Generic Flow API简介

原创 DPDK开源社区 2017-04-14

作者 邢蓓蕾

Generic Flow API
简介

Classification功能

Classification功能是指网卡在收包时，将符合某种规则的包放入指定的队列。网卡一般支持一种或多种classification功能，以intel 700系列网卡为例，其支持MAC/VLAN filter、Ethertype filter、Cloud filter、flow director等等。不同的网卡可能支持不同种类的filter，例如intel 82599系列网卡支持n-tuple、L2_tunnel、flow director等，下表列出了DPDK中不同驱动对filter的支持。即使intel 82599系列网卡和intel 700系列网卡都支持flow director，它们支持的方式也不一样。那DPDK是如何支持不同网卡的filter的呢？

Driver	MACVLAN	ETHERTYPE	FLEXIBLE	SYN	NYUPLE	TUNNEL	FDIR	HASH	L2_TUNNEL
bnx2x									
cxgbe									
e1000		yes	yes	yes	yes				
ena									
enic							yes		
fm10k									
i40e	yes	yes				yes	yes	yes	
ixgbe		yes		yes	yes		yes		yes
mlx4									
mlx5							yes		
szedata2									

现有flow API

DPDK定义所有网卡的filter类型以及每种filter的基本属性，并提供相应的接口给上层应用，所以要求用户对filter属性有一定的概念。以flow director为例，其中有两个数据结构为：

（可左右滑动查看以下代码）↓↓↓

```
/**
 * A structure used to contain extend input of flow
 */
struct rte_eth_fdir_flow_ext {
        uint16_t vlan_tci;
        uint8_t flexbytes[RTE_ETH_FDIR_MAX_FLEXLEN];
        /**< It is filled by the flexible payload to match. */
        uint8_t is_vf;    /**< 1 for VF, 0 for port dev */
        uint16_t dst_id; /**< VF ID, available when is_vf is 1*/
};
/**
 * A structure used to define the input for a flow director filter entry
 */
struct rte_eth_fdir_input {
        uint16_t flow_type;
        union rte_eth_fdir_flow flow;
        /**< Flow fields to match, dependent on flow_type */
        struct rte_eth_fdir_flow_ext flow_ext;
        /**< Additional fields to match */
};
```

用户在添加flow director流规则的时候，需要填写上述信息。但并不是所有的网卡都关注flow_type 和is_vf，比如intel 82599系列网卡不需要flow_type及is_vf。其它类型的filter也有类似的情况。

从上可以看出，现有方案有很多缺陷：首先，DPDK为所有网卡抽象出统一的属性，但是某些属性只对一种网卡有意义；其次，随着DPDK支持的网卡越来越多，DPDK需要定义的filter类型要增加，网卡filter功能升级也需要DPDK作相应修改，这样很容易导致API/ABI的破坏；

另外，从应用角度来看，现有的方案也有诸多不便，使得API比较难用，不够友好：那些可选或者可缺省的属性容易让用户产生疑惑；经常在某种filter类型中随意插入一些某个网卡特有的属性；设计复杂，也没有比较详细的说明文档。

鉴于上述原因，一个generic flow API必不可少。

Generic flow API

DPDK v17.02 推出了generic flow API方案，DPDK把一条流规则抽象为pattern和actions两部分。

Pattern由一定数目的item组成。Item主要和协议相关，支持ETH, IPV4, IPV6, UDP, TCP, VXLAN等等。item也包括一些标志符，比如PF, VF, END等，目前DPDK支持的item类型定义在rte_flow.h的enum rte_flow_item_type。在描述一个item的时候可以添加spec和mask，告诉驱动哪些需要匹配。下面以以太网包的流规则为例，该item描述的是精确匹配二层包头的目的地址11:22:33:44:55:66。

Ethernet		
spec	src	00:11:22:33:44:55
	dst	11:22:33:44:55:66
mask	src	00:00:00:00:00:00
	dst	ff:ff:ff:ff:ff:ff

Actions表示流规则的动作，比如QUEUE, DROP, PF, VF,END等等，DPDK支持的action类型定义在rte_flow.h的enum rte_flow_action_type。以下action表示符合某种pattern的包放入队列3。

QUEUE	
queue	3

该方案把复杂的区分filter类型的事情交给驱动处理，用户再也不需要关注硬件的能力，这样使得上层应用能够方便添加或者删除流规则。

如果想添加一条流规则，上层应用只需要调用rte_flow_create()这个接口，并填好相应的pattern和actions；如果要删除一条流规则，上层应用只需要调用rte_flow_destroy()。还是以flow director为例，流规则定义如下所示，对于用户来说，这种方式更易操作。

Pattern			Actions
0	ETH	spec	QUEUE,DROP,PASSTHRU
		mask	
1	IPV4,IPV6	spec	
		mask	
2	TCP,UDP,SCTP	spec	
		mask	
3	PF,VF,SIGNATURE	spec	
		mask	

综上所述，generic flow API明显方便很多。

> > > 作 者 简 介 < < <

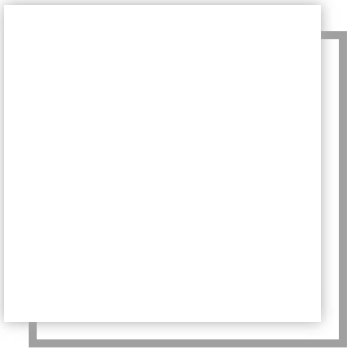
邢蓓蕾， 英特尔软件工程师，主要从事DPDK PMD的开发工作。

往期精选文章

DPDK开源社区

- 基于virtio-user的新exception path方案
- DPDK Release 17.02
- Hyperscan Release 4.4.0
- DPDK Release 16.11
- 无锁队列详细分解——Lock与Cache，到底有没有锁？
- 从计算机架构师的角度看DPDK性能
- 欢迎搭乘Hyperscan号极速列车~
- 无锁队列详细分解 — 顶层设计
- VMware Player 搭建DPDK实验平台
- Qemu/Kvm 搭建DPDK实验平台

长按扫描二维码关注我们



投诉