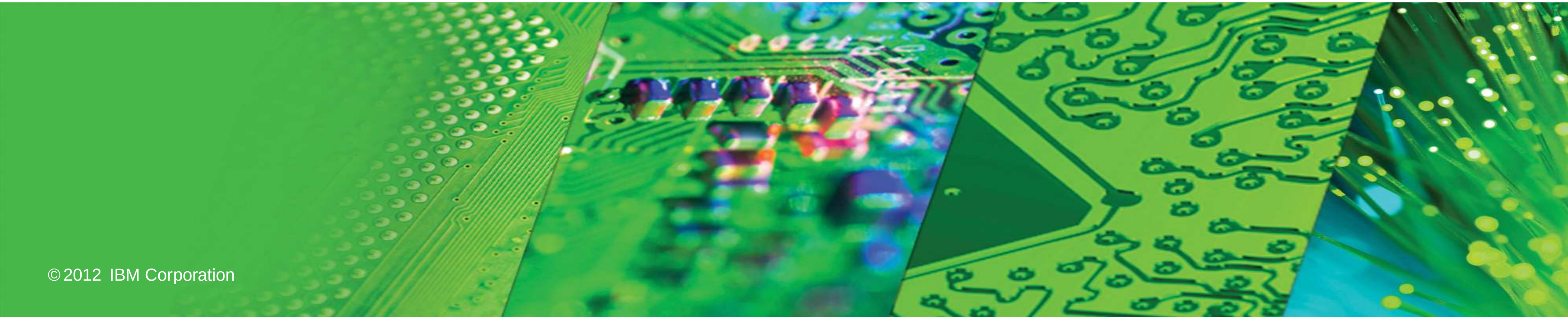
2012
IBM System z Technical University
Enabling the infrastructure for smarter computing

Advanced Networking with Linux on System z

zLG22

Dr. Stefan Reimbold



Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at Copyright and trademark information at www.ibm.com/legal/copytrade.shtml.

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply. All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions. This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area. All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

Agenda

- Overview
- Basic Network Setup
- Advanced Network Setup
- Networking Tools
- Problem Determination / Debugging

Agenda

- Overview
- Advanced Network Setup
- Networking Tools
- Problem Determination / Debugging

Networking Options

- OSA
- HiperSockets
- Virtual NIC
  - Guest LAN
  - VSWITCH
- LCS
- CTC
- NETIUCV

Networking Drivers

- QETH
- LCS
- CTC (functionally stable)
- NETIUCV (functionally stable)

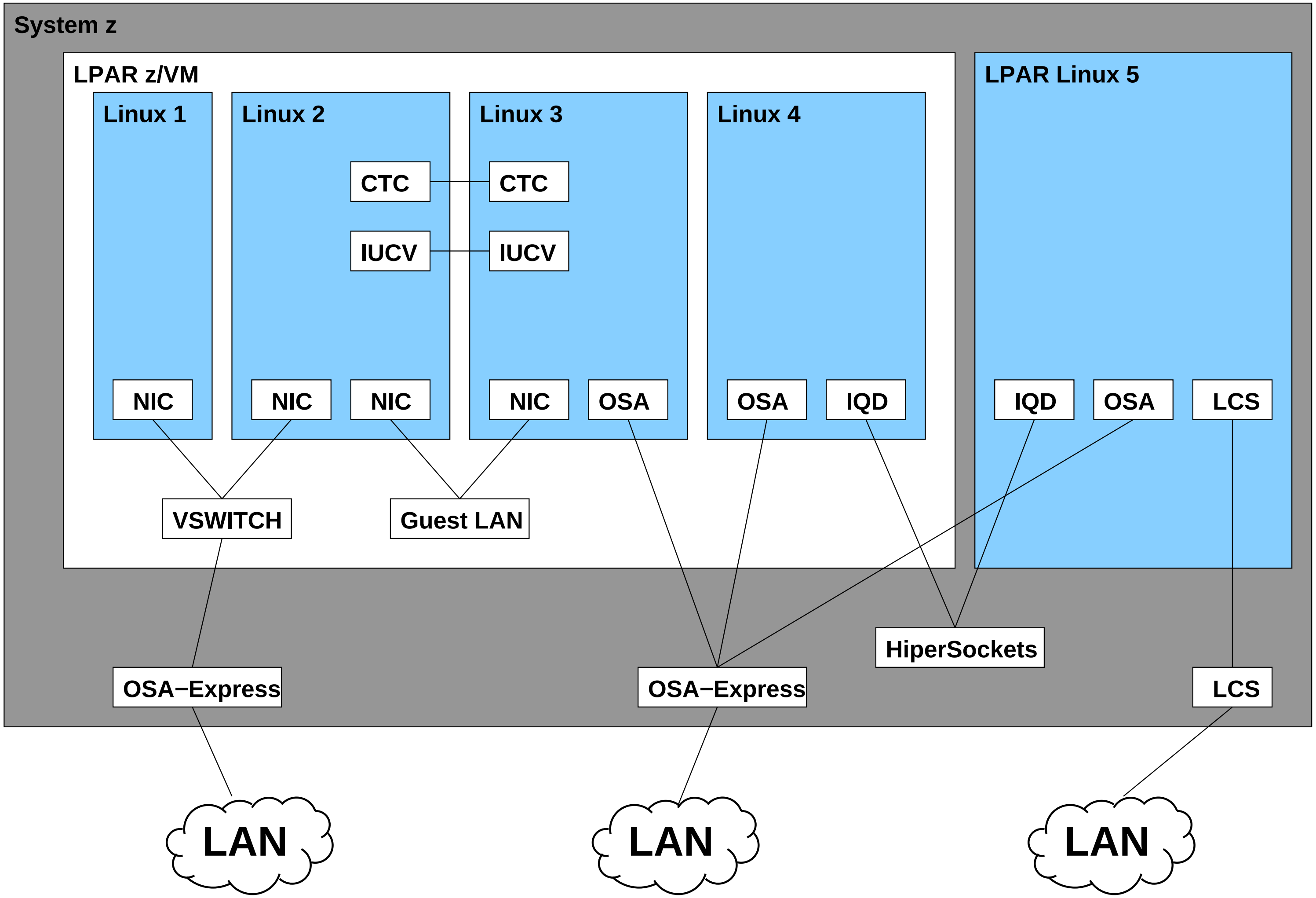
Networking Drivers

- CTC - Channel-To-Channel Connection
- IUCV - Inter User Communication Vehicle
- device driver deprecated (kernel 2.6)
- still available for backwards comptibility
- migration path
  - Virtual CTC and IUCV ⇒ Guest LAN
  - CTC in LPAR ⇒ HiperSockets
  - CTC ⇒ OSA-Express

QETH Device Driver

- supports
  - OSA-Express
  - HiperSockets
  - Guest LAN
  - VSWITCH
- primary network driver for Linux on System z

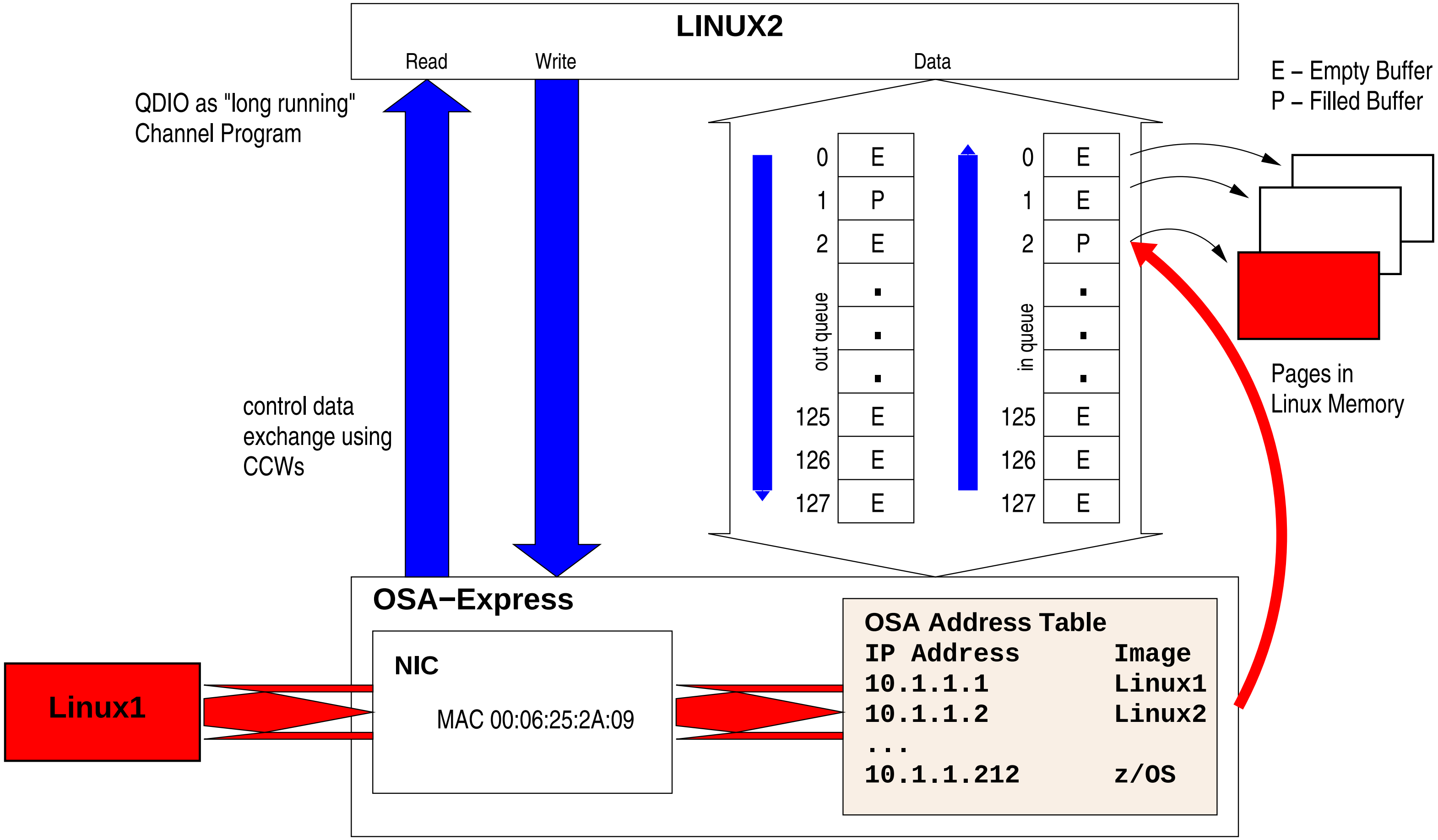
System z Network



QDIO - Queued Direct I/O

- OSA
- HiperSockets
- FCP

QDIO Architecture



E – Empty Buffer
P – Filled Buffer

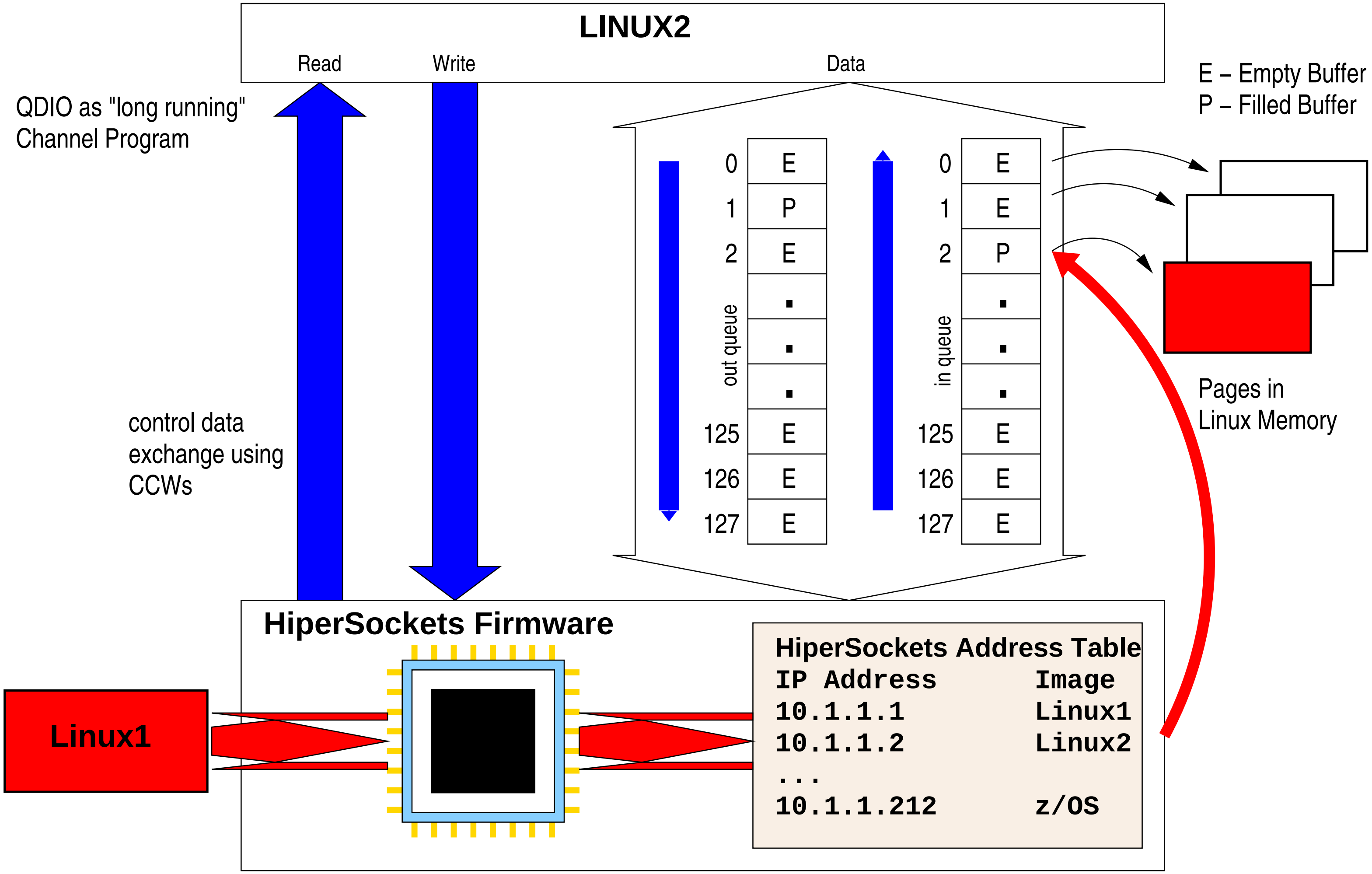
out queue	
0	E
1	P
2	E
	.
	.
	.
125	E
126	E
127	E

in queue	
0	E
1	E
2	P
	.
	.
	.
125	E
126	E
127	E

OSA Address Table

IP Address	Image
10.1.1.1	Linux1
10.1.1.2	Linux2
...	
10.1.1.212	z/OS

QDIO Architecture



out queue	
0	E
1	P
2	E
	.
	.
	.
125	E
126	E
127	E

in queue	
0	E
1	E
2	P
	.
	.
	.
125	E
126	E
127	E

HiperSockets Address Table

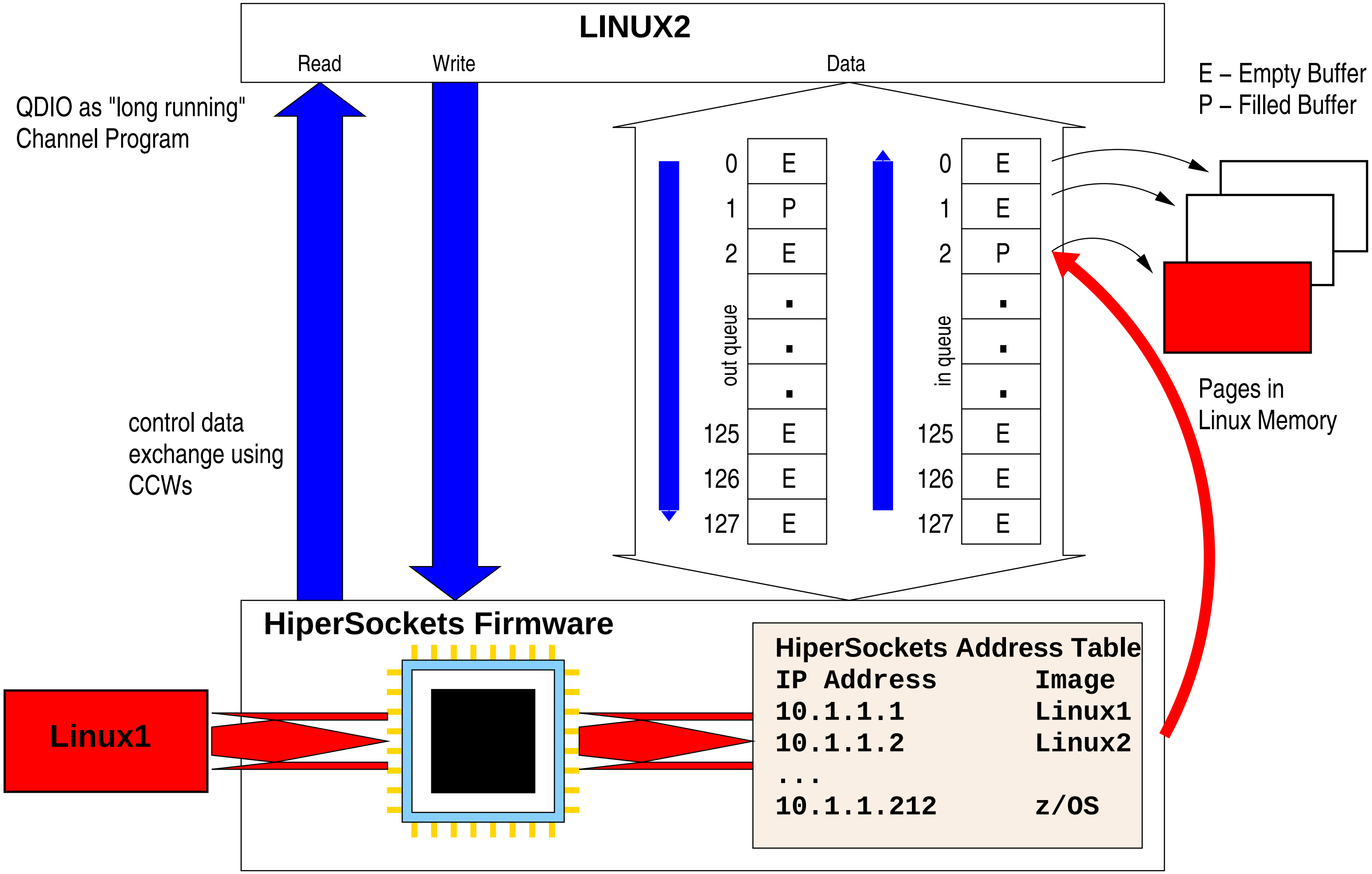
IP Address	Image
10.1.1.1	Linux1
10.1.1.2	Linux2
...	
10.1.1.212	z/OS

Layer 2 Mode

OSI Model	TCP/IP Model		
	Layer	Protocol	
7 Application	Application	FTP, HTTP	
6 Presentation			
5 Session			
4 Transport	Transport	TCP, UDP	
3 Network	Internet	IP, ICMP	arp
2 Data Link	Network Access	Ethernet	
1 Physical			

Layer 3 Frame			IP	TCP	Data
Layer 2 Frame	TGT MAC	SRC MAC	IP	TCP	Data

Layer 2 Mode



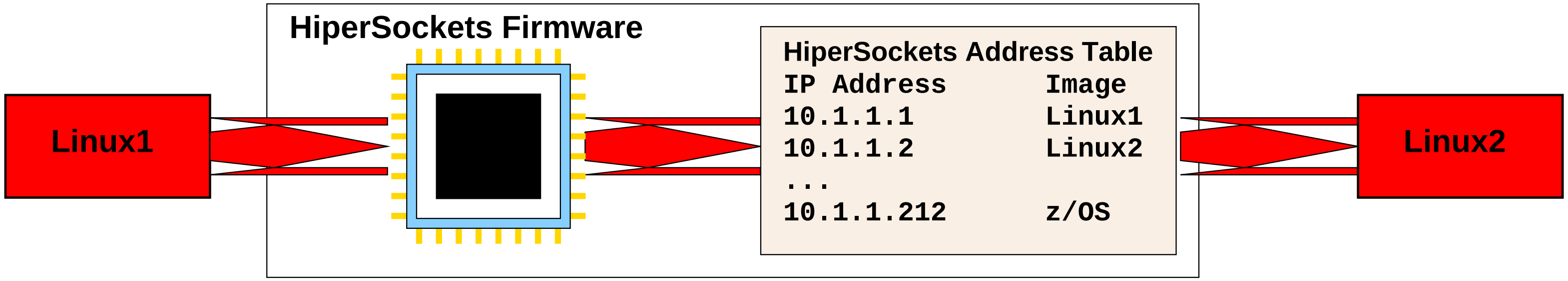
out queue	
0	E
1	P
2	E
	.
	.
	.
125	E
126	E
127	E

in queue	
0	E
1	E
2	P
	.
	.
	.
125	E
126	E
127	E

HiperSockets Address Table

IP Address	Image
10.1.1.1	Linux1
10.1.1.2	Linux2
...	
10.1.1.212	z/OS

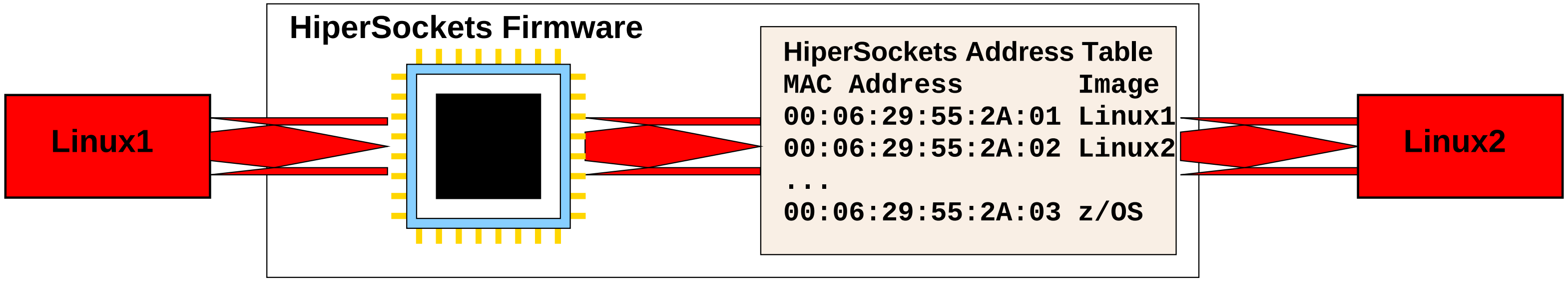
Layer 2 Mode



HiperSockets Address Table

IP Address	Image
10.1.1.1	Linux1
10.1.1.2	Linux2
...	
10.1.1.212	z/OS

Layer 2 Mode



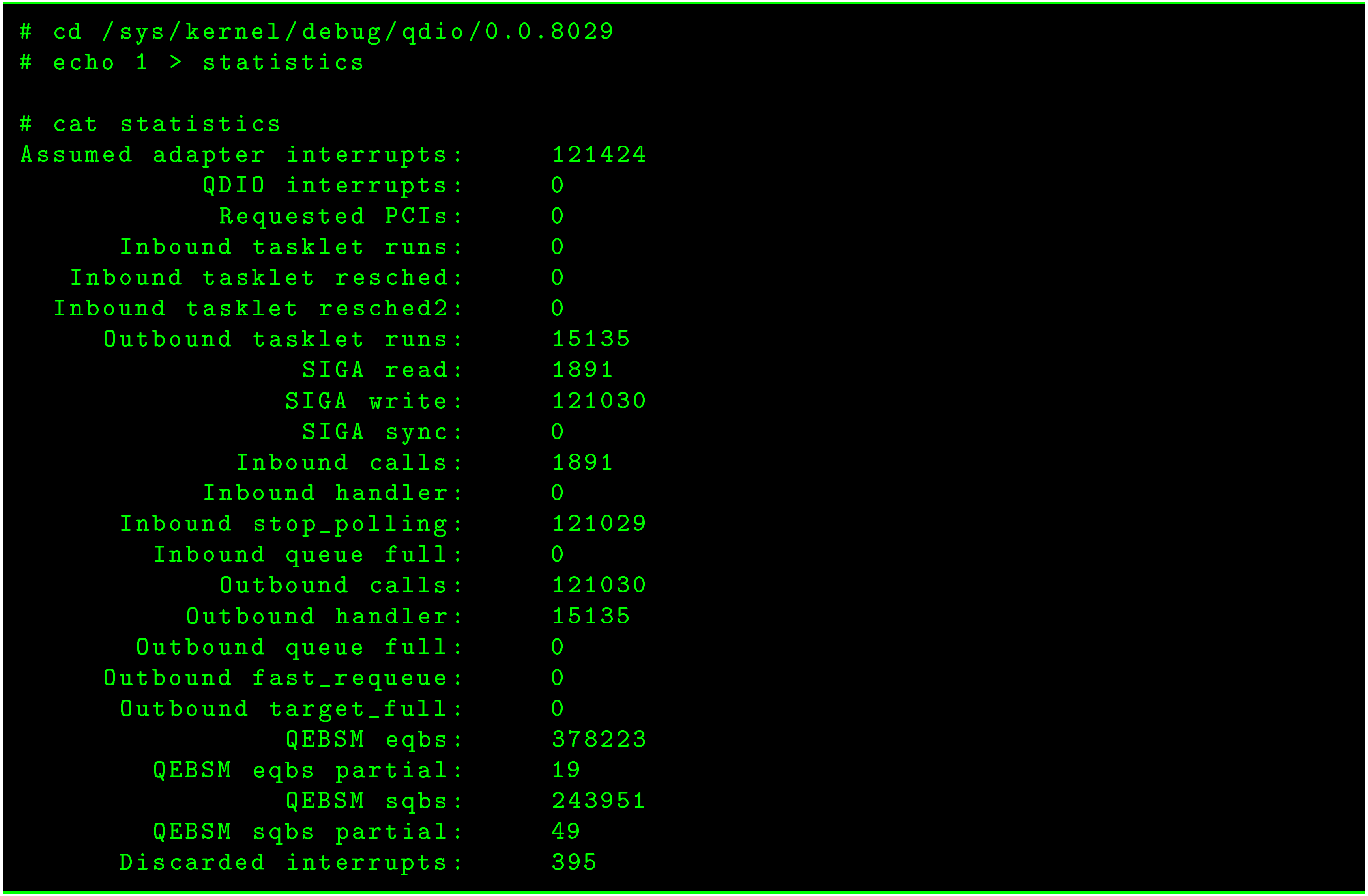
HiperSockets Address Table

MAC Address	Image
00:06:29:55:2A:01	Linux1
00:06:29:55:2A:02	Linux2
...	
00:06:29:55:2A:03	z/OS

Layer 2 Mode

- HiperSockets
  - Layer 2 and Layer 3 seperated
- Guest LAN
- VSWITCH

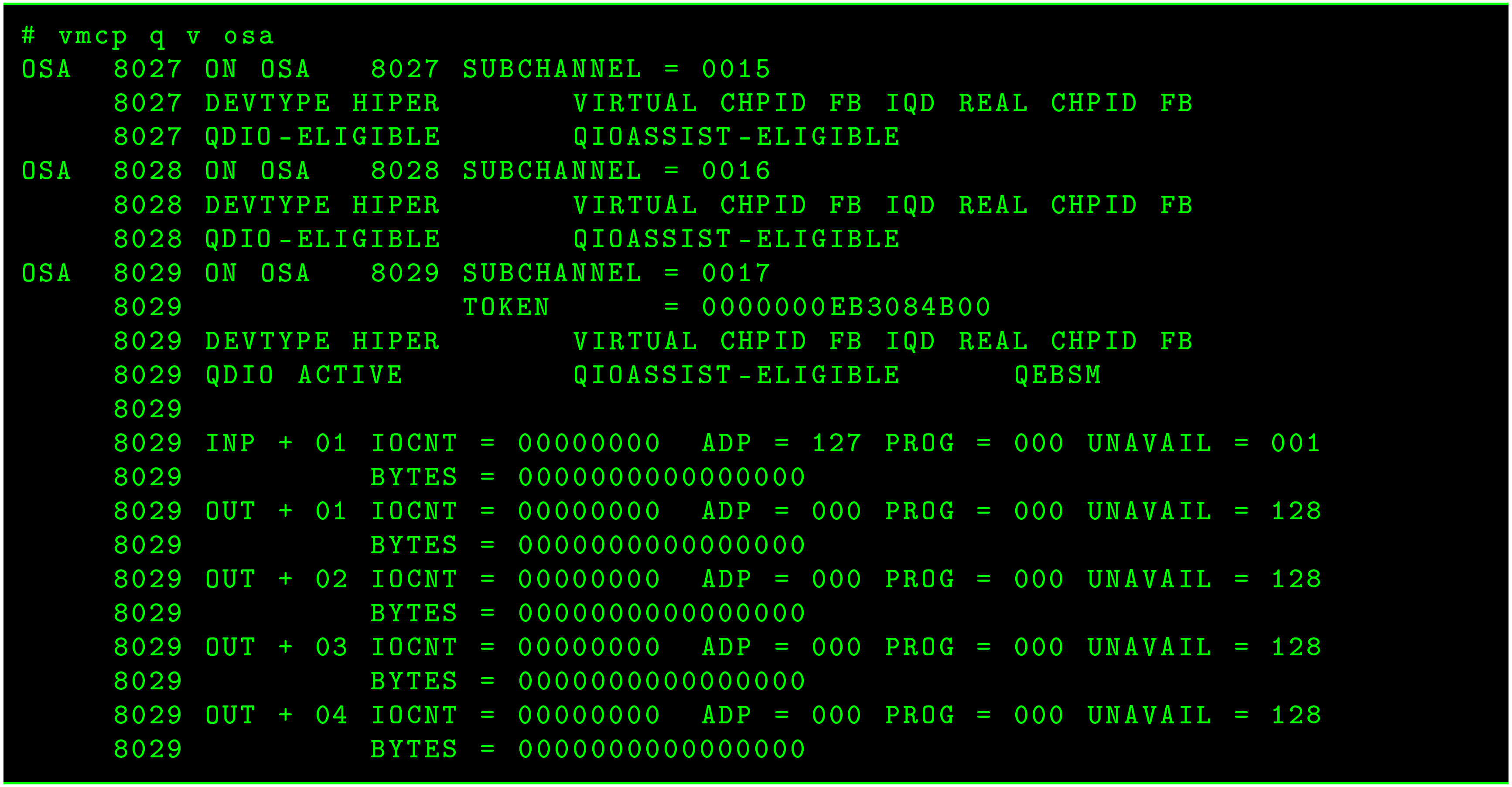
Statistics



```
# cd /sys/kernel/debug/qdio/0.0.8029
# echo 1 > statistics

# cat statistics
Assumed adapter interrupts:     121424
           QDIO interrupts:     0
            Requested PCIs:     0
      Inbound tasklet runs:     0
   Inbound tasklet resched:     0
  Inbound tasklet resched2:     0
     Outbound tasklet runs:     15135
                 SIGA read:     1891
                SIGA write:     121030
                 SIGA sync:     0
             Inbound calls:     1891
           Inbound handler:     0
      Inbound stop_polling:     121029
        Inbound queue full:     0
            Outbound calls:     121030
          Outbound handler:     15135
       Outbound queue full:     0
     Outbound fast_requeue:     0
      Outbound target_full:     0
                QEBSM eqbs:     378223
        QEBSM eqbs partial:     19
                QEBSM sqbs:     243951
        QEBSM sqbs partial:     49
      Discarded interrupts:     395
```

QDIO Assist



```
# vmcp q v osa
OSA  8027 ON OSA   8027 SUBCHANNEL = 0015
     8027 DEVTYPE HIPER       VIRTUAL CHPID FB IQD REAL CHPID FB
     8027 QDIO-ELIGIBLE       QIOASSIST-ELIGIBLE
OSA  8028 ON OSA   8028 SUBCHANNEL = 0016
     8028 DEVTYPE HIPER       VIRTUAL CHPID FB IQD REAL CHPID FB
     8028 QDIO-ELIGIBLE       QIOASSIST-ELIGIBLE
OSA  8029 ON OSA   8029 SUBCHANNEL = 0017
     8029               TOKEN      = 0000000EB3084B00
     8029 DEVTYPE HIPER       VIRTUAL CHPID FB IQD REAL CHPID FB
     8029 QDIO ACTIVE         QIOASSIST-ELIGIBLE      QEBSM
     8029
     8029 INP + 01 IOCNT = 00000000  ADP = 127 PROG = 000 UNAVAIL = 001
     8029          BYTES = 0000000000000000
     8029 OUT + 01 IOCNT = 00000000  ADP = 000 PROG = 000 UNAVAIL = 128
     8029          BYTES = 0000000000000000
     8029 OUT + 02 IOCNT = 00000000  ADP = 000 PROG = 000 UNAVAIL = 128
     8029          BYTES = 0000000000000000
     8029 OUT + 03 IOCNT = 00000000  ADP = 000 PROG = 000 UNAVAIL = 128
     8029          BYTES = 0000000000000000
     8029 OUT + 04 IOCNT = 00000000  ADP = 000 PROG = 000 UNAVAIL = 128
     8029          BYTES = 0000000000000000
```

QDIO Assist

- avoids interception to z/VM
- SQBS and EQBS are used to synchronize buffer states with z/VM storage
  - SQBS to write buffer state
  - EQBS to read buffer state
- big performance improvement because CPU always in guest level

Bonding

```
# ifconfig eth0 hw ether 00:06:29:55:2A:01
# ifconfig eth1 hw ether 00:06:29:55:2A:02

# modprobe bonding

# ifconfig bond0 10.40.33.38 netmask 255.255.255.0

# ifenslave bond0 eth0
# ifenslave bond0 eth1

# ifenslave bond0
The result of SIOCGIFFLAGS on bond0 is 1443.
The result of SIOCGIFADDR is 0a.28.21.26.
The result of SIOCGIFHWADDR is type 1  06:00:fb:0f:00:2c.
```

Bonding

```
cat /proc/net/bonding/bond0
Ethernet Channel Bonding Driver: v3.7.1 (April 27, 2011)

Bonding Mode: load balancing (round-robin)
MII Status: up
MII Polling Interval (ms): 0
Up Delay (ms): 0
Down Delay (ms): 0

Slave Interface: hsi0
MII Status: up
Speed: 10000 Mbps
Duplex: full
Link Failure Count: 0
Permanent HW addr: 06:00:fb:0f:00:2c
Slave queue ID: 0
```

QETH - large_send

- offload TCP segmentation from TCP/IP stack to OSA card
- move workload to OSA-Express adapter
- better performance with large outgoing packets

```
# echo TSO > /sys/devices/qeth/0.0.8027/large_send
```

```
QETH_OPTIONS='large_send=TSO'
```

- offload TCP segmentation from TCP/IP stack to device driver

```
# echo EDDP > /sys/devices/qeth/0.0.8027/large_send
```

```
QETH_OPTIONS='large_send=EDDP'
```

QETH - checksumming

- offload checksum calculation for incoming packets from TCP/IP stack to OSA card
- move workload to OSA-Express adapter
- Available for OSA devices in Layer 3 mode

```
# echo hw_checksumming > /sys/devices/qeth/0.0.8027/checksumming
```

```
QETH_OPTIONS='checksumming=hw_checksumming'
```

- remove checksum calculation for trusted HiperSockets connections
- reduce CPU load of TCP/IP stack

```
# echo no_checksumming > /sys/devices/qeth/0.0.8027/checksumming
```

```
QETH_OPTIONS='checksumming=no_checksumming'
```

QETH - buffer_count

- reduce buffers to reduce memory usage
- increase buffers to increase performance

```
# echo 64 > /sys/devices/qeth/0.0.8027/buffer_count
```

- need to set device offline

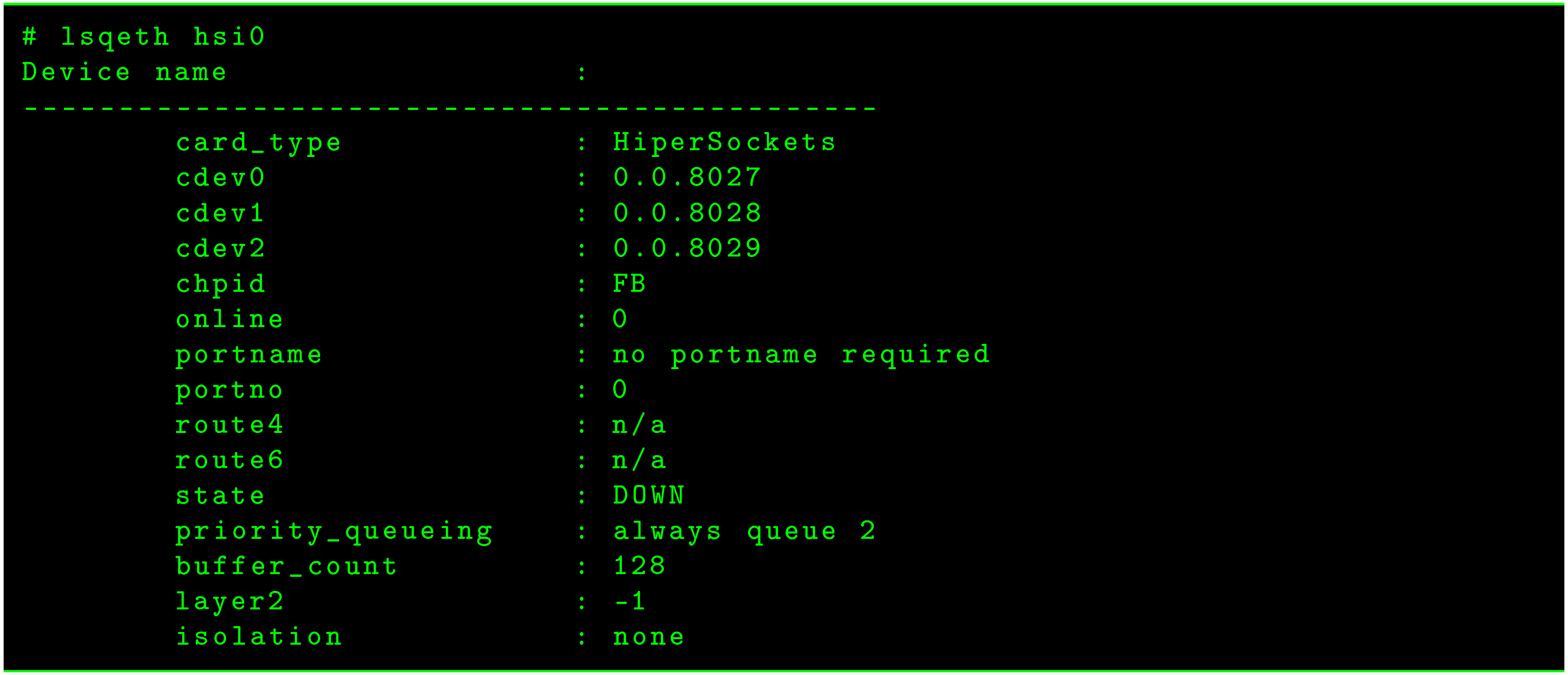
Network Tools

- lsqeth
- /proc/qeth
- ifconfig
- ping
- route
- netstat
- traceroute
- znetconf
- qetharp
- tcpdump / wireshark

lsqeth

```
 # lsqeth -p
 devices                    CHPID interface   cardtype
port chksum prio-q'ing rtr4 rtr6 lay'2 cnt
 -------------------------- ----- ---------- -------------- ---- ------ ---------- ---- ---- ----- -----
 0.0.8027/0.0.8028/0.0.8029 xFB   hsi0        HiperSockets   0    sw     always_q_2 no   no
0      128
 0.0.f503/0.0.f504/0.0.f505 x02   eth0        GuestLAN QDIO  0    sw     always_q_2 no   no
0      64
```

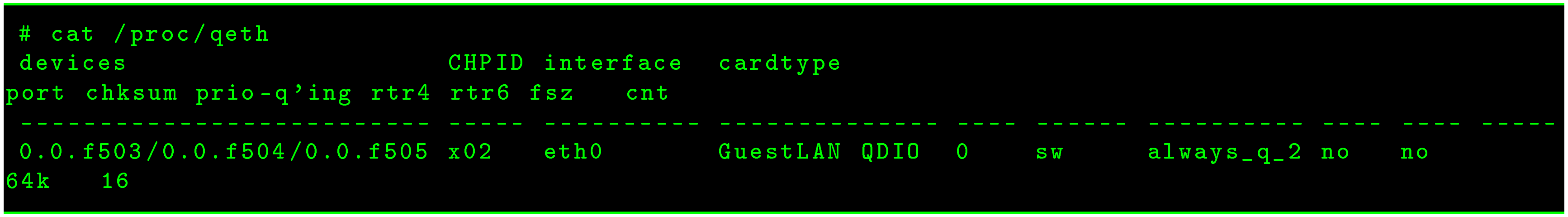
lsqeth



```
# lsqeth hsi0
Device name                   :
---------------------------------------------
        card_type             : HiperSockets
        cdev0                 : 0.0.8027
        cdev1                 : 0.0.8028
        cdev2                 : 0.0.8029
        chpid                 : FB
        online                : 0
        portname              : no portname required
        portno                : 0
        route4                : n/a
        route6                : n/a
        state                 : DOWN
        priority_queueing     : always queue 2
        buffer_count          : 128
        layer2                : -1
        isolation             : none
```

/proc/qeth

- RHEL5 and SLES10



```
 # cat /proc/qeth
 devices                    CHPID interface  cardtype
port chksum prio-q'ing rtr4 rtr6 fsz   cnt
 -------------------------- ----- ---------- -------------- ---- ------ ---------- ---- ---- ----- -----
 0.0.f503/0.0.f504/0.0.f505 x02   eth0       GuestLAN QDIO  0    sw     always_q_2 no   no
64k   16
```

ifconfig

- `ifconfig hsi0`

```
# ifconfig hsi0
hsi0      Link encap:Ethernet  HWaddr 06:00:FB:0F:00:29
          inet addr:10.40.33.38  Bcast:10.40.33.255  Mask:255.255.255.0
          inet6 addr: fe80::400:fbff:fe0f:29/64 Scope:Link
          UP BROADCAST RUNNING NOARP MULTICAST  MTU:8192  Metric:1
          RX packets:8681429 errors:0 dropped:0 overruns:0 frame:0
          TX packets:8681440 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:34265850338 (32678.4 Mb)  TX bytes:34387387822 (32794.3 Mb)
```

ifconfig

- `ifconfig hsi0 down`

```
# ifconfig hsi0 down
# ifconfig hsi0
hsi0      Link encap:Ethernet  HWaddr 06:00:FB:0F:00:29
          inet addr:10.40.33.38  Bcast:10.40.33.255  Mask:255.255.255.0
          BROADCAST NOARP MULTICAST  MTU:8192  Metric:1
          RX packets:8681429 errors:0 dropped:0 overruns:0 frame:0
          TX packets:8681440 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:34265850338 (32678.4 Mb)  TX bytes:34387387822 (32794.3 Mb)
```

- `ifconfig hsi0 up`

```
# ifconfig hsi0 up
```

ifconfig

- configure IP address

```
# ifconfig hsi0 10.40.33.38
```

- configure subnet

```
# ifconfig hsi0 10.40.33.38 network 255.255.255.0
```

```
# ifconfig hsi0 10.40.33.38/24
```

ifconfig

- configure MTU size

```
# ifconfig hsi0 mtu 1024
# ifconfig hsi0
hsi0      Link encap:Ethernet  HWaddr 06:00:FB:0F:00:29
          inet addr:10.40.33.38  Bcast:10.40.33.255  Mask:255.255.255.0
          UP BROADCAST RUNNING NOARP MULTICAST  MTU:1024  Metric:1
          RX packets:8681429 errors:0 dropped:0 overruns:0 frame:0
          TX packets:8681443 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:34265850338 (32678.4 Mb)  TX bytes:34387388032 (32794.3 Mb)
```

ping

- `ping -s`
- `ping -c`
- `ping -f`

ping

- `ping -s`
  - set size of ping packets
- `ping -c`
  - set number of ping packets to send

```
# ping -s 1024 -c 2 10.40.33.39
PING 10.40.33.39 (10.40.33.39) 1024(1052) bytes of data.
1032 bytes from 10.40.33.39: icmp_seq=1 ttl=64 time=0.118 ms
1032 bytes from 10.40.33.39: icmp_seq=2 ttl=64 time=0.073 ms

--- 10.40.33.39 ping statistics ---
2 packets transmitted, 2 received, 0% packet loss, time 999ms
rtt min/avg/max/mdev = 0.073/0.095/0.118/0.024 ms
```

ping

- `ping -f`
  - flood ping
  - send packets as fast as you can
  - print dot for sent packet - backspace for received
  - need to be root

```
# ping -f -c 1000 10.40.33.39
PING 10.40.33.39 (10.40.33.39) 56(84) bytes of data.
...
```

ping

- `ping -f`
  - flood ping
  - send packets as fast as you can
  - print dot for sent packet - backspace for received
  - need to be root

```
# ping -f -c 1000 10.40.33.39
PING 10.40.33.39 (10.40.33.39) 56(84) bytes of data.

--- 10.40.33.39 ping statistics ---
1000 packets transmitted, 1000 received, 0% packet loss, time 35ms
rtt min/avg/max/mdev = 0.012/0.026/0.126/0.008 ms, ipg/ewma 0.035/0.024 ms
```

ping

```
# ping -f -c 1000000 -s 8192 10.40.33.39
PING 10.40.33.39 (10.40.33.39) 8192(8220) bytes of data.

--- 10.40.33.39 ping statistics ---
1000000 packets transmitted, 1000000 received, 0% packet loss, time 55057ms
rtt min/avg/max/mdev = 0.018/0.039/0.264/0.009 ms, ipg/ewma 0.055/0.041 ms
```

$$\frac{8\,\text{MB}\cdot 1,000,000}{55\ \text{s}} = 145\ \text{kB}/\text{s}$$

ping

```
# ping -f -c 1000000 -s 8192 localhost
PING localhost (127.0.0.1) 8192(8220) bytes of data.

--- localhost ping statistics ---
1000000 packets transmitted, 1000000 received, 0% packet loss, time 10386ms
rtt min/avg/max/mdev = 0.002/0.002/0.297/0.002 ms, ipg/ewma 0.010/0.002 ms
```

$$\frac{8\,\mathsf{kB}\cdot 1,000,000}{10\ \mathsf{s}} = 800\ \mathsf{MB}/\mathsf{s}$$

ping

```
# ping -f -c 100000 -s 56000 10.40.33.39
PING 10.40.33.39 (10.40.33.39) 56000(56028) bytes of data.

--- 10.40.33.39 ping statistics ---
100000 packets transmitted, 100000 received, 0% packet loss, time 19222ms
rtt min/avg/max/mdev = 0.076/0.135/3.171/0.025 ms, ipg/ewma 0.192/0.139 ms
```

$$\frac{56,000\,\mathsf{B}\cdot 100,000}{19\ \mathsf{s}} = 295\ \mathsf{MB/s}$$

route

- route
- route -n

```
# route -n
Kernel IP routing table
Destination     Gateway         Genmask         Flags Metric Ref    Use Iface
0.0.0.0         9.152.108.1     0.0.0.0         UG    0      0        0 eth0
9.152.108.0     0.0.0.0         255.255.252.0   U     0      0        0 eth0
10.40.33.0      0.0.0.0         255.255.255.0   U     0      0        0 hsi0
127.0.0.0       0.0.0.0         255.0.0.0       U     0      0        0 lo
169.254.0.0     0.0.0.0         255.255.0.0     U     0      0        0 eth0
```

route

- route add

```
# route add -net 10.40.33.0 netmask 255.255.255.0 dev eth0
# route -n
Kernel IP routing table
Destination     Gateway         Genmask         Flags Metric Ref    Use Iface
0.0.0.0         9.152.108.1     0.0.0.0         UG    0      0        0 eth0
9.152.108.0     0.0.0.0         255.255.252.0   U     0      0        0 eth0
10.40.33.0      0.0.0.0         255.255.255.0   U     0      0        0 eth0
10.40.33.0      0.0.0.0         255.255.255.0   U     0      0        0 hsi0
127.0.0.0       0.0.0.0         255.0.0.0       U     0      0        0 lo
169.254.0.0     0.0.0.0         255.255.0.0     U     0      0        0 eth0
```

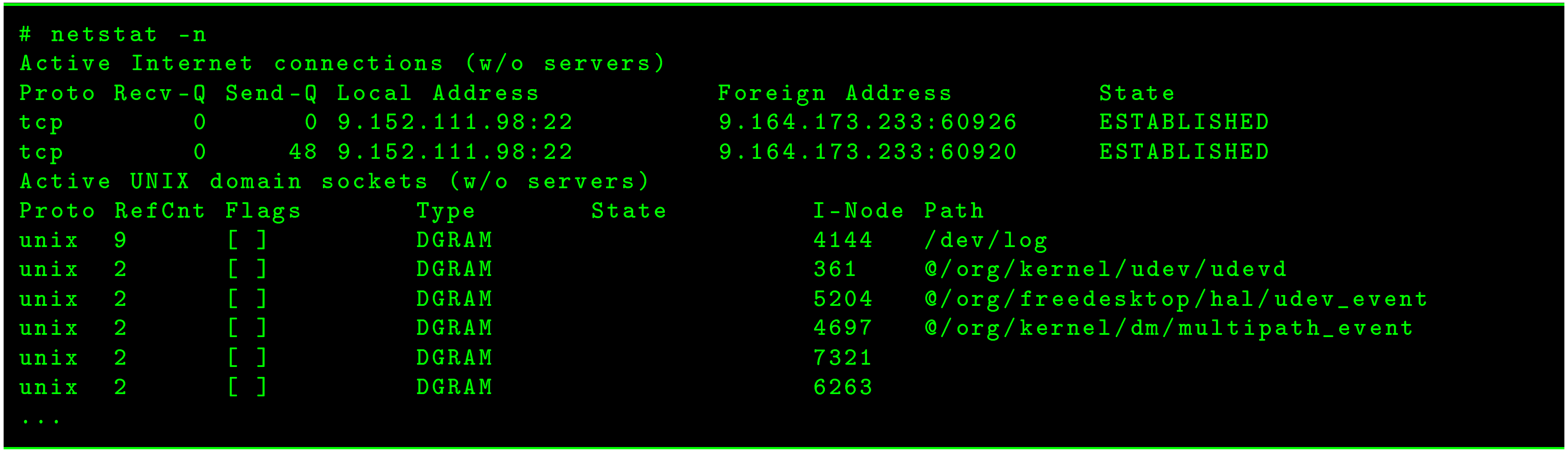
route

- route del

```
# route del -net 10.40.33.0 netmask 255.255.255.0 dev eth0
# route -n
Kernel IP routing table
Destination     Gateway         Genmask         Flags Metric Ref    Use Iface
0.0.0.0         9.152.108.1     0.0.0.0         UG    0      0        0 eth0
9.152.108.0     0.0.0.0         255.255.252.0   U     0      0        0 eth0
10.40.33.0      0.0.0.0         255.255.255.0   U     0      0        0 hsi0
127.0.0.0       0.0.0.0         255.0.0.0       U     0      0        0 lo
169.254.0.0     0.0.0.0         255.255.0.0     U     0      0        0 eth0
```

netstat

- `netstat -n`
  - always use `-n`



```
# netstat -n
Active Internet connections (w/o servers)
Proto Recv-Q Send-Q Local Address           Foreign Address         State
tcp        0      0 9.152.111.98:22         9.164.173.233:60926     ESTABLISHED
tcp        0     48 9.152.111.98:22         9.164.173.233:60920     ESTABLISHED
Active UNIX domain sockets (w/o servers)
Proto RefCnt Flags       Type       State         I-Node Path
unix  9      [ ]         DGRAM                    4144   /dev/log
unix  2      [ ]         DGRAM                    361    @/org/kernel/udev/udevd
unix  2      [ ]         DGRAM                    5204   @/org/freedesktop/hal/udev_event
unix  2      [ ]         DGRAM                    4697   @/org/kernel/dm/multipath_event
unix  2      [ ]         DGRAM                    7321
unix  2      [ ]         DGRAM                    6263
...
```

traceroute

- traceroute

```
# traceroute r3515039
traceroute to r3515039 (9.152.111.99), 30 hops max, 40 byte packets using UDP
 1  r3515039.boeblingen.de.ibm.com (9.152.111.99)  0.096 ms   0.013 ms   0.012 ms
```

```
# traceroute 10.40.33.38
traceroute to 10.40.33.38 (10.40.33.38), 30 hops max, 40 byte packets using UDP
 1  10.40.33.38 (10.40.33.38)  0.000 ms   0.000 ms   0.000 ms
```

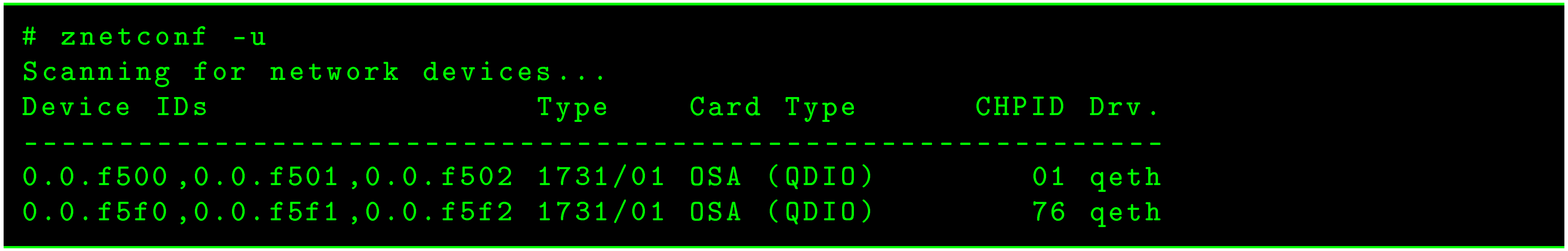
traceroute

```
# traceroute www.ibm.com
traceroute to www.ibm.com (129.42.58.158), 30 hops max, 40 byte packets
 1  10.21.0.1 (10.21.0.1)  1.178 ms  1.124 ms  1.090 ms
 2  10.20.0.3 (10.20.0.3)  0.388 ms  0.371 ms  0.341 ms
 3  AFS-Boise-216-222-81-129.afsnetworks.com (216.222.81.129)  0.594 ms  0.554 ms  0.519 ms
 4  AFS-Boise-216-222-87-5.afsnetworks.com (216.222.87.5)  4.253 ms  4.228 ms  4.195 ms
 5  AFS-Boise-216-222-87-2.afsnetworks.com (216.222.87.2)  2.133 ms  2.081 ms  2.041 ms
 6  AFS-LasVegas-216-222-65-97.afsnetworks.com (216.222.65.97)  3.939 ms  4.195 ms  4.164 ms
 7  g7-2.las-esw03.switchnap.com (66.209.87.93)  4.581 ms  4.554 ms  4.535 ms
 8  te2-7.las-core2-1.switchnap.com (66.209.64.113)  4.479 ms  4.588 ms  4.615 ms
 9  xe0-3-0.0.border7-1.switchnap.com (66.209.64.146)  10.111 ms  10.094 ms  10.063 ms
10  GigabitEthernet6-1-0.GW3.VEG2.ALTER.NET (208.222.10.173)  10.031 ms  10.968 ms  10.926 ms
11  0.xe-1-2-0.XL4.VEG2.ALTER.NET (152.63.113.174)  9.910 ms  10.122 ms  10.094 ms
12  0.so-5-0-0.XT4.STL3.ALTER.NET (152.63.4.253)  55.446 ms  55.666 ms  55.654 ms
13  POS7-0.GW8.STL3.ALTER.NET (152.63.92.41)  55.581 ms  55.553 ms  55.521 ms
14  ibm-gw.customer.alter.net (65.206.180.74)  56.609 ms  57.181 ms  56.651 ms
15  * * *
16  * * *
17  * * *
```

znetconf

- show unconfigured devices

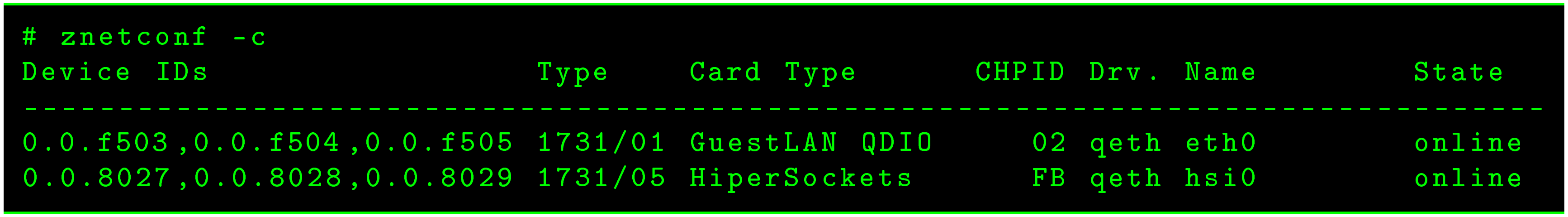
  `znetconf -u`



```
# znetconf -u
Scanning for network devices...
Device IDs                 Type    Card Type      CHPID Drv.
------------------------------------------------------------
0.0.f500,0.0.f501,0.0.f502 1731/01 OSA (QDIO)        01 qeth
0.0.f5f0,0.0.f5f1,0.0.f5f2 1731/01 OSA (QDIO)        76 qeth
```

- show configures devices

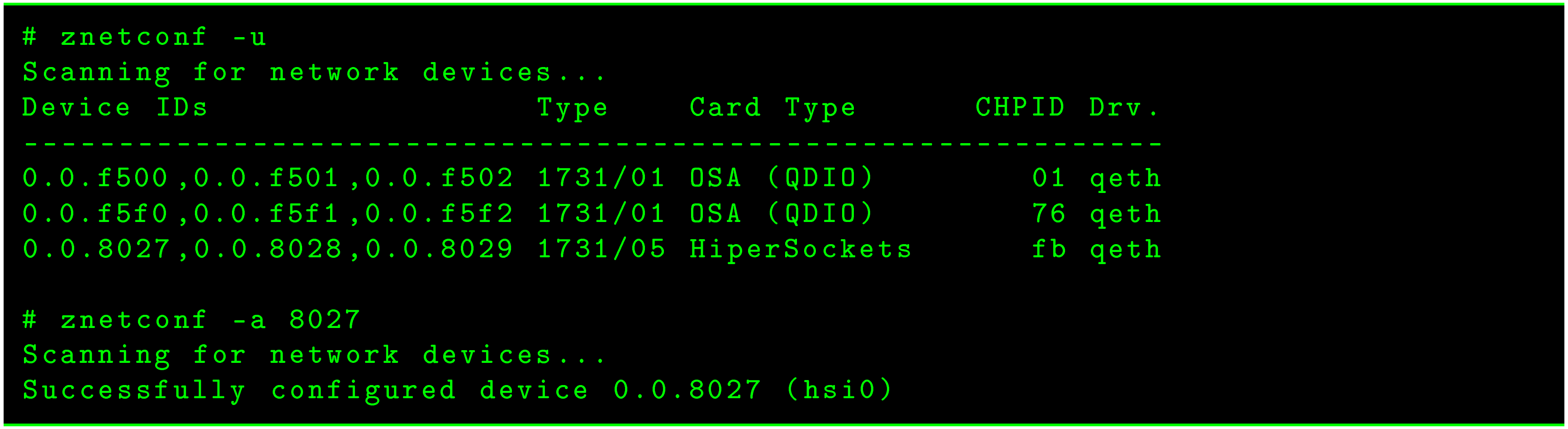
  `znetconf -c`



```
# znetconf -c
Device IDs                 Type    Card Type      CHPID Drv. Name        State
-------------------------------------------------------------------------------
0.0.f503,0.0.f504,0.0.f505 1731/01 GuestLAN QDIO     02 qeth eth0        online
0.0.8027,0.0.8028,0.0.8029 1731/05 HiperSockets      FB qeth hsi0        online
```

znetconf

- add new device
  `znetconf -a`



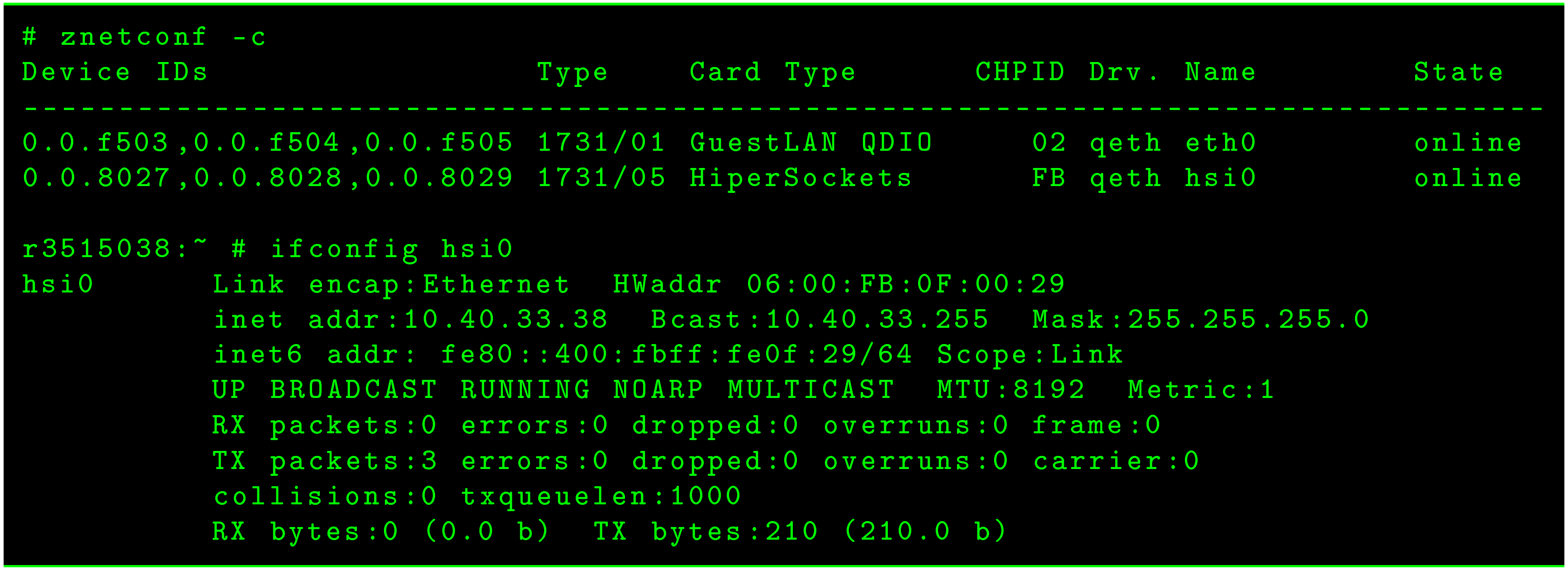
```
# znetconf -u
Scanning for network devices...
Device IDs                 Type    Card Type     CHPID Drv.
------------------------------------------------------------
0.0.f500,0.0.f501,0.0.f502 1731/01 OSA (QDIO)        01 qeth
0.0.f5f0,0.0.f5f1,0.0.f5f2 1731/01 OSA (QDIO)        76 qeth
0.0.8027,0.0.8028,0.0.8029 1731/05 HiperSockets      fb qeth

# znetconf -a 8027
Scanning for network devices...
Successfully configured device 0.0.8027 (hsi0)
```

znetconf

- add new device

  `znetconf -c`



```
# znetconf -c
Device IDs                 Type    Card Type      CHPID Drv. Name        State
--------------------------------------------------------------------------------
0.0.f503,0.0.f504,0.0.f505 1731/01 GuestLAN QDIO     02 qeth eth0        online
0.0.8027,0.0.8028,0.0.8029 1731/05 HiperSockets      FB qeth hsi0        online

r3515038:~ # ifconfig hsi0
hsi0      Link encap:Ethernet  HWaddr 06:00:FB:0F:00:29
          inet addr:10.40.33.38  Bcast:10.40.33.255  Mask:255.255.255.0
          inet6 addr: fe80::400:fbff:fe0f:29/64 Scope:Link
          UP BROADCAST RUNNING NOARP MULTICAST  MTU:8192  Metric:1
          RX packets:0 errors:0 dropped:0 overruns:0 frame:0
          TX packets:3 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:0 (0.0 b)  TX bytes:210 (210.0 b)
```

znetconf

- remove device
  `znetconf -r`

```
# znetconf -r 8027
Remove network device 0.0.8027 (0.0.8027,0.0.8028,0.0.8029)?
Warning: this may affect network connectivity!
Do you want to continue (y/n)?y
Successfully removed device 0.0.8027 (hsi0)
```

qetharp

- read Addresses from QDIO network

```
# qetharp -q hsi0
Address                                   HWaddress            HWType     Iface
10.40.33.38                                                    hiper      hsi0
10.40.33.39                                                    hiper      hsi0
10.40.33.52                                                    hiper      hsi0
10.40.39.2                                                     hiper      hsi0
10.40.39.3                                                     hiper      hsi0
10.40.30.10                                                    hiper      hsi0
```

tcpdump

- tcpdump -i

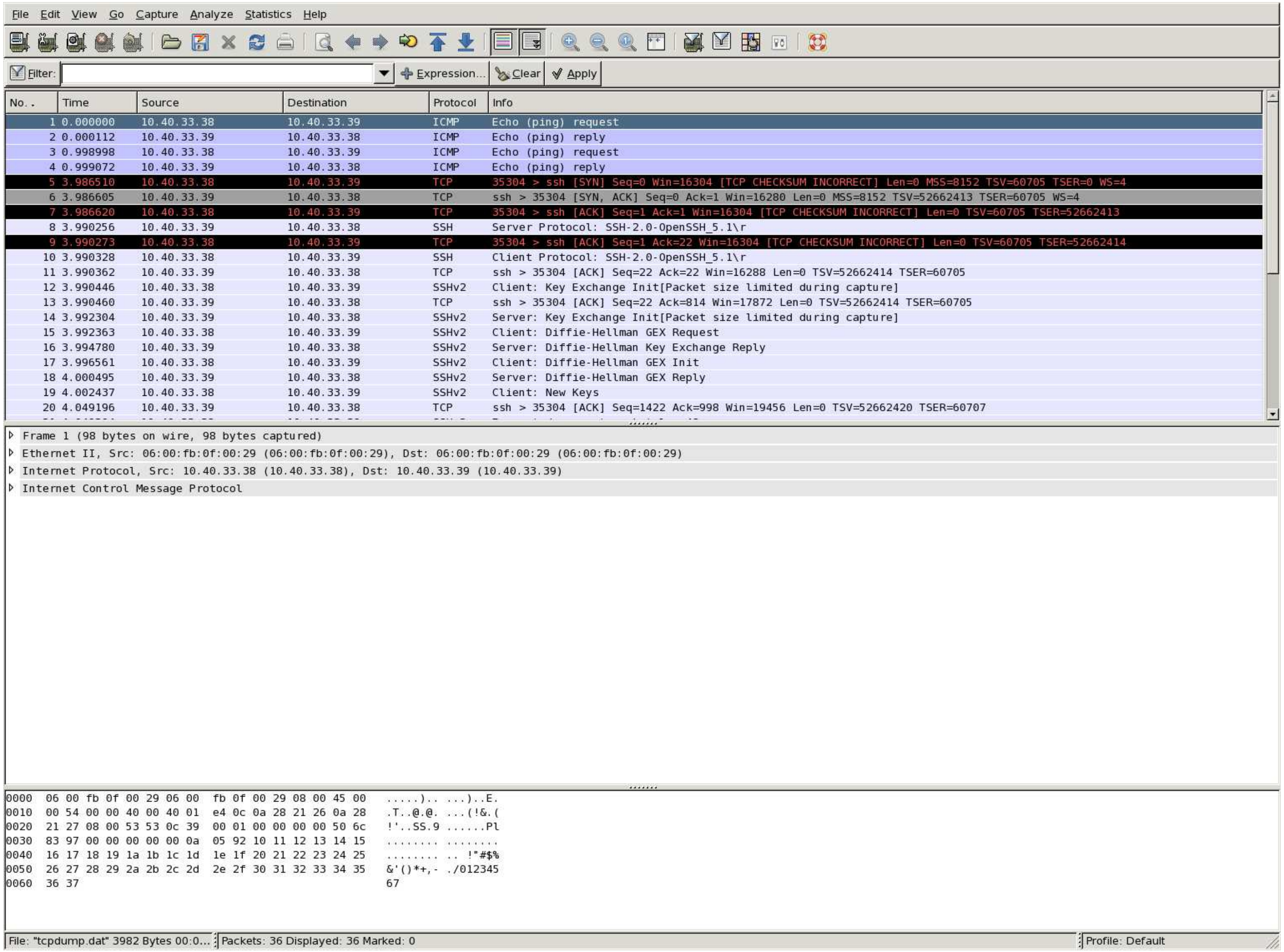
```
# tcpdump -i hsi0
tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
listening on hsi0, link-type EN10MB (Ethernet), capture size 96 bytes
20:22:50.131139 IP 10.40.33.38 > 10.40.33.39: ICMP echo request, id 2939, seq 1, lengt
20:22:50.131260 IP 10.40.33.39 > 10.40.33.38: ICMP echo reply, id 2939, seq 1, length
20:22:51.130135 IP 10.40.33.38 > 10.40.33.39: ICMP echo request, id 2939, seq 2, lengt
20:22:51.130192 IP 10.40.33.39 > 10.40.33.38: ICMP echo reply, id 2939, seq 2, length
^C
4 packets captured
4 packets received by filter
0 packets dropped by kernel
```

tcpdump

- `tcpdump -w`
  - saves dump in binary format
  - can be analyzed using wireshark
  - reduces size - good when sent to IBM support
  - please tell us how many packets were dropped

```
# tcpdump -i hsi0 -w tcpdump.dat
tcpdump: listening on hsi0, link-type EN10MB (Ethernet), capture size 96 bytes
^C47 packets captured
47 packets received by filter
0 packets dropped by kernel
```

wireshark



```
No. .  Time      Source       Destination  Protocol  Info
    1 0.000000   10.40.33.38  10.40.33.39  ICMP      Echo (ping) request
    2 0.000112   10.40.33.39  10.40.33.38  ICMP      Echo (ping) reply
    3 0.998998   10.40.33.38  10.40.33.39  ICMP      Echo (ping) request
    4 0.999072   10.40.33.39  10.40.33.38  ICMP      Echo (ping) reply
    5 3.986510   10.40.33.38  10.40.33.39  TCP       35304 > ssh [SYN] Seq=0 Win=16304 [TCP CHECKSUM INCORRECT] Len=0 MSS=8152 TSV=60705 TSER=0 WS=4
    6 3.986605   10.40.33.39  10.40.33.38  TCP       ssh > 35304 [SYN, ACK] Seq=0 Ack=1 Win=16280 Len=0 MSS=8152 TSV=52662413 TSER=60705 WS=4
    7 3.986620   10.40.33.38  10.40.33.39  TCP       35304 > ssh [ACK] Seq=1 Ack=1 Win=16304 [TCP CHECKSUM INCORRECT] Len=0 TSV=60705 TSER=52662413
    8 3.990256   10.40.33.39  10.40.33.38  SSH       Server Protocol: SSH-2.0-OpenSSH_5.1\r
    9 3.990273   10.40.33.38  10.40.33.39  TCP       35304 > ssh [ACK] Seq=1 Ack=22 Win=16304 [TCP CHECKSUM INCORRECT] Len=0 TSV=60705 TSER=52662414
   10 3.990328   10.40.33.38  10.40.33.39  SSH       Client Protocol: SSH-2.0-OpenSSH_5.1\r
   11 3.990362   10.40.33.39  10.40.33.38  TCP       ssh > 35304 [ACK] Seq=22 Ack=22 Win=16288 Len=0 TSV=52662414 TSER=60705
   12 3.990446   10.40.33.38  10.40.33.39  SSHv2     Client: Key Exchange Init[Packet size limited during capture]
   13 3.990460   10.40.33.39  10.40.33.38  TCP       ssh > 35304 [ACK] Seq=22 Ack=814 Win=17872 Len=0 TSV=52662414 TSER=60705
   14 3.992304   10.40.33.39  10.40.33.38  SSHv2     Server: Key Exchange Init[Packet size limited during capture]
   15 3.992363   10.40.33.38  10.40.33.39  SSHv2     Client: Diffie-Hellman GEX Request
   16 3.994780   10.40.33.39  10.40.33.38  SSHv2     Server: Diffie-Hellman Key Exchange Reply
   17 3.996561   10.40.33.38  10.40.33.39  SSHv2     Client: Diffie-Hellman GEX Init
   18 4.000495   10.40.33.39  10.40.33.38  SSHv2     Server: Diffie-Hellman GEX Reply
   19 4.002437   10.40.33.38  10.40.33.39  SSHv2     Client: New Keys
   20 4.049196   10.40.33.39  10.40.33.38  TCP       ssh > 35304 [ACK] Seq=1422 Ack=998 Win=19456 Len=0 TSV=52662420 TSER=60707
```

```
▷ Frame 1 (98 bytes on wire, 98 bytes captured)
▷ Ethernet II, Src: 06:00:fb:0f:00:29 (06:00:fb:0f:00:29), Dst: 06:00:fb:0f:00:29 (06:00:fb:0f:00:29)
▷ Internet Protocol, Src: 10.40.33.38 (10.40.33.38), Dst: 10.40.33.39 (10.40.33.39)
▷ Internet Control Message Protocol
```

```
0000  06 00 fb 0f 00 29 06 00  fb 0f 00 29 08 00 45 00   .....)..  ...)..E.
0010  00 54 00 00 40 00 40 01  e4 0c 0a 28 21 26 0a 28   .T..@.@.  ...(!&.(
0020  21 27 08 00 53 53 0c 39  00 01 00 00 00 00 50 6c   !'..SS.9  ......Pl
0030  83 97 00 00 00 00 00 0a  05 92 10 11 12 13 14 15   ........  ........
0040  16 17 18 19 1a 1b 1c 1d  1e 1f 20 21 22 23 24 25   ........  .. !"#$%
0050  26 27 28 29 2a 2b 2c 2d  2e 2f 30 31 32 33 34 35   &'()*+,-  ./012345
0060  36 37                                              67
```

File: "tcpdump.dat" 3982 Bytes 00:0... | Packets: 36 Displayed: 36 Marked: 0 | Profile: Default

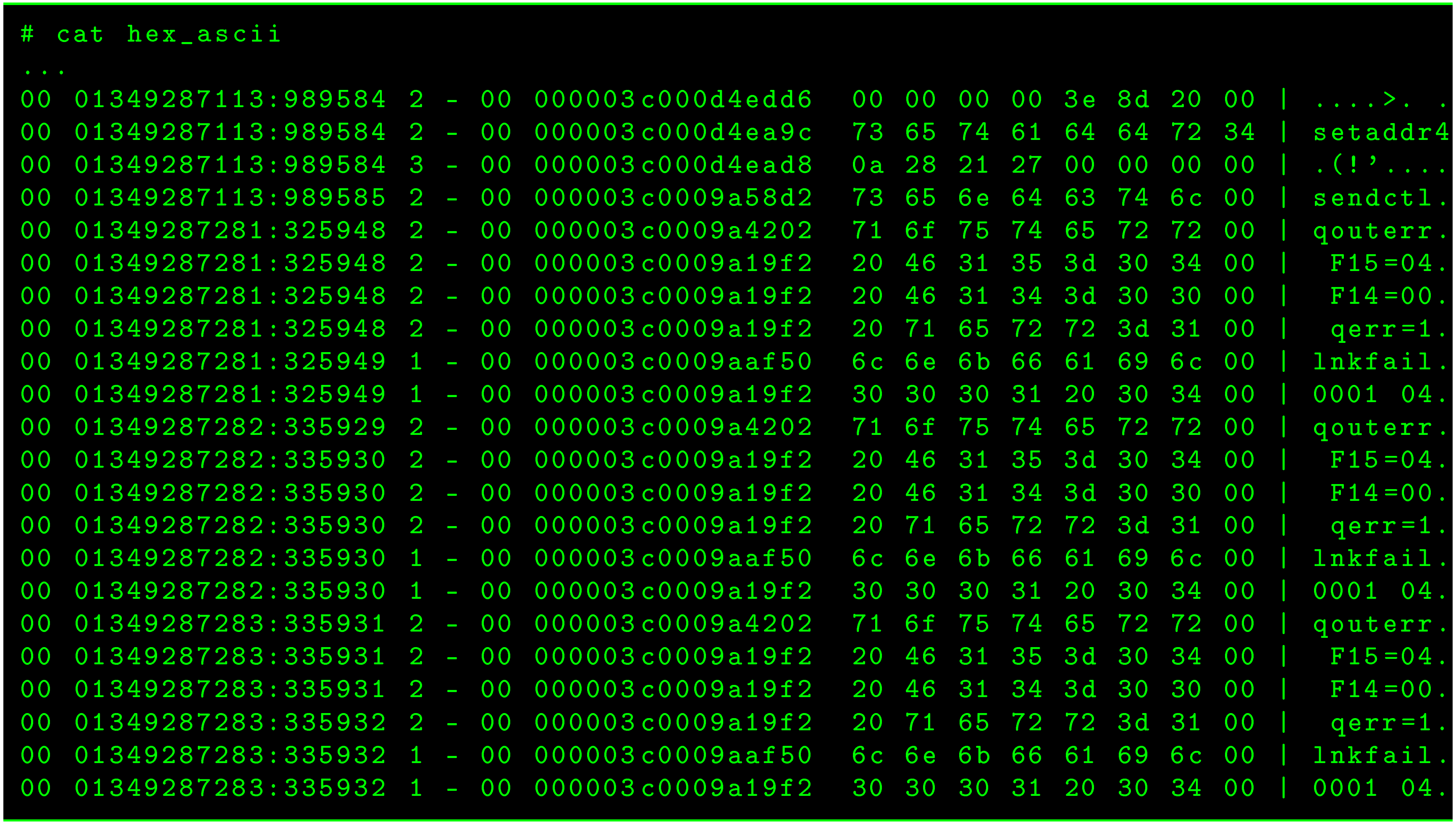
Problem Determination

- QETH errors
- QDIO statistics

QETH Errors

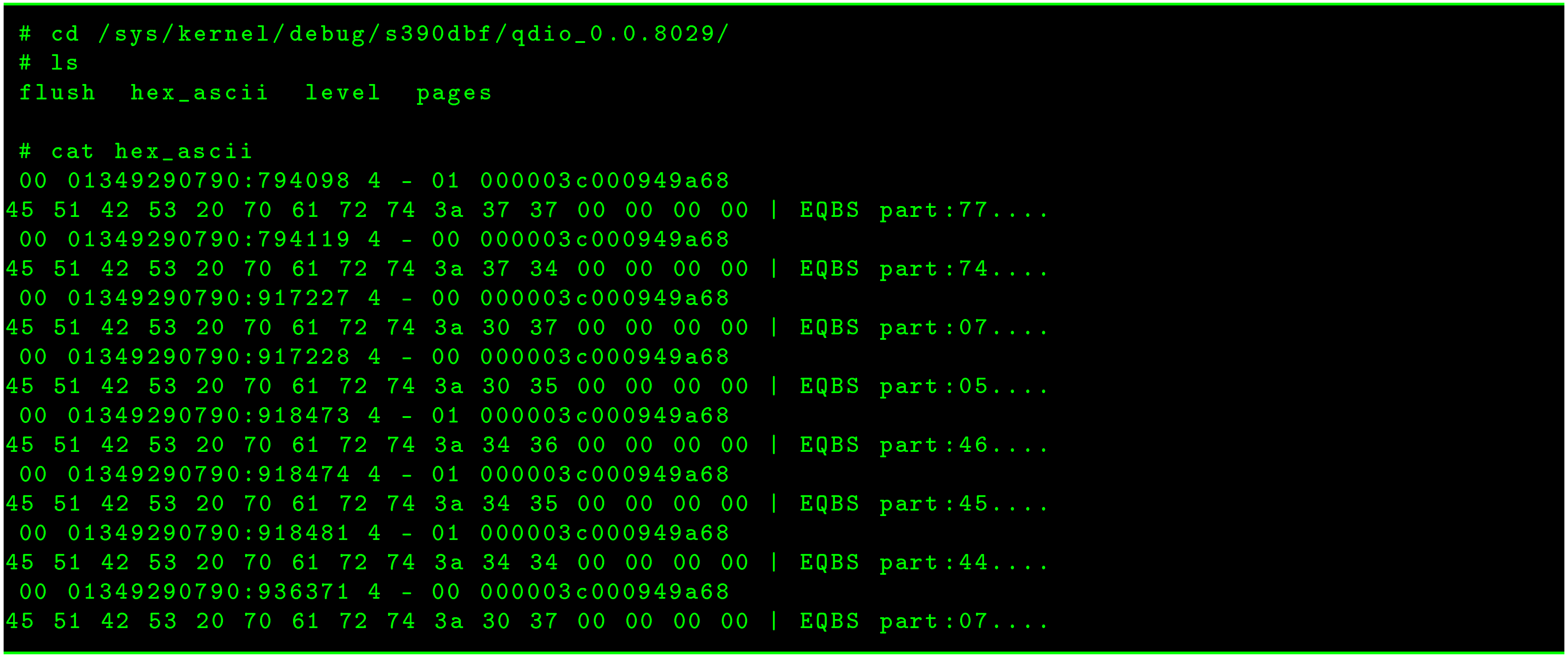
```
# cd /sys/kernel/debug/s390dbf/qeth_card_0.0.8027/
# ls
flush  hex_ascii  level  pages
```

QETH Errors



```
# cat hex_ascii
...
00 01349287113:989584 2 - 00 000003c000d4edd6  00 00 00 00 3e 8d 20 00 | ....>. .
00 01349287113:989584 2 - 00 000003c000d4ea9c  73 65 74 61 64 64 72 34 | setaddr4
00 01349287113:989584 3 - 00 000003c000d4ead8  0a 28 21 27 00 00 00 00 | .(!'....
00 01349287113:989585 2 - 00 000003c0009a58d2  73 65 6e 64 63 74 6c 00 | sendctl.
00 01349287281:325948 2 - 00 000003c0009a4202  71 6f 75 74 65 72 72 00 | qouterr.
00 01349287281:325948 2 - 00 000003c0009a19f2  20 46 31 35 3d 30 34 00 |  F15=04.
00 01349287281:325948 2 - 00 000003c0009a19f2  20 46 31 34 3d 30 30 00 |  F14=00.
00 01349287281:325948 2 - 00 000003c0009a19f2  20 71 65 72 72 3d 31 00 |  qerr=1.
00 01349287281:325949 1 - 00 000003c0009aaf50  6c 6e 6b 66 61 69 6c 00 | lnkfail.
00 01349287281:325949 1 - 00 000003c0009a19f2  30 30 30 31 20 30 34 00 | 0001 04.
00 01349287282:335929 2 - 00 000003c0009a4202  71 6f 75 74 65 72 72 00 | qouterr.
00 01349287282:335930 2 - 00 000003c0009a19f2  20 46 31 35 3d 30 34 00 |  F15=04.
00 01349287282:335930 2 - 00 000003c0009a19f2  20 46 31 34 3d 30 30 00 |  F14=00.
00 01349287282:335930 2 - 00 000003c0009a19f2  20 71 65 72 72 3d 31 00 |  qerr=1.
00 01349287282:335930 1 - 00 000003c0009aaf50  6c 6e 6b 66 61 69 6c 00 | lnkfail.
00 01349287282:335930 1 - 00 000003c0009a19f2  30 30 30 31 20 30 34 00 | 0001 04.
00 01349287283:335931 2 - 00 000003c0009a4202  71 6f 75 74 65 72 72 00 | qouterr.
00 01349287283:335931 2 - 00 000003c0009a19f2  20 46 31 35 3d 30 34 00 |  F15=04.
00 01349287283:335931 2 - 00 000003c0009a19f2  20 46 31 34 3d 30 30 00 |  F14=00.
00 01349287283:335932 2 - 00 000003c0009a19f2  20 71 65 72 72 3d 31 00 |  qerr=1.
00 01349287283:335932 1 - 00 000003c0009aaf50  6c 6e 6b 66 61 69 6c 00 | lnkfail.
00 01349287283:335932 1 - 00 000003c0009a19f2  30 30 30 31 20 30 34 00 | 0001 04.
```

QDIO Errors



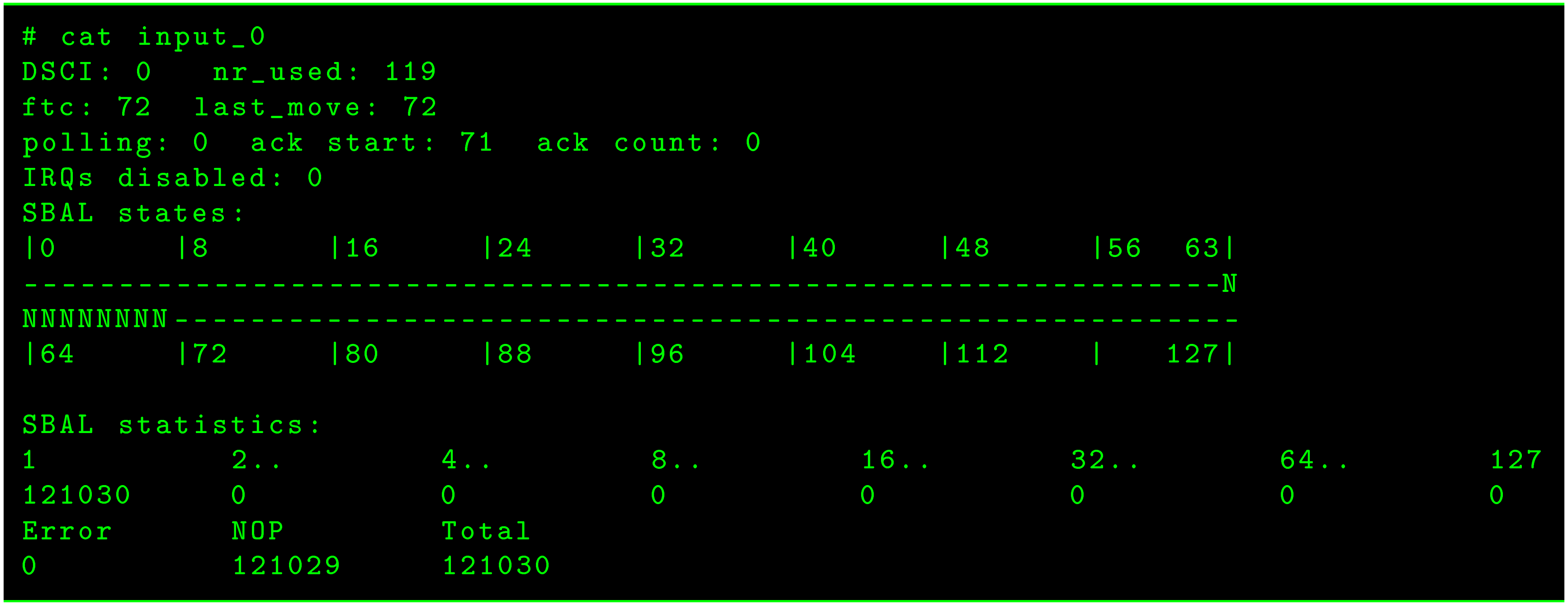
```
# cd /sys/kernel/debug/s390dbf/qdio_0.0.8029/
# ls
flush  hex_ascii  level  pages

# cat hex_ascii
 00 01349290790:794098 4 - 01 000003c000949a68
45 51 42 53 20 70 61 72 74 3a 37 37 00 00 00 00 | EQBS part:77....
 00 01349290790:794119 4 - 00 000003c000949a68
45 51 42 53 20 70 61 72 74 3a 37 34 00 00 00 00 | EQBS part:74....
 00 01349290790:917227 4 - 00 000003c000949a68
45 51 42 53 20 70 61 72 74 3a 30 37 00 00 00 00 | EQBS part:07....
 00 01349290790:917228 4 - 00 000003c000949a68
45 51 42 53 20 70 61 72 74 3a 30 35 00 00 00 00 | EQBS part:05....
 00 01349290790:918473 4 - 01 000003c000949a68
45 51 42 53 20 70 61 72 74 3a 34 36 00 00 00 00 | EQBS part:46....
 00 01349290790:918474 4 - 01 000003c000949a68
45 51 42 53 20 70 61 72 74 3a 34 35 00 00 00 00 | EQBS part:45....
 00 01349290790:918481 4 - 01 000003c000949a68
45 51 42 53 20 70 61 72 74 3a 34 34 00 00 00 00 | EQBS part:44....
 00 01349290790:936371 4 - 00 000003c000949a68
45 51 42 53 20 70 61 72 74 3a 30 37 00 00 00 00 | EQBS part:07....
```

Problem Determination

```
# cd /sys/kernel/debug/qdio/0.0.8029
# echo 1 > statistics
```

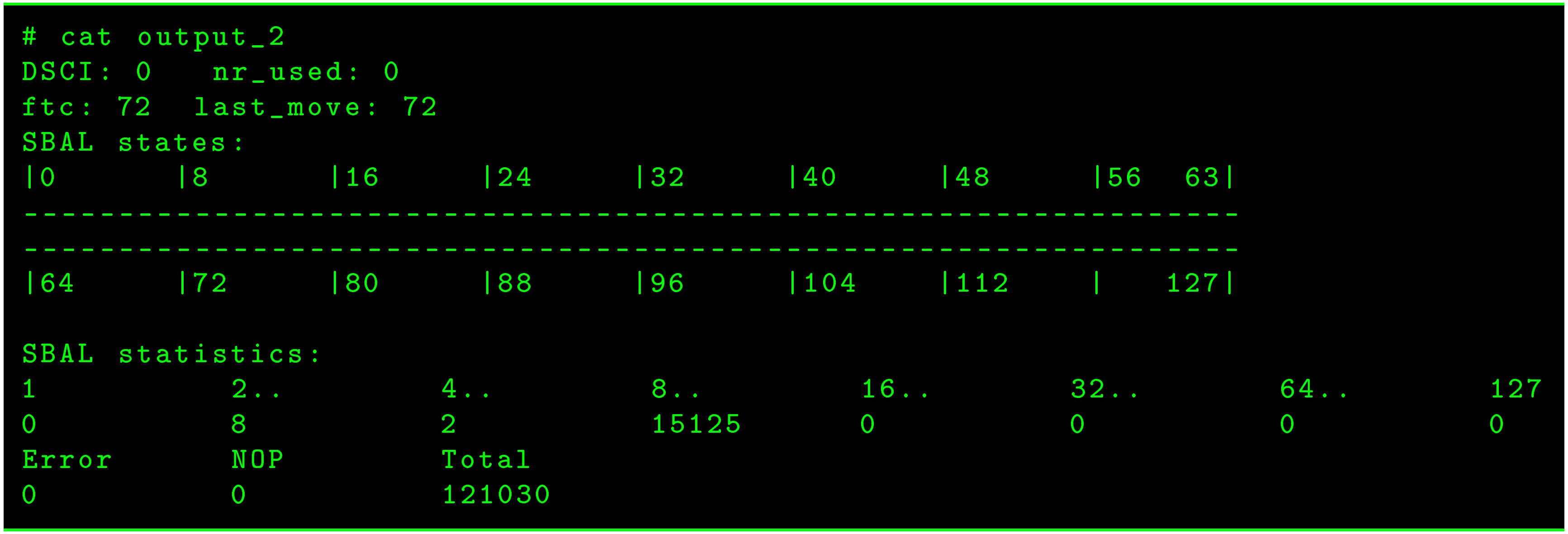
Problem Determination



```
# cat input_0
DSCI: 0   nr_used: 119
ftc: 72  last_move: 72
polling: 0  ack start: 71  ack count: 0
IRQs disabled: 0
SBAL states:
|0      |8      |16     |24     |32     |40     |48     |56  63|
---------------------------------------------------------------N
NNNNNNNN--------------------------------------------------------
|64     |72     |80     |88     |96     |104    |112    |   127|

SBAL statistics:
1          2..        4..        8..        16..       32..       64..       127
121030     0          0          0          0          0          0          0
Error      NOP        Total
0          121029     121030
```

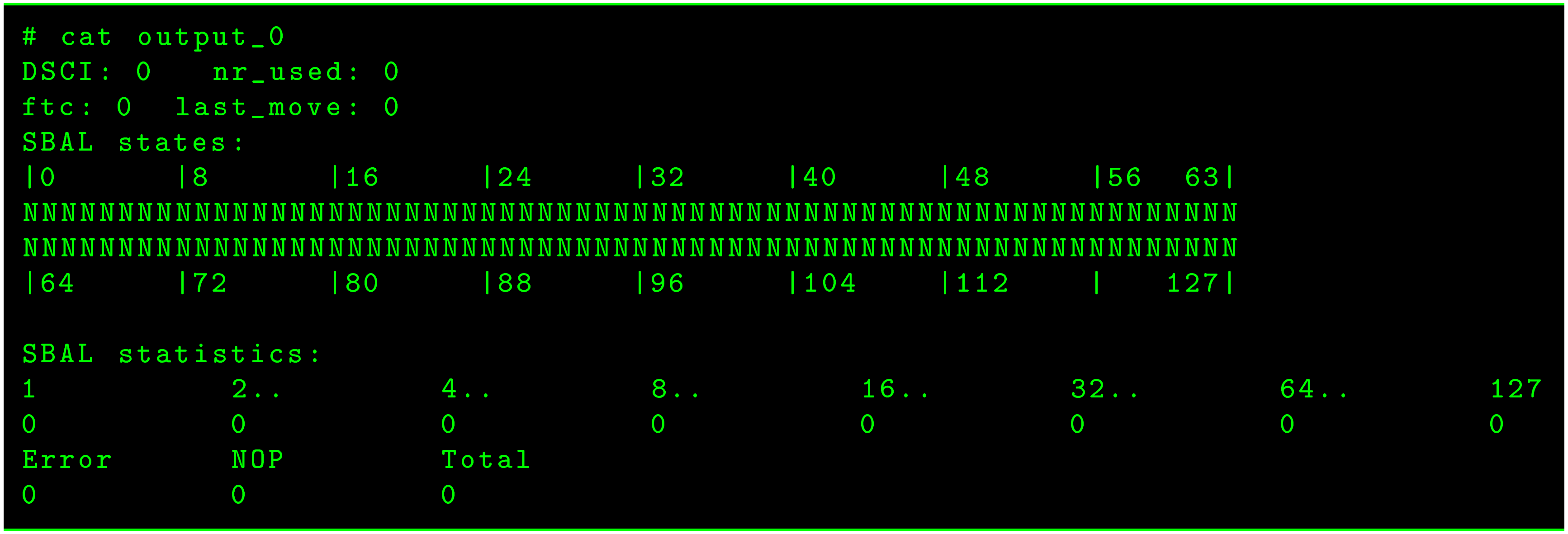
Problem Determination



```
# cat output_2
DSCI: 0   nr_used: 0
ftc: 72  last_move: 72
SBAL states:
|0       |8       |16      |24      |32      |40      |48      |56  63|
----------------------------------------------------------------
----------------------------------------------------------------
|64      |72      |80      |88      |96      |104     |112     |   127|

SBAL statistics:
1          2..        4..        8..        16..       32..       64..       127
0          8          2          15125      0          0          0          0
Error      NOP        Total
0          0          121030
```

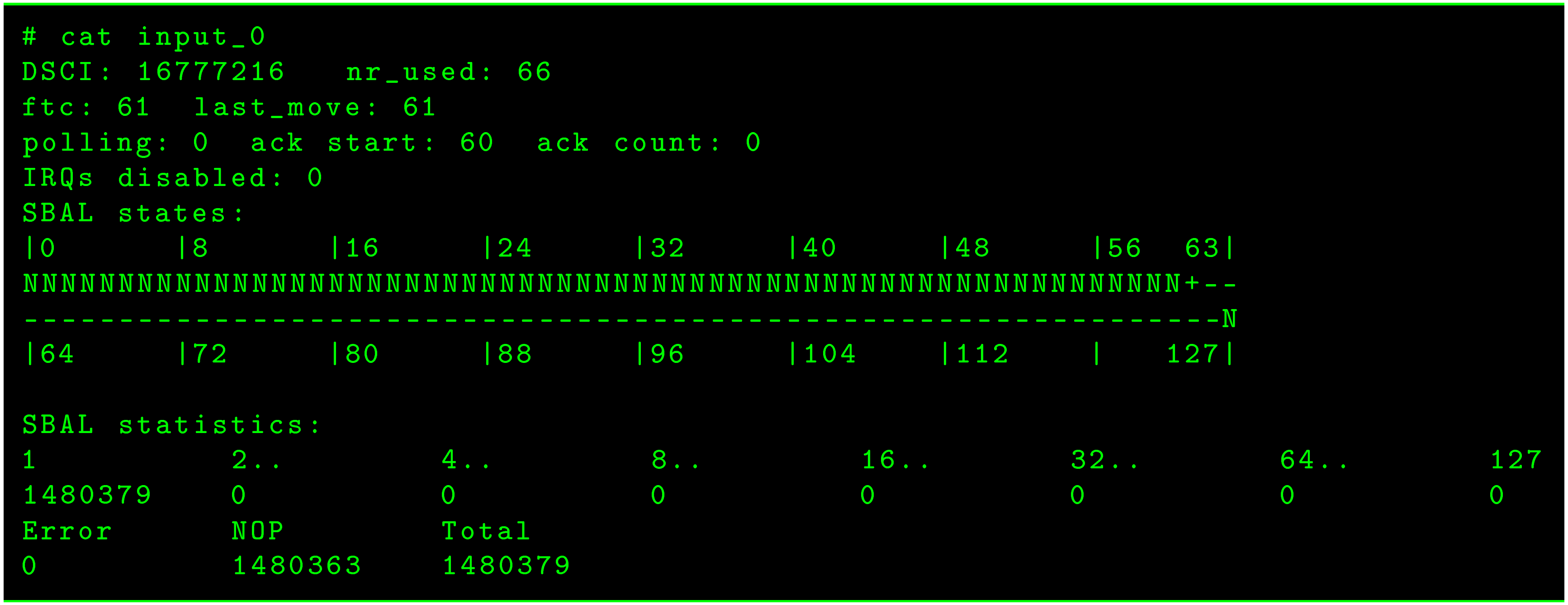
Problem Determination



```
# cat output_0
DSCI: 0   nr_used: 0
ftc: 0  last_move: 0
SBAL states:
|0       |8       |16      |24      |32      |40      |48      |56   63|
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
|64      |72      |80      |88      |96      |104     |112     |    127|

SBAL statistics:
1          2..        4..        8..        16..       32..       64..       127
0          0          0          0          0          0          0          0
Error      NOP        Total
0          0          0
```

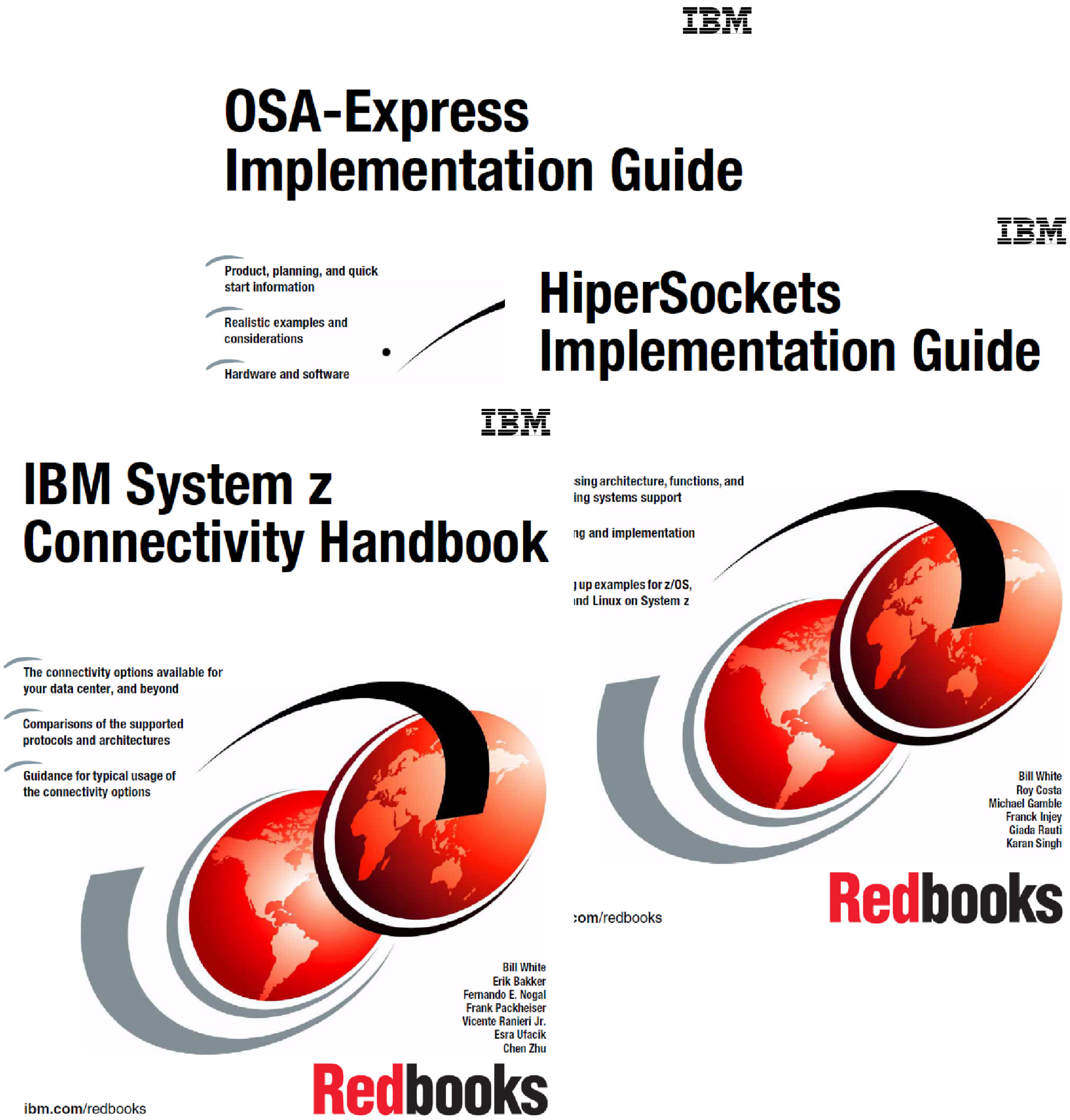
Problem Determination



```
# cat input_0
DSCI: 16777216   nr_used: 66
ftc: 61  last_move: 61
polling: 0  ack start: 60  ack count: 0
IRQs disabled: 0
SBAL states:
|0      |8      |16     |24     |32     |40     |48     |56  63|
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN+--
---------------------------------------------------------------N
|64     |72     |80     |88     |96     |104    |112    |   127|

SBAL statistics:
1          2..        4..        8..        16..       32..       64..       127
1480379    0          0          0          0          0          0          0
Error      NOP        Total
0          1480363    1480379
```

Redbooks



IBM

OSA-Express
Implementation Guide

Product, planning, and quick start information

Realistic examples and considerations

Hardware and software

IBM

HiperSockets
Implementation Guide

sing architecture, functions, and
ing systems support

ng and implementation

up examples for z/OS,
nd Linux on System z

Bill White
Roy Costa
Michael Gamble
Franck Injey
Giada Rauti
Karan Singh

com/redbooks

Redbooks

IBM

IBM System z
Connectivity Handbook

The connectivity options available for your data center, and beyond

Comparisons of the supported protocols and architectures

Guidance for typical usage of the connectivity options

Bill White
Erik Bakker
Fernando E. Nogal
Frank Packheiser
Vicente Ranieri Jr.
Esra Ufacik
Chen Zhu

ibm.com/redbooks

Redbooks

Links

- Linux on System z - Tuning Hints & Tips
  `http://www.ibm.com/developerworks/linux/linux390/perf/index.html`
- developerWorks
  `http://www.ibm.com/developerworks/linux/linux390`
- IBM Redbooks
  `http://www.redbooks.ibm.com`

Thank You !

For starting out with their very good presentations

- Susanne Wintenberger
- Mario Held

Questions ?

Dr. Stefan Reimbold
Diplom-Physiker

Linux on System z Service

Schoenaicher Strasse 220
D-71032 Boeblingen
Mail: Postfach 1380
D-71003 Boeblingen

Phone +49-7031-16-2368
Stefan.Reimbold@de.ibm.com