OPENFABRICS ALLIANCE

10TH ANNUAL INTERNATIONAL DEVELOPER WORKSHOP

OFVWG: IPoIB LLD acceleration support

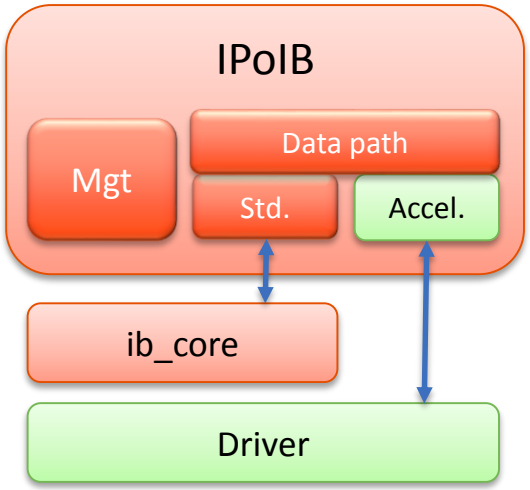
Rony Efraim

Goals

- Allow vendors to optimize IPoIB data path
  - Buffer management
  - Interrupt moderation
  - Tunneling offloads
- Avoid bloating kAPI
- Leverage IPoIB ULP code as much as possible
- Support UD mode only

Architecture

- Separate functionality into management and data path
- Management roles (IPoIB)
  - Interface registration and APIs
  - Multicast management
  - Event processing
  - Logical link state
  - Path resolution
  - ARP
- Data path roles (Driver)
  - Multi-queue support
  - Buffering
  - Receiving packets
  - Sending packets



Packet Flows

- Tx
  - IPoIB
    - Resolve address path
    - Pass AH attributes
  - Driver
    - Select Tx queue
    - Packet transmission
    - Free skb
- Rx
  - Driver
    - Receive ring processing
      - NAPI
      - skb allocation
    - Call netif_receive_skb

IPoIB Operation Template

```
struct ipoib_ops {
     void (*get_mq_param)(struct ib_device *hca, int* priv_size,
          int *rx_queues, int *tx_queues);
     int (*ipoib_ops_update_netdev_settings) (struct net_device *dev);
     int (*ib_dev_init)(struct net_device *dev, struct ib_device *hca, uint pkey,
          uint *qp_num);

     void (*ib_dev_open) (struct net_device *dev);
     void (*ib_dev_stop) (struct net_device *dev);

     int (*set_qkey)(struct net_device *dev)
     int (*attach_mcast)(struct net_device *dev , union ib_gid *gid, u16 lid);
     int (*dettach_mcast)(struct net_device *dev , union ib_gid *gid, u16 lid);
     int (*send) (struct net_device *dev, struct sk_buff *skb,
          struct ib_ah *address, u32 qpn);

     void (*ib_dev_cleanup)(struct net_device *dev);
};
```

- Note
  - QP passed by number only
  - netdev providers driver context

Event Responsibility

Event	ULP	Driver
Interface up/down	V	
Physical port up/down	V	
Carrier up/down	V	
Client Reregister	V	
PKey change	V	
MTU change	V	
Tx timeout	V	
Select queue		V
Tx queue management (netif_wake/stop_queue, polling)		V
Rx queue management (NAPI, polling, buffer allocation)		V

Thank You

OPENFABRICS ALLIANCE

10TH ANNUAL INTERNATIONAL DEVELOPER WORKSHOP