggest new names

The system BIOS indicates the order of Onboard Network Devices to the OS via SMBIOS type 41 records. The system BIOS provides 'system slot information' to the OS via type 9 records. The biosdevname makes use of SMBIOS type 41 to suggest names to the onboard network devices and type 9 records to suggest new names for PCI add-in adapters.

The biosdevname falls back on PCI IRQ Routing Table in the absence of the above mentioned information.

2.2.   Dell Systems that support the new network naming

Table 1: Dell Systems that support the new network naming

System Model	RHEL 6.1	Fedora 15 Beta
Dell PowerEdge		
M610	X	X
M610X	X	X
M710	X	X
M710HD	X	X
M910	X	X
R210	X	X
R310	X	X
R410	X	X
R415	X	X
R510	X	X
R515	X	X
R610	X	X
R710	X	X
R715	X	X
R810	X	X
R815	X	X
R910	X	X
T110	X	X
T310	X	X
T410	X	X
T610	X	X
T710	X	X
M600	-	X
M605	-	X
M805	-	X
M905	-	X
R200	-	X
R300	-	X
R805	-	X
R900	-	X
R905	-	X
T100	-	X
T105	-	X
T300	-	X
T605	-	X
1955	-	X
1950	-	X
2950	-	X
2970	-	X
6950	-	X
1900	-	X
2900	-	X
Dell Precision		
R5400	-	X
T3500	-	X
T5500	-	X
T7500	-	X
T1600	X	X
R5500	-	X
M4600	X	X
M6600	X	X
T7600	-	X
T5600	-	X

T3600	-	X
PowerEdge C Series		
C1100	-	X
C2100	-	X
C6100	-	X
C6105	-	X
C6145	-	X

Note:

1. Systems with '-' under RHEL 6.1 have their system BIOS implementing SMBIOS version 2.5 and 2.4, hence biosdevname will not suggest new names in RHEL 6.1

3. RHEL 6.1 NoteS

In RHEL 6.1, system BIOS should implement SMBIOS Version 2.6 Type 41 and Type 9 records for the biosdevname to suggest new names. biosdevname will not fall back to PCI IRQ Routing Table and new names will not be suggested in the absence of SMBIOS Version 2.6 compliant system BIOS.

4. Getting the new naming scheme to work on your system

4.1. Prerequisites

The system BIOS should implement SMBIOS Version 2.6 type 41 and type 9 tables. If they are not available, the system BIOS should expose a PCI IRQ Routing Table. This can be checked with these commands:

To check if BIOS implements SMBIOS type 41

> ***# /sbin/dmidecode –t 41***

To check if BIOS implements SMBIOS type 9

> ***# /sbin/dmidecode –t 9***

To check if BIOS implements PCI IRQ Routing Table

> ***# /sbin/biosdecode***

Note: RHEL 6.1 requires SMBIOS version 2.6 to use the new naming scheme.

4.2. Building and installing biosdevname from source tar ball

biosdevname can be built from sources. These sources are available for download here:

http://linux.dell.com/files/biosdevname/permalink/

git://linux.dell.com/biosdevname.git

   a) Prerequisites: pciutils-devel and zlib-devel
   b) Build and install biosdevname

```
$ tar zxvf biosdevname-<latest version>.tar.gz
$ cd biosdevname-<latest version>
$ ./configure
$ sudo make && make install
```

c) Remove /etc/udev/rules.d/70-persistent-net.rules

This file contains existing interface names for persistency across reboots. If this file exists, the biosdevname will not kick in

d) Rename ifcfg-ethN files to ifcfg-emN and ifcfg-p*. The content of these files is distribution specific. Please modify the fields that contain the interface name to reflect the new names

As an example ifcfg-ethN files in RHEL 6.1 contain a 'DEVICE=' field which will be assigned with the ethN name. Please modify it to suit the new naming scheme such as 'DEVICE=em1'

If the ifcfg-ethN file embeds interface hardware address, please comment it as biosdevname suggests a name based on the location of the adapter on the mother board

e) Reboot the system

4.3. Installing Biosdevname from RPMs

The biosdevname can be installed using RPMSs built for the distribution you are using:

Fedora 15:

> **# rpm –ivh biosdevname-0.3.8-1.fc15.x86_64.rpm**

RHEL 6.1:

> **# rpm –ivh biosdevname-0.3.6-11.el6.x86_64.rpm**

4.4. Building biosdevname from source rpm

The biosdevname rpm can be built from the source rpm from:

http://linux.dell.com/biosdevname/

Pre-requisites: pciutils-devel and zlib-devel

> **# rpm –ivh biosdevname-<latest version>-1.src.rpm**

> **# rpmbuild –bb <path-to-biosdevname.spec file>**

On Fedora based distributions, this is under ~/rpmbuild/SPECS. On SuSE based distributions, this is /usr/src/packages/SPECS

> **# rpm –ivh <path-to-biosdevname-binary-rpm>**

4.5. New naming built into Operating System

If the OS has biosdevname support built in, then the new naming scheme is enforced by default during install time and run time on supported Dell systems

For a list of Linux Distributions where biosdevname is integrated into the installer, see Table 1.

5. Enable/Disable the new naming scheme

If you choose not to use the new naming scheme, it can be turned off. You could do it during install time or post install

5.1. During Install time

To disable the use of the new naming scheme during install (attended or automated), pass the kernel command line parameter biosdevname=0.

Certain distributions (see Table 1) have enabled the new naming scheme on all Dell systems by default and disabled for others. If you choose to use the new naming scheme on non-Dell systems with such distributions, set biosdevname=1 in the kernel command line during install.

5.2. Reverting to the traditional 'ethN' naming scheme

In case you have a system with the new naming scheme and would like to revert to the traditional naming scheme:

a) Rename all ifcfg-emN and ifcfg-p* configuration files and modify the contents of the files accordingly. The content of these files is distribution specific (see /usr/share/doc/initscripts-<version> for details) As an example ifcfg-ethN files in RHEL 6.1 contain a 'DEVICE=' field which will be assigned with the emN name. Modify it to suit the new naming scheme such as 'DEVICE=eth0'.

b) Comment the HWADDR variable in the ifcfg-eth* files if present as it is not possible to predict here which of the network devices will be named eth0, eth1 etc.

c) Reboot the system

d) Login and you could see the ethN names.

The **70-persistent-net.rules** file will be created under **/etc/udev/rules.d/** to make the names persistent across reboots.

5.3. Applications/Scripts using hardcoded 'ethN' names

If an application expects network interface names to be of the form ethN, then it might fail. Please report it via the distribution's bug tracking system

5.4. Examples:

On a PowerEdge R710 with

Four Broadcom BCM 5709 Lan-On-Motherboard ports

One Intel 82572EI single port network adapter on PCI slot 4

One Intel 82576 dual port network adapter on PCI slot 3

The naming of the network interfaces would be:

[root@per710~]# ls /sys/class/net/

em1 em2 em3 em4 lo p3p1 p3p2 p4p1

Using the new naming scheme

[root@per710 ~]# ifconfig em1

em1   Link encap:Ethernet HWaddr 00:24:E8:2E:DE:FF

          inet addr:172.16.66.166 Bcast:172.16.127.255 Mask:255.255.192.0

          …

[root@per710 ~]# ethtool -i p3p2_0

driver: igbvf

version: 1.0.8-k0

firmware-version: N/A

bus-info: 0000:07:10.1

Name for interfaces of Intel 82572EI single port network adapter on PCI slot 4

   p4p1

This conveys that the interface is from the network adapter on PCI slot 4 and port 1

5.5.   Naming scheme for network interfaces from SRIOV virtual functions

The interface names for virtual functions exposed by Intel 82576 network adapter would be (Assuming that we enable 2 virtual functions per physical function):

[root@per710 ~]# ls /sys/class/net/

em1 em2 em3 em4 lo p3p1 p3p1_0 p3p1_1 p3p2 p3p2_0 p3p2_1 p4p1

Physical function interface name	p3p1	p<slot 3>p<port number 1>
Virtual function interface names exposed by 'p3p1'	p3p1_0	p<slot 3>p<port number 1>_<virtual function instance 0> Virtual function instance 0 of port 1 of PCI add-in network adapter on SLOT 3
	p3p1_1	p<slot 3>p<port number 1>_<virtual function instance 1> (Virtual function instance 1 of port 1 of PCI add-in network adapter on SLOT 3)

6. Contribution

Your contribution in making biosdevname achieve its purpose comprehensively would be invaluable.

   a)   Test biosdevname and report bugs on the bug tracking systems corresponding to the distribution

        a.   "biosdevname -d" provides extensive debugging information that can assist with bug resolution

   b)   Contribute to enabling 'biosdevname' by fixing applications that depend on hardcoded eth names

7. More Information

Biosdevname manpage

8. Mailing Lists

Queries relating to 'biosdevname' can be discussed on

https://lists.us.dell.com/mailman/listinfo/linux-poweredge

linux-hotplug mailing list (linux-hotplug@vger.kernel.org)

9. References

Presentation at Texas Linux Fest – April 2011

> http://marc.info/?l=linux-hotplug&m=128892593821639&w=2

Presentation at Linux Plumbers Conference - 2010

> http://marc.info/?l=linux-hotplug&m=128892593821639&w=2

Fedora Feature Description

> http://fedoraproject.org/wiki/Features/ConsistentNetworkDeviceNaming

Fedora Test Day – January 2011

> https://fedoraproject.org/wiki/Test_Day:2011-01-

27_Network_Device_Naming_With_Biosdevname

SMBIOS v2.6 Reference Specification

> http://www.dmtf.org/sites/default/files/standards/documents/DSP0134_2.6.0.pdf

ACPI DSM for Naming a PCI or PCI Express Device Under Operating Systems

> http://www.pcisig.com/members/downloads/pcifw_r3_1_13Dec10.pdf