Lab 2 — send and receive packets with D...

Handout: March 9, 2021
Deadline: March 24 23:00, 2021 (No extension)

Assignment overview:

In this assignment, you will learn what DPDK is and what it's used for, you will learn how to set up a DPDK development environment and understand the implementation of DPDK modules and programs. In part 1, you are asked to answer some questions about DPDK. In part 2, you need to write a DPDK application to construct UDP packets and send them to NIC using DPDK.

Prerequisite:

In this lab, DPDK must be installed in your virtual machine. VMware Workstation and Ubuntu 20.04 are suggested.

The steps for building DPDK development environment are as follows:

1. Get Source Code

```
$ git clone git://dpdk.org/dpdk # Get DPDK
$ git clone http://dpdk.org/git/dpdk-kmods # Get igb_uio
$ cp -r ./dpdk-kmods/linux/igb_uio ./dpdk/kernel/linux/ # Copy dpdk-kmods/linux/igb_uio/ to dpdk/kernel/linux/
```

- add igb_uio in dpdk/kernel/linux/meson.build subdirs as below:

```
subdirs = ['kni', 'igb_uio']
```

- create a file of meson.build in dpdk/kernel/linux/igb_uio/ as below:

```
# SPDX-License-Identifier: BSD-3-Clause
# Copyright(c) 2017 Intel Corporation
mkfile = custom_target('igb_uio_makefile',
        output: 'Makefile',
        command: ['touch', '@OUTPUT@'])
custom_target('igb_uio',
        input: ['igb_uio.c', 'Kbuild'],
        output: 'igb_uio.ko',
        command: ['make', '-C', kernel_dir + '/build',
```

```
                'M=' + meson.current_build_dir(),
                'src=' + meson.current_source_dir(),
                'EXTRA_CFLAGS=-I' + meson.current_source_dir() +
                        '/../../../lib/librte_eal/include',
                'modules'],
        depends: mkfile,
        install: true,
        install_dir: kernel_dir + '/extra/dpdk',
        build_by_default: get_option('enable_kmods'))
```

2. Configure and build DPDK runtime environment

```
$ sudo apt-get install python3
$ sudo apt-get install python3-pip
$ sudo pip3 install meson
$ sudo pip3 install ninja
$ sudo pip3 install pwntools
$ cd dpdk # 进入DPDK文件夹
$ sudo meson -D examples=all build # 使用选项 -D examples 指定编译所有样例程序
$ cd build
$ sudo ninja install
#以下命令每次开机都要执行
$ mkdir -p /dev/hugepages # 创建 hugetlbfs 挂载点，每次开机都要执行
$ mountpoint -q /dev/hugepages || mount -t hugetlbfs nodev /dev/hugepages # 挂载 hugetlbfs，每次开机都要执行
$ echo 64 > /sys/devices/system/node/node0/hugepages/hugepages-2048kB/nr_hugepages # 分配大页，每次开机都要执行
```

3. Run the sample program to make sure you have successfully installed DPDK

```
$ cd examples # 进入 DPDK 中的例子
$ sudo ./dpdk-helloworld # 运行
```

For more information about installing DPDK,refer to the guidance (http://doc.dpdk.org/guides/linux_gsg/).

Part 1: Get familiar with DPDK

Take DPDK official website http://www.dpdk.org/ and Chapter 1 of **”深入浅出DPDK”** as references, answer the questions below:

Q1: What’s the purpose of using hugepage?
Q2: Take examples/helloworld as an example, describe the execution flow of DPDK programs?
Q3: Read the codes of examples/skeleton, describe DPDK APIs related to sending and receiving packets.
Q4: Describe the data structure of ’rte_mbuf’.

Part 2: send packets with DPDK

1.Construct UDP packets with DPDK according to the definition of UDP/IP/Ethernet header. Contents of packets can be filled as you wish.

(Tips: refer to codes of examples/skeleton and construct UDP packets in the array of rte_mbuf, you may use the function rte_pktmbuf_mtod https://doc.dpdk.org/api/rte__mbuf_8h.html#a2a8b10263496c7b580e9d0c7f2a1f073)

```
/* Ethernet 协议头的定义在 lib/librte_net/rte_ether.h 中： */
/**
 * Ethernet header: Contains the destination address, source address
 * and frame type.
 */
struct rte_ether_hdr {
        struct rte_ether_addr d_addr; /**< Destination address. */
        struct rte_ether_addr s_addr; /**< Source address. */
        uint16_t ether_type;      /**< Frame type. */
} __rte_aligned(2);
/* IP 协议头的定义在 lib/librte_net/rte_ip.h 中： */
/**
 * IPv4 Header
 */
struct rte_ipv4_hdr {
        uint8_t  version_ihl;           /**< version and header length */
        uint8_t  type_of_service;       /**< type of service */
        rte_be16_t total_length;        /**< length of packet */
        rte_be16_t packet_id;           /**< packet ID */
        rte_be16_t fragment_offset;     /**< fragmentation offset */
        uint8_t  time_to_live;          /**< time to live */
        uint8_t  next_proto_id;         /**< protocol ID */
        rte_be16_t hdr_checksum;        /**< header checksum */
        rte_be32_t src_addr;            /**< source address */
        rte_be32_t dst_addr;            /**< destination address */
} __rte_packed;
/* UDP 协议头的定义在 lib/librte_net/rte_udp.h 中： */
/**
 * UDP Header
 */
struct rte_udp_hdr {
        rte_be16_t src_port;    /**< UDP source port. */
        rte_be16_t dst_port;    /**< UDP destination port. */
        rte_be16_t dgram_len;   /**< UDP datagram length */
        rte_be16_t dgram_cksum; /**< UDP datagram checksum */
} __rte_packed;
```

2. Write a DPDK application to construct and send UDP packets. (Refer to examples/skeleton). Before writing any code, you need to enable the virtual NIC in your

virtual machine. If you are using VMware, refer to this website to enable a host-only mode virtual NIC: https://blog.csdn.net/ka_ka314/article/details/78936105

After virtual NIC is enabled, use following command to bind it to DPDK:

```
#以下命令每次开机都要执行
sudo ifconfig ens33 down #停止 ens33 网卡的工作， ens33 为刚激活的 host-only 模式下的虚拟网卡，
#（ens33 可替换成其他网卡， 在绑定网卡到 DPDK 前需要使其从活跃状态变为不活跃状态）
sudo modprobe uio
cd ../../kernel/linux/igb_uio
make
sudo insmod igb_uio.ko
cd ../../..
sudo usertools/dpdk-devbind.py --bind=igb_uio ens33
sudo usertools/dpdk-devbind.py -s #验证虚拟网卡是否被成功绑定到 DPDK 上
```

For binding ports to DPDK modules, refer to the contents of chapter 5.4 of http://doc.dpdk.org/guides/linux_gsg/linux_drivers.html

3. To verify the correctness of your program, you need to capture the packets and display the contents of them. Wireshark is the world's foremost and widely-used network protocol analyzer. It can help to capture the packets sent to NIC. Install it in your physical machine (NOT in virtual machine) and test your program. https://www.wireshark.org/

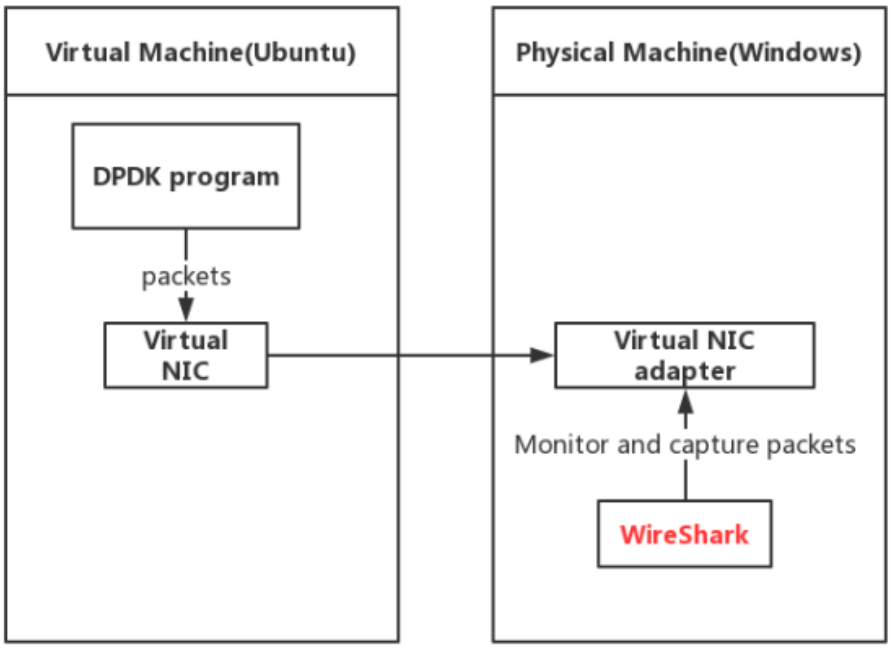


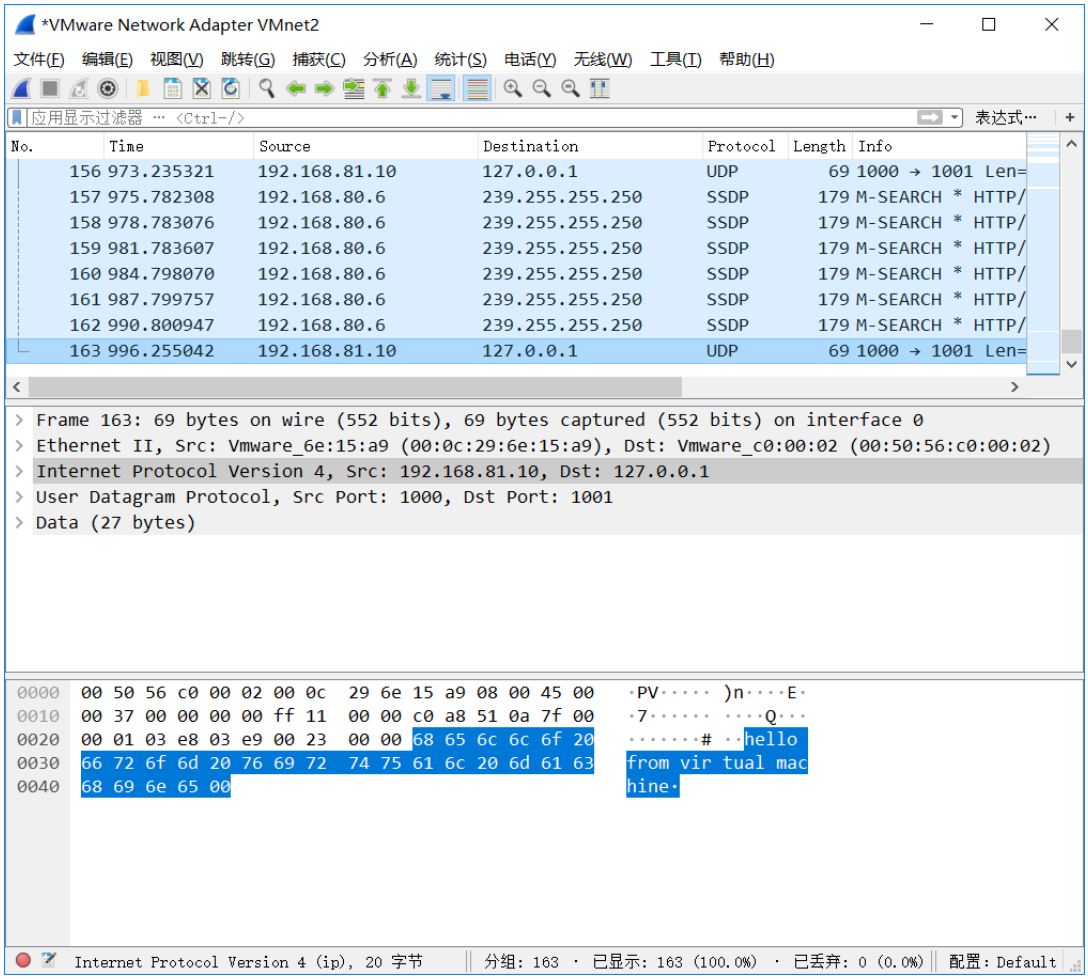
Diagram used to show how to test your program with Wireshark

Tips: If your enabled virtual NIC is in host-only mode, you need to capture packets from the same virtual NIC in your physical machine using wireshark. If your enabled virtual NIC is in bridgeconnection mode, you need to capture packets from the NIC your virtual NIC bridged to in your physical machine using wireshark. NAT mode is not advised.

Handin procedure:

You need to submit the source code of the DPDK application, and a README file including the answers to the questions in part 1 and a description of how you verify the correctness of the application with the screenshot of Wireshark in part 2.

The screenshot is supposed to look like this:



Send your report and source code as a gzipped tar file, named as {Your studentID}.tar.gz, to CANVAS

Grading:

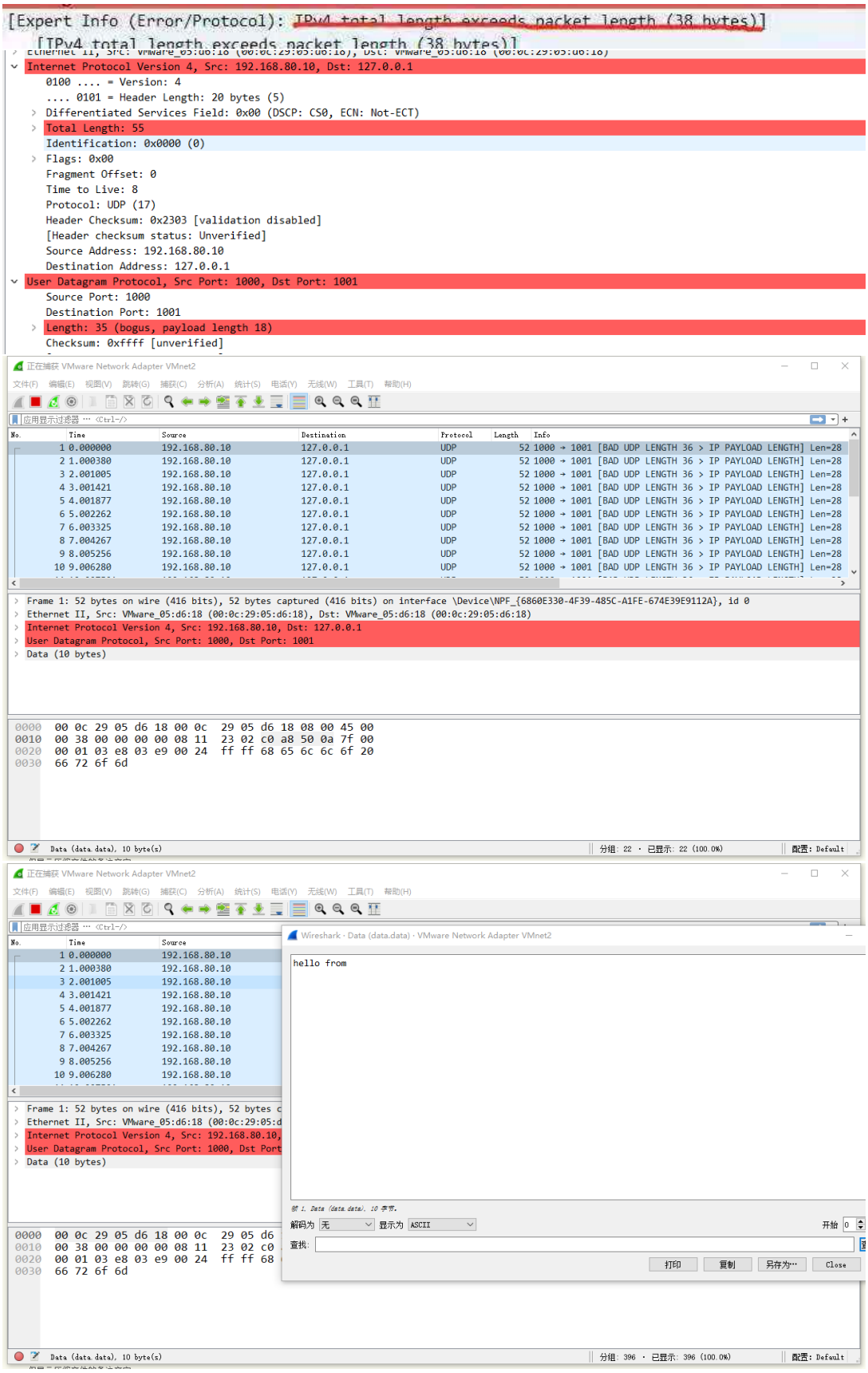
- 40% for your document totally. 20% for the answers to the questions in part 1, and another 20% for a description of how you verify the correctness of the application with the screenshot of Wireshark in part 2.
- 60% for the code of DPDK program. Code will be checked manually and results will be verified by wireshark.

Reference:

1. DPDK official website: http://www.dpdk.org/
2. Helloworld 例程 http://doc.dpdk.org/guides/sample_app_ug/hello_world.html
3. 深入浅出DPDK https://book.douban.com/subject/26798275/
4. DPDK APIs http://doc.dpdk.org/api/

FAQ:

Q1: Error when display packets captured



A1: Check your code to ensure allocate enough space for buffer

Q2: Helloworld cannot get hugepage information

```
za@ubuntu:~/Documents/env/dpdk/build/examples$ sudo ./dpdk-helloworld
EAL: Detected 2 lcore(s)
EAL: Detected 1 NUMA nodes
EAL: Detected static linkage of DPDK
EAL: Multi-process socket /var/run/dpdk/rte/mp_socket
EAL: Selected IOVA mode 'PA'
EAL: No free 2048 kB hugepages reported on node 0
EAL: No available 2048 kB hugepages reported
EAL: No available 1048576 kB hugepages reported
EAL: FATAL: Cannot get hugepage information.
EAL: Cannot get hugepage information.
PANIC in main():
Cannot init EAL
5: [./dpdk-helloworld(_start+0x2e) [0x55de97877c0e]]
4: [/lib/x86_64-linux-gnu/libc.so.6(__libc_start_main+0xf3) [0x7fee88b7d0b3]]
3: [./dpdk-helloworld(+0x10af6c) [0x55de97402f6c]]
2: [./dpdk-helloworld(__rte_panic+0xcd) [0x55de97418f48]]
1: [./dpdk-helloworld(rte_dump_stack+0x32) [0x55de979e31e2]]
Aborted
za@ubuntu:~/Documents/env/dpdk/build/examples$
```

A2: Hugepage not mounted

Q3: Basicfwd must be even error

```
os@ubuntu:~/dpdk/dpdk/examples/skeleton$ cd build
os@ubuntu:~/dpdk/dpdk/examples/skeleton/build$ ls
basicfwd  basicfwd-shared
os@ubuntu:~/dpdk/dpdk/examples/skeleton/build$ sudo ./basicfwd
EAL: Detected 2 lcore(s)
EAL: Detected 1 NUMA nodes
EAL: Detected shared linkage of DPDK
EAL: Multi-process socket /var/run/dpdk/rte/mp_socket
EAL: Selected IOVA mode 'PA'
EAL: No available 1048576 kB hugepages reported
EAL: Probing VFIO support...
EAL: VFIO support initialized
EAL:   Invalid NUMA socket, default to 0
EAL: Probe PCI driver: net_e1000_em (8086:100f) device: 0000:02:01.0 (socket 0)
EAL: Error reading from file descriptor 20: Input/output error
EAL: No legacy callbacks, legacy socket not created
EAL: Error - exiting with code: 1
  Cause: Error: number of ports must be even
os@ubuntu:~/dpdk/dpdk/examples/skeleton/build$
```

A3: Basicfwd needs even number of ports, refer to https://www.cnblogs.com/quinn-yann/p/9722138.html and https://www.wandouip.com/t5i136742/

Q4: How to set virtual NIC for mac

A4: Windows and VMware Workstation are suggested in this lab, https://my.oschina.net/u/3606160/blog/3015720 may be helpful for mac.

Q5: Can I organize my packet the way I like, since there is no spercific requirements

A5: Yes, but make sure no error when display, at least no red lines, situations like Q1 are not acceptable

Q6: I make a new directory under dpdk/examples, how to run it

A6: modify dpdk/examples/meson.build

Q7: Can I use existing hash for checksum

A7: Yes

Q8: Does dpdk provide hash function and network address get function

A8: rte_ipv4_cksum and rte_eth_macaddr_get

Q9: Do checksum for ipv4 and udp overlap

A9: No

Q10: There is no ens33 in my environment

A10: Make sure you are using the right VM

Q11: Error when executing dpdk-devbind.py

```
make[1]: 离开目录"/usr/src/linux-headers-5.8.0-44-generic"
alan@ubuntu:~/Desktop/dpdk/kernel/linux/igb_uio$ sudo insmod igb_uio.ko
alan@ubuntu:~/Desktop/dpdk/kernel/linux/igb_uio$ cd ../../..
alan@ubuntu:~/Desktop/dpdk$ sudo usertools/dpdk-devbind.py --bind=igb_uio ens33
alan@ubuntu:~/Desktop/dpdk$ sudo usertools/dpdk-devbind.py -s
Error: No action specified for devices. Please give a --bind, --ubind or --statu
s option
usage: dpdk-devbind.py [-h] [-s]
                       [--status-dev {baseband,compress,crypto,event,mempool,mis
c,net,regex}]
                       [-b DRIVER | -u] [--force]
                       [DEVICE [DEVICE ...]]
alan@ubuntu:~/Desktop/dpdk$ sudo ifconfig ens33 up
ens33: 获取接口标志时出错: 没有那个设备
alan@ubuntu:~/Desktop/dpdk$
```

A11: Make sure you are using English – rather than Chinese –

Q12: Git clone to slow

A12: Use http

Q13: VM lose network connection after virtual NIC setup

A13: This is normal, refer to https://jingyan.baidu.com/article/f79b7cb333361a9144023e0f.html if you need to add NIC

Q14: Still lose network connection after NIC added

A14: Use NAT

Q15: How to set IP address

A15: Refer to https://zhuanlan.zhihu.com/p/84698725

Q16: Error when initialize vfio, numa and pcie

```
EAL: Detected 2 lcore(s)
EAL: Detected 1 NUMA nodes
EAL: Detected shared linkage of DPDK
EAL: Multi-process socket /var/run/dpdk/rte/mp_socket
EAL: Selected IOVA mode 'PA'
EAL: No available 1048576 kB hugepages reported
EAL: Probing VFIO support...
EAL: VFIO support initialized
EAL:   Invalid NUMA socket, default to 0
EAL: Probe PCI driver: net_e1000_em (8086:100f) device: 0000:02:01.0 (soc
EAL: Error reading from file descriptor 20: Input/output error
EAL: No legacy callbacks, legacy socket not created
EAL: Error enabling interrupts for fd 20 (Input/output error)
```

```
root@ubuntu:/home/yg/cloudComputing/dpdk/build/examples# ./dpdk-skeleton -l 1 -n
EAL: Detected 4 lcore(s)
EAL: Detected 1 NUMA nodes
EAL: Detected static linkage of DPDK
EAL: Multi-process socket /var/run/dpdk/rte/mp_socket
EAL: Selected IOVA mode 'PA'
EAL: No available 1048576 kB hugepages reported
EAL: Probing VFIO support...
EAL: VFIO support initialized
EAL:   Invalid NUMA socket, default to 0
EAL: Probe PCI driver: net_e1000_em (8086:100f) device: 0000:02:01.0 (socket 0)
EAL:   Invalid NUMA socket, default to 0
EAL:   Invalid NUMA socket, default to 0
EAL: Error reading from file descriptor 10: Input/output error
EAL: Probe PCI driver: net_e1000_em (8086:100f) device: 0000:02:07.0 (socket 0)
EAL: Error reading from file descriptor 19: Input/output error
EAL: No legacy callbacks, legacy socket not created
EAL: Error enabling interrupts for fd 10 (Input/output error)
Port 0 MAC: 00 0c 29 3c 46 05
EAL: Error enabling interrupts for fd 19 (Input/output error)
Port 1 MAC: 00 0c 29 3c 46 19

Core 1 forwarding packets. [Ctrl+C to quit]
```

```
EAL: Detected 1 NUMA nodes
EAL: Detected static linkage of DPDK
EAL: Multi-process socket /var/run/dpdk/rte/mp_socket
EAL: Selected IOVA mode 'PA'
EAL: No available 1048576 kB hugepages reported
EAL: Probing VFIO support...
EAL: VFIO support initialized
EAL:   Invalid NUMA socket, default to 0
EAL: Probe PCI driver: net_e1000_em (8086:100f) device: 0000:02:01.0 (socket 0)
EAL:   Invalid NUMA socket, default to 0
EAL: Probe PCI driver: net_e1000_em (8086:100f) device: 0000:02:05.0 (socket 0)
EAL: Error reading from file descriptor 10: Input/output error
EAL: Error reading from file descriptor 19: Input/output error
EAL: No legacy callbacks, legacy socket not created
ports num: 2
EAL: Error enabling interrupts for fd 10 (Input/output error)
Port 0 MAC: 00 0c 29 b8 22 9e
EAL: Error enabling interrupts for fd 19 (Input/output error)
Port 1 MAC: 00 0c 29 b8 22 a8

Core 1 forwarding packets. [Ctrl+C to quit]
```

A16: This is normal, just go ahead

Q17: Cannot init port 0

```
root@ubuntu:/home/ics/dpdk/examples/lab2/build# ./basicfwd
EAL: Detected 4 lcore(s)
EAL: Detected 1 NUMA nodes
EAL: Detected shared linkage of DPDK
EAL: Multi-process socket /var/run/dpdk/rte/mp_socket
EAL: Selected IOVA mode 'PA'
EAL: No available 1048576 kB hugepages reported
EAL: Probing VFIO support...
EAL: VFIO support initialized
EAL:   Invalid NUMA socket, default to 0
EAL:   Invalid NUMA socket, default to 0
EAL: No legacy callbacks, legacy socket not created
Invalid port_id=0
EAL: Error - exiting with code: 1
  Cause: Cannot init port 0
```

A17: NIC not bind to dpdk

Q18: Error when helloworld init pci and only get one line output

```
root@iZuf62zy5jnv4oqd1f1kgtZ:~/feax/cloud/dpdk/build/examples# ./dpdk-helloworld
EAL: Detected 1 lcore(s)
EAL: Detected 1 NUMA nodes
EAL: Detected static linkage of DPDK
EAL: Multi-process socket /var/run/dpdk/rte/mp_socket
EAL: Selected IOVA mode 'PA'
EAL: No available 1048576 kB hugepages reported
EAL: Probing VFIO support...
EAL:   Invalid NUMA socket, default to 0
EAL: Probe PCI driver: net_virtio (1af4:1000) device: 0000:00:03.0 (socket 0)
eth_virtio_pci_init(): Failed to init PCI device

EAL: Requested device 0000:00:03.0 cannot be used
EAL: No legacy callbacks, legacy socket not created
hello from core 0
root@iZuf62zy5jnv4oqd1f1kgtZ:~/feax/cloud/dpdk/build/examples#
```

A18: NIC not bind to dpdk. One line output is normal