

# Performance of New 8020 Backend Board

Version 0.0.1 \*

Meihui Fan

范美辉

<[mhfan@ustc.edu](mailto:mhfan@ustc.edu)>

Copyright © 2005 HHTech. Co., Ltd. †  
All rights reserved.

华恒科技 版权所有

2005年6月7日

\*revised version

†<http://www.hhcn.com>

# 目录

<b>1</b>	<b>8020 后端板的千兆网口</b>	<b>3</b>
1.1	任务描述 . . . . .	3
1.2	调试记录 . . . . .	3
1.3	性能测试 . . . . .	4
1.3.1	测试环境 . . . . .	4
1.3.2	测试报告 . . . . .	5
1.3.3	结果分析 . . . . .	6
1.4	总结思考 . . . . .	6
<b>2</b>	<b>8020 后端板的RAID 模块</b>	<b>7</b>
2.1	任务描述 . . . . .	7
2.2	调试记录 . . . . .	7
2.3	性能测试 . . . . .	7
2.3.1	测试环境 . . . . .	7
2.3.2	测试报告 . . . . .	8
2.3.3	结果分析 . . . . .	10
2.4	总结思考 . . . . .	11

# 1 8020 后端板的千兆网口

## 1.1 任务描述

调通新8020 后端板的千兆网口(RTL8169)——缺少EEPROM，并评测网络性能。

## 1.2 调试记录

### 1. 现象描述：

PCI 总线检测到的RTL8169 的设备号是0x8129，不同于正常的0x8169。

### 原因分析：

未知，可能设备号也由EEPROM 中加载。

### 解决方法：

将RTL8169 驱动的PCI 设备匹配表的设备ID 改为，PCI\_ANY\_ID——即只匹配vendor ID。

### 2. 现象描述：

pci\_find\_capability 函数返回失败，导致内核恐慌。

### 原因分析：

EEPROM 中存有电源管理相关的信息，这个函数取不到RTL8169 的电源状态（可能是访问非法地址），所以出错。但是这不是必须的，事实上RealTek 发布的驱动中就没有这个功能。

### 解决方法：

comment out.

### 3. 现象描述：

驱动正常加载，发包正常，但没有收包中断；LED 显示，物理层 / 链路层工作正常。

### 原因分析：

可能是默认的参数不能正常工作。

The EEPROM interface provides the ability for the RTL8169S to read from and write data to an external serial EEPROM device. Values in the external EEPROM allow default fields in PCI configuration space and I/O space to be overridden following a reboot or software EEPROM auto load command. The RTL8169S will auto load values from the EEPROM. [If the EEPROM is not present, the RTL8169S initialization uses default values](#)

for the appropriate Configuration and Operational Registers. Software can read and write to the EEPROM using “bit-bang” accesses via the 9346CR Register. The interface consists of EESK, EECS, EEDO, and EEDI.

#### 解决方法：

焊上EEPROM 后，一切正常。另外，向RealTek 公司申请RTL8169 的软件编程资料；希望能够手动设置RTL8169 芯片的工作状态。

## 1.3 性能测试

### 1.3.1 测试环境

#### 硬件

	[TARGET]	[HOST]
CPU	PowerPC 7410, altivec supported	AMD Sempron(tm) 2500+
MEM	128MB SDRAM	256MB DDRAM
NET	RTL8169	RTL8169

用对接双绞线连接新8020 后端板上的千兆网口与 PC 机上的千兆网口，弃用板上的百兆网口。

#### 软件——LMbench

LMbench 的移植可以在命令行上完成，不需要调整任何代码：

```
$ cd /path/to/lmbench-root-directory/src
$ make OS=powerpc-7410-linux-gnu \
      CC=powerpc-7410-linux-gnu-gcc \
      LDFLAGS=-static
```

如果没有编译除错，会在<lmbench-root-directory>/bin 目录下对应OS 目录下生成一套系统性能测试工具集：

```
$ cd /path/to/lmbench-root-directory/bin
$ ls -R

.:
powerpc-7410-linux-gnu/

./powerpc-7410-linux-gnu:
bw_file_rd*  flushdisk*  lat_mmap*    lat_udp*     lmd*
bw_mem*      hello*      lat_pagefault*  lat_unix*    lmhttp*
```

bw_mmap_rd*	lat_connect*	lat_pipe*	lat_unix_connect*	loop_o*
bw_pipe*	lat_ctx*	lat_proc*	lib_stats.o	memsize*
bw_tcp*	lat_fcntl*	lat_rpc*	lib_tcp.o	mhz*
bw_unix*	lat_fifo*	lat_select*	lib_timing.o	msleep*
clock*	lat_fs*	lat_sig*	lib_udp.o	timing_o*
disk*	lat_http*	lat_syscall*	lib_unix.o	
enough*	lat_mem_rd*	lat_tcp*	lmbench.a	

这些程序可以运行于新8020 目标板用于性能测试。

### 1.3.2 测试报告

#### 测试命令

[TARGET]	[HOST]
# /path/to/bw_tcp -s	\$ /path/to/bw_tcp target-hostname
# /path/to/bw_tcp pc-hostname	\$ /path/to/bw_tcp -s
# /path/to/lat_tcp -s	\$ /path/to/lat_tcp target-hostname
# /path/to/lat_tcp pc-hostname	\$ /path/to/lat_tcp -s
# /path/to/lat_udp -s	\$ /path/to/lat_udp target-hostname
# /path/to/lat_udp pc-hostname	\$ /path/to/lat_udp -s
# /path/to/bw_unix	\$ /path/to/bw_unix
# /path/to/lat_unix	\$ /path/to/lat_unix
# /path/to/lat_connect -s	\$ /path/to/lat_connect target-hostname
# /path/to/lat_connect pc-hostname	\$ /path/to/lat_connect -s

#### Flood ping

```
$ ping -f 192.168.0.3
PING 192.168.0.3 (192.168.0.3) 56(84) bytes of data.

--- 192.168.0.3 ping statistics ---
96438482 packets transmitted, 96438482 received,
0% packet loss, time 7186346ms
rtt min/avg/max/mdev = 0.027/0.031/2948.612/0.560 ms,
pipe 5, ipg/ewma 0.074/0.048 ms
```

#### 带宽和延时

	TCP		UDP		UNIX SOCKET		CONN <sup>①</sup>	
H → T	58.46MB/s	58.85us	-	9.27us	740.94MB/s	7.26us	-	122.63us
T → H	57.08MB/s	59.08us	-	22.81us	172.28MB/s	14.70us	-	158.22us

### 1.3.3 结果分析

测试结果表明板子的网络性能表现良好，至少不应该会成为磁盘性能的瓶颈；但是网络延时劣于PC 的表现。另外，LMbench 测试套件中没有UDP 带宽的测试工具，可能需要找其它的测试套件。

## 1.4 总结思考

...

---

<sup>①</sup>TCP/IP connection

## 2 8020 后端板的RAID 模块

### 2.1 任务描述

评测新8020 后端板的RAID-5 性能。

### 2.2 调试记录

#### 1. 现象描述：

测试过程中在写磁盘的时候频繁报错：

```
drivers/scsi/gd31244/drv/gd31244.lld.c#3128:gd31244_isr:
*** ERROR: Ignoring command 0 that is not active (0)
```

#### 原因分析：

可能与磁盘供电有关，电压不稳或不足。

#### 解决方法：

...

#### 2. 现象描述：

多线程上传同时NFS 回放时内核打印内存分配错误：

```
swapper: page allocation failure. order:0, mode:0x20
```

#### 原因分析：

...

#### 解决方法：

...

### 2.3 性能测试

#### 2.3.1 测试环境

硬件

CPU	PowerPc 7410, altivec supported
DISK	4 × ST3160023AS(160GB, 7200.7, 8MB, SATA)
MEM	128MB SDRAM
NET	RTL8169

## 软件—— RAID HOWTO

1. 创建磁盘分区<sup>①</sup>:

```
# fdisk /dev/sda
```

2. Setup the /etc/raidtab file like this:

```
raiddev /dev/md0
raid-level          5
nr-raid-disks       4
nr-spare-disks      0
persistent-superblock 1
parity-algorithm    left-symmetric
chunk-size          32
device              /dev/sda1
raid-disk            0
device              /dev/sdb1
raid-disk            1
device              /dev/sdc1
raid-disk            2
device              /dev/sdd1
raid-disk            3
```

3. 创建RAID 卷:

```
# mkraid /dev/md0
```

4. 在RAID 卷上创建文件系统:

```
# mke2fs -b 4096 -R stride=8 /dev/md0
```

5. 加载RAID 卷:

```
# mount /dev/md0 /hd
```

### 2.3.2 测试报告

#### 测试命令

```
# /path/to/bw_mem 8m
# /path/to/lat_fs /path/to/directory
# /path/to/disk /path/to/raw-disk # e.g.: /dev/md0
# /path/to/bw_file.rd 8m open2close /path/to/bench-stuff
```

<sup>①</sup>单主分区每磁盘，分区格式 0xfd( Linux RAID auto)。



```
# /path/to/bw mmap rd 8m io_only /path/to/bench-stuff
$ time lftp -e "get /path/to/bench-stuff; quit" ftp-url-path
$ time lftp -e "put /path/to/bench-stuff; quit" ftp-url-path
$ time cp /path/to/nfs-mount-point /path/to/bench-stuff
$ time cp /path/to/bench-stuff /path/to/nfs-mount-point
# /path/to/time cp /path/to/nfs-mount-point /path/to/bench-stuff
# /path/to/time cp /path/to/bench-stuff /path/to/nfs-mount-point
$ time cp /path/to/smb-mount-point /path/to/bench-stuff
$ time cp /path/to/bench-stuff /path/to/smb-mount-point
# /path/to/lmdd if=internal of=/path/to/bench-stuff count=??? fsync=1
# /path/to/lmdd if=/path/to/bench-stuff of=internal
```

### RAID-5 测试结果

memory copy speeds: (MB/s)

rd	wr	rdwr	cp	fwr	frd	fcy	bzero	bcopy
395.84	204.32	201.10	126.10	257.33	236.01	112.94	276.30	112.70

file system create/delete performance:

size	files	create	delete
0k	1000	27134/s	119346/s
1k	1000	9251/s	48414/s
4k	1000	9300/s	48555/s
10k	1000	6633/s	36880/s

disk performance<sup>②</sup>: (MB/s)

	zone	seek
raw disk perf	1.21/35.759/146.78 (MB/s)	1.57/14.2448/78.25 (ms)
	read	mmap
file bandwidth	145.49	235.99
	read	write
lmdd bandwidth	39.18	35.84
ftpd bandwidth	22.78	15.57
nfsd bandwidth	21.99	14.35
nfs bandwidth	15.89	14.23
smbd bandwidth	11.42	2.42

<sup>②</sup>测试文件大小: 1073565696 B。

## 单块硬盘测试结果

Single ST380013AS, 80GB, 7200.7, 8MB, SATA.

file system create/delete performance:

size	files	create	delete
0k	1000	9339/s	43170/s
1k	1000	5128/s	19192/s
4k	1000	5139/s	19369/s
10k	1000	3724/s	16474/s

disk performace<sup>③</sup>: (MB/s)

	zone	seek
raw disk perf	30.05/39.9131/47.53 (MB/s)	4.50/15.203/29.52 (ms)
	read	mmap
file bandwidth	144.89	236.08
	read	write
lmdd bandwidth	56.12	47.36
ftpd bandwidth	25.54	16.78
nfsd bandwidth	25.34	14.79
nfs bandwidth	15.38	14.78
smbd bandwidth	10.05	2.34

## 上传 / 下载和NFS 回放

FTP: 8 线程同时上传 / 下载, 没线程带宽: 1.2 MB/s 1.4 MB/s; 回放基本流畅, 偶尔丢帧, 这应该是极限情况。

SAMBA: 6 进程上传 / 下载时回放正常; 再有进程上传 / 下载过程中没有问题, 但当文件上传 / 下载完成时会有丢帧。

## 2.3.3 结果分析

测试结果说明在新8020 后端板上, RAID-5 似乎没有起到磁盘访问加速的作用, 尤其与单块硬盘的性能相比, 在某些测试上RAID-5 还不如单块磁盘的结果。我确定我的操作没有问题, 所以可能的原因如下:

- ❁ 板子的磁盘性能在物理参数上优于PC 机, 所以对于联网测试性能瓶颈在于PC 机; 所以如果要测试出板子的真实性能必须采用在物理参数至少是相同的硬件平台。

<sup>③</sup>测试文件大小: 1073565696 B。

- ✿ Lmbench 测试套件可能测试不出RAID-5 的性能（这似乎不应该，只是怀疑），可能需要找一些其它的测试软件，最好是寻找和采用专门的磁盘性能测试软件（RAID 磁盘性能测试软件？）。

## 2.4 总结思考

...

## 文献

- ☞ “LMbench document and manual”, <http://www.bitmover.com/lmbench/>.
- ☞ “RTL8169S-32/RTL8169S-64 DATASHEET”, <http://www.realtek.com.tw>.
- ☞ “RAID 技术简介”, <http://www.forecast.com.cn/network/jishuzhichi-xt-4.htm>
- ☞ “Software-RAID-HOWTO”, <http://www.tldp.org/HOWTO/Software-RAID-HOWTO.html>
- ☞ “RAID 磁盘阵列技术全面介绍”, [http://www.ccw.com.cn/server/jssc/htm2005/20050520\\_14RU3.htm](http://www.ccw.com.cn/server/jssc/htm2005/20050520_14RU3.htm)
- ☞ “Redundant array of independent disks”, [http://en.wikipedia.org/wiki/Redundant\\_array\\_of\\_independent\\_disks](http://en.wikipedia.org/wiki/Redundant_array_of_independent_disks)