Virtio网络发包过程分析

2018-08-24 | 2.虚拟化， virtio | 1596 | 1
19k | 18 分钟

前言

本文将分析Virtio网络发包过程，包括Frontend如何发送数据给Backend，以及Backend收到数据后的处理。

正文

Frontend： Transmit packet

查看driver注册的ops函数，其中指定的发送函数为**start_xmit**，我们从这里开始分析发包过程。

> virtio_net.c : virtnet_netdev

```
1  static const struct net_device_ops virtnet_netdev = {
2          .ndo_open           = virtnet_open,
3          .ndo_stop           = virtnet_close,
4          .ndo_start_xmit     = start_xmit,
5          .ndo_validate_addr  = eth_validate_addr,
6          .ndo_set_mac_address = virtnet_set_mac_address,
7          .ndo_set_rx_mode    = virtnet_set_rx_mode,
8          .ndo_change_mtu     = virtnet_change_mtu,
9          .ndo_get_stats64    = virtnet_stats,
10         .ndo_vlan_rx_add_vid = virtnet_vlan_rx_add_vid,
11         .ndo_vlan_rx_kill_vid = virtnet_vlan_rx_kill_vid,
12         .ndo_select_queue   = virtnet_select_queue,
13 #ifdef CONFIG_NET_POLL_CONTROLLER
14         .ndo_poll_controller = virtnet_netpoll,
15 #endif
16         .ndo_features_check   = passthru_features_check,
17 };
```

start_xmit函数首先获取sk_buff对应的send_queue，**free_old_xmit_skbs**释放backend处理过的desc，然后调用**xmit_skb**发包，之后调用**skb_orphan**孤立skb,因为对于guest来说**xmit_skb**之后skb的内容已经是发送的了，最后调用**virtqueue_kick**通知Host。

> virtio_net.c : start_xmit

```
1  static netdev_tx_t start_xmit(struct sk_buff *skb, struct net_device *dev)
2  {
3      struct virtnet_info *vi = netdev_priv(dev);
4      int qnum = skb_get_queue_mapping(skb);
5      struct send_queue *sq = &vi->sq[qnum];  //对应的send_queue
6      int err;
7      struct netdev_queue *txq = netdev_get_tx_queue(dev, qnum);
8      bool kick = !skb->xmit_more;
9
10     /* Free up any pending old buffers before queueing new ones. */
11 /*释放backend处理过的desc*/
12     free_old_xmit_skbs(sq);
13
14     /* Try to transmit */
15 /*发包*/
16     err = xmit_skb(sq, skb);
17
18     /* This should not happen! */
19     if (unlikely(err)) {
20         dev->stats.tx_fifo_errors++;
21         if (net_ratelimit())
22             dev_warn(&dev->dev,
23                  "Unexpected TXQ (%d) queue failure: %d\n", qnum, err);
24         dev->stats.tx_dropped++;
25         kfree_skb(skb);
26         return NETDEV_TX_OK;
27     }
28
29     /* Don't wait up for transmitted skbs to be freed. */
30 /*孤立skb,因为此时对guest来说此skb的内容已发送*/
31     skb_orphan(skb);
32     nf_reset(skb);
33
34     /* Apparently nice girls don't return TX_BUSY; stop the queue
35      * before it gets out of hand.  Naturally, this wastes entries. */
36     if (sq->vq->num_free < 2+MAX_SKB_FRAGS) {
37         netif_stop_subqueue(dev, qnum);
38         if (unlikely(!virtqueue_enable_cb_delayed(sq->vq))) {
39 ↑ 90%       /* More just got used, free them then recheck. */
```

```
40                  free_old_xmit_skbs(sq);
41                  if (sq->vq->num_free >= 2+MAX_SKB_FRAGS) {
42                      netif_start_subqueue(dev, qnum);
43                      virtqueue_disable_cb(sq->vq);
44                  }
45              }
46          }
47
48          if (kick || netif_xmit_stopped(txq))
49      /*通知host端*/
50              virtqueue_kick(sq->vq);
51
52          return NETDEV_TX_OK;
53      }
```

下面我们来分析**xmit_skb**函数是如何发包的，主要就是将数据包头部和数据包填入scatterlist，然后调用**virtqueue_add_outbuf**将sg table 写入desc描述符表，并将head desc信息写入vring.avail。

> virtio_net.c : xmit_skb

```
1   static int xmit_skb(struct send_queue *sq, struct sk_buff *skb)
2   {
3       struct virtio_net_hdr_mrg_rxbuf *hdr;  //数据包头部
4       const unsigned char *dest = ((struct ethhdr *)skb->data)->h_dest;
5       struct virtnet_info *vi = sq->vq->vdev->priv;
6       unsigned num_sg;
7       unsigned hdr_len = vi->hdr_len;       //Packet virtio header size
8       bool can_push;
9
10      pr_debug("%s: xmit %p %pM\n", vi->dev->name, skb, dest);
11
12      can_push = vi->any_header_sg &&
13          !((unsigned long)skb->data & (__alignof__(*hdr) - 1)) &&
14          !skb_header_cloned(skb) && skb_headroom(skb) >= hdr_len;
15      /* Even if we can, don't push here yet as this would skew
16       * csum_start offset below. */
17      if (can_push)
18          hdr = (struct virtio_net_hdr_mrg_rxbuf *)(skb->data - hdr_len);
19      else
20          hdr = skb_vnet_hdr(skb);
21  /*校验和相关设置*/
22      if (skb->ip_summed == CHECKSUM_PARTIAL) {
23          hdr->hdr.flags = VIRTIO_NET_HDR_F_NEEDS_CSUM;
24          hdr->hdr.csum_start = cpu_to_virtio16(vi->vdev,
```

```
25                                  skb_checksum_start_offset(skb));
26              hdr->hdr.csum_offset = cpu_to_virtio16(vi->vdev,
27                                                     skb->csum_offset);
28          } else {
29              hdr->hdr.flags = 0;
30              hdr->hdr.csum_offset = hdr->hdr.csum_start = 0;
31          }
32      /*如果是GSO数据包需要重新设置包头*/
33          if (skb_is_gso(skb)) {
34              hdr->hdr.hdr_len = cpu_to_virtio16(vi->vdev, skb_headlen(skb));
35              hdr->hdr.gso_size = cpu_to_virtio16(vi->vdev,
36                                                  skb_shinfo(skb)->gso_size);
37              if (skb_shinfo(skb)->gso_type & SKB_GSO_TCPV4)
38                  hdr->hdr.gso_type = VIRTIO_NET_HDR_GSO_TCPV4;
39              else if (skb_shinfo(skb)->gso_type & SKB_GSO_TCPV6)
40                  hdr->hdr.gso_type = VIRTIO_NET_HDR_GSO_TCPV6;
41              else if (skb_shinfo(skb)->gso_type & SKB_GSO_UDP)
42                  hdr->hdr.gso_type = VIRTIO_NET_HDR_GSO_UDP;
43              else
44                  BUG();
45              if (skb_shinfo(skb)->gso_type & SKB_GSO_TCP_ECN)
46                  hdr->hdr.gso_type |= VIRTIO_NET_HDR_GSO_ECN;
47          } else {
48              hdr->hdr.gso_type = VIRTIO_NET_HDR_GSO_NONE;
49              hdr->hdr.gso_size = hdr->hdr.hdr_len = 0;
50          }
51      /*判断设备是否支持合并buffer*/
52          if (vi->mergeable_rx_bufs)
53              hdr->num_buffers = 0;
54          sg_init_table(sq->sg, MAX_SKB_FRAGS + 2);
55          if (can_push) {
56              __skb_push(skb, hdr_len);
57              num_sg = skb_to_sgvec(skb, sq->sg, 0, skb->len);
58              /* Pull header back to avoid skew in tx bytes calculations. */
59              __skb_pull(skb, hdr_len);
60          } else {
61      /*数据包头部填入scatterlist*/
62              sg_set_buf(sq->sg, hdr, hdr_len);
63      /*数据包填入scatterlist*/
64              num_sg = skb_to_sgvec(skb, sq->sg + 1, 0, skb->len) + 1;
65          }
66      /*sg table 写入desc描述符表，head desc信息写vring.avail*/
67          return virtqueue_add_outbuf(sq->vq, sq->sg, num_sg, skb, GFP_ATOMIC);
68      }
```

可以看到**sg_set_buf**和**skb_to_sgvec**最后都调用了**sg_set_page**，**sg_set_page**中把sk_buffer(逻辑buffer)中的物理块的page信息、offset、len信息放入sg中。

> scatterlist.h : sg_set_buf

```
1  static inline void sg_set_buf(struct scatterlist *sg, const void *buf,
2                                unsigned int buflen)
3  {
4  #ifdef CONFIG_DEBUG_SG
5      BUG_ON(!virt_addr_valid(buf));
6  #endif
7      sg_set_page(sg, virt_to_page(buf), buflen, offset_in_page(buf));
8  }
```

skb_to_sgvec

> skbuff.c : skb_to_sgvec

```
1  int skb_to_sgvec(struct sk_buff *skb, struct scatterlist *sg, int offset, int len)
2  {
3      int nsg = __skb_to_sgvec(skb, sg, offset, len);
4
5      sg_mark_end(&sg[nsg - 1]);
6
7      return nsg;
8  }
```

> skbuff.c : __skb_to_sgvec

```
1   static int
2   __skb_to_sgvec(struct sk_buff *skb, struct scatterlist *sg, int offset, int len)
3   {
4       int start = skb_headlen(skb);
5       int i, copy = start - offset;
6       struct sk_buff *frag_iter;
7       int elt = 0;
8
9       if (copy > 0) {
10          if (copy > len)
11              copy = len;
12          sg_set_buf(sg, skb->data + offset, copy);
13          elt++;
14          if ((len -= copy) == 0)
15              return elt;
```

```
16              offset += copy;
17          }
18
19          for (i = 0; i < skb_shinfo(skb)->nr_frags; i++) {
20              int end;
21
22              WARN_ON(start > offset + len);
23
24              end = start + skb_frag_size(&skb_shinfo(skb)->frags[i]);
25              if ((copy = end - offset) > 0) {
26                  skb_frag_t *frag = &skb_shinfo(skb)->frags[i];
27
28                  if (copy > len)
29                      copy = len;
30                  sg_set_page(&sg[elt], skb_frag_page(frag), copy,
31                              frag->page_offset+offset-start);
32                  elt++;
33                  if (!(len -= copy))
34                      return elt;
35                  offset += copy;
36              }
37              start = end;
38          }
39
40          skb_walk_frags(skb, frag_iter) {
41              int end;
42
43              WARN_ON(start > offset + len);
44
45              end = start + frag_iter->len;
46              if ((copy = end - offset) > 0) {
47                  if (copy > len)
48                      copy = len;
49                  elt += __skb_to_sgvec(frag_iter, sg+elt, offset - start,
50                                        copy);
51                  if ((len -= copy) == 0)
52                      return elt;
53                  offset += copy;
54              }
55              start = end;
56          }
57          BUG_ON(len);
58          return elt;
59      }
```

```
static inline void sg_set_page(struct scatterlist *sg, struct page *page,
                               unsigned int len, unsigned int offset)
{
    sg_assign_page(sg, page);
    sg->offset = offset;
    sg->length = len;
}
```

下面看**virtqueue_add_outbuf**如何将sg table 写入desc描述符表。

```
int virtqueue_add_outbuf(struct virtqueue *vq,
                         struct scatterlist *sg, unsigned int num,
                         void *data,
                         gfp_t gfp)
{
    return virtqueue_add(vq, &sg, num, 1, 0, data, gfp);
}
```

virtqueue_add首先判断是否支持间接描述符表并且total_sg > 1(否则也不需要使用间接描述符表了)，如果支持indirect把descs_used设置为1，否则设为实际的sg entry总数total_sg。如果主描述符表的空闲表项数小于total_sg就错误返回。然后在for循环中先调用**vring_map_one_sg**函数得到scatterlist中数据的GPA,然后将sg table中entry信息分别对应到描述符表中。标记last desc信息，更新空闲desc数量以及free_head，最后将head desc信息写入vring.avail->ring[]并更新idx。

```
static inline int virtqueue_add(struct virtqueue *_vq,
                                struct scatterlist *sgs[],
                                unsigned int total_sg,
                                unsigned int out_sgs,
                                unsigned int in_sgs,
                                void *data,
                                gfp_t gfp)
{
    struct vring_virtqueue *vq = to_vvq(_vq);
    struct scatterlist *sg;
    struct vring_desc *desc;
    unsigned int i, n, avail, descs_used, uninitialized_var(prev), err_idx;
    int head;
    bool indirect;

    START_USE(vq);
    ......
    head = vq->free_head;
```

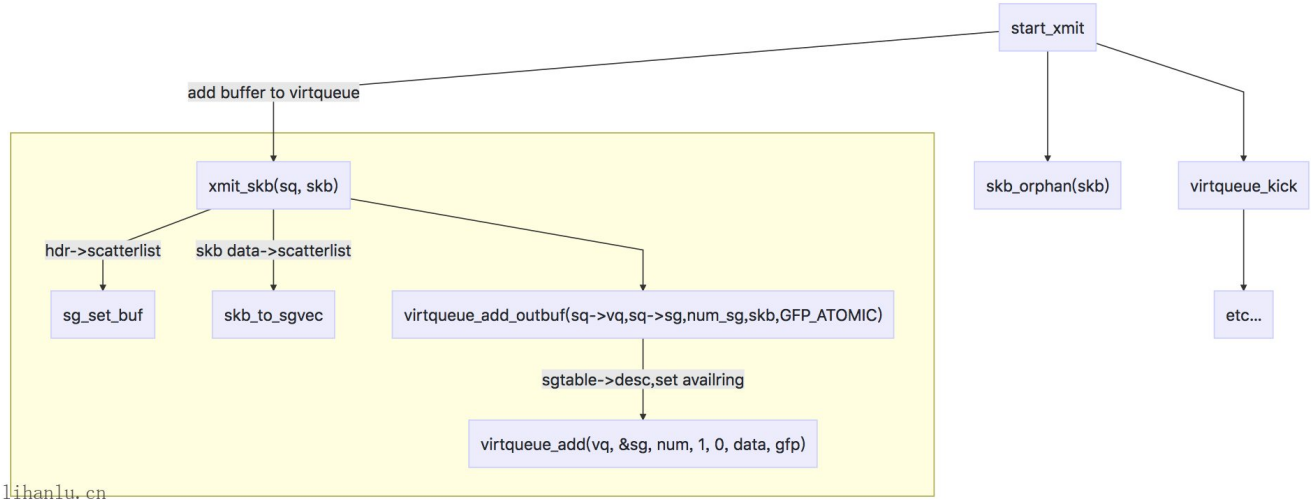
```
19  //判断是否支持间接描述符并且total_sg > 1
20      /* If the host supports indirect descriptor tables, and we have multiple
21       * buffers, then go indirect. FIXME: tune this threshold */
22      if (vq->indirect && total_sg > 1 && vq->vq.num_free)
23          desc = alloc_indirect(_vq, total_sg, gfp);
24      else
25          desc = NULL;
26
27      if (desc) {
28          /* Use a single buffer which doesn't continue */
29          indirect = true;
30          /* Set up rest to use this indirect table. */
31          i = 0;
32          descs_used = 1;
33      } else {
34          indirect = false;
35          desc = vq->vring.desc;
36          i = head;
37          descs_used = total_sg;
38      }
39  /*主描述符表的空闲表项数小于total_sg，错误*/
40      if (vq->vq.num_free < descs_used) {
41          ……
42          return -ENOSPC;
43      }
44  /*sg table中entry信息记录到对应的desc表中*/
45      for (n = 0; n < out_sgs; n++) {
46          for (sg = sgs[n]; sg; sg = sg_next(sg)) {
47  /*得到总线地址addr，也就是GPA*/
48              dma_addr_t addr = vring_map_one_sg(vq, sg, DMA_TO_DEVICE);
49              if (vring_mapping_error(vq, addr))
50                  goto unmap_release;
51  /*GPA等信息填入desc[]中*/
52              desc[i].flags = cpu_to_virtio16(_vq->vdev, VRING_DESC_F_NEXT);
53              desc[i].addr = cpu_to_virtio64(_vq->vdev, addr);
54              desc[i].len = cpu_to_virtio32(_vq->vdev, sg->length);
55              prev = i;
56              i = virtio16_to_cpu(_vq->vdev, desc[i].next);
57          }
58      }
59      for (; n < (out_sgs + in_sgs); n++) {
60          for (sg = sgs[n]; sg; sg = sg_next(sg)) {
61              dma_addr_t addr = vring_map_one_sg(vq, sg, DMA_FROM_DEVICE);
62              if (vring_mapping_error(vq, addr))
63                  goto unmap_release;
64
65              desc[i].flags = cpu_to_virtio16(_vq->vdev, VRING_DESC_F_NEXT | VRING_DESC_F_WRIT
```

```
66                  desc[i].addr = cpu_to_virtio64(_vq->vdev, addr);
67                  desc[i].len = cpu_to_virtio32(_vq->vdev, sg->length);
68                  prev = i;
69                  i = virtio16_to_cpu(_vq->vdev, desc[i].next);
70              }
71          }
72      /*标记last desc信息*/
73          /* Last one doesn't continue. */
74          desc[prev].flags &= cpu_to_virtio16(_vq->vdev, ~VRING_DESC_F_NEXT);
75
76          if (indirect) {
77              /* Now that the indirect table is filled in, map it. */
78              dma_addr_t addr = vring_map_single(
79                  vq, desc, total_sg * sizeof(struct vring_desc),
80                  DMA_TO_DEVICE);
81              if (vring_mapping_error(vq, addr))
82                  goto unmap_release;
83              vq->vring.desc[head].flags = cpu_to_virtio16(_vq->vdev, VRING_DESC_F_INDIRECT);
84              vq->vring.desc[head].addr = cpu_to_virtio64(_vq->vdev, addr);
85              vq->vring.desc[head].len = cpu_to_virtio32(_vq->vdev, total_sg * sizeof(struct vring_desc));
86          }
87      /*更新空闲desc数量*/
88          /* We're using some buffers from the free list. */
89          vq->vq.num_free -= descs_used;
90      /*更新free_head指针*/
91          /* Update free pointer */
92          if (indirect)
93              vq->free_head = virtio16_to_cpu(_vq->vdev, vq->vring.desc[head].next);
94          else
95              vq->free_head = i;
96
97          /* Store token and indirect buffer state. */
98          vq->desc_state[head].data = data;
99          if (indirect)
100             vq->desc_state[head].indir_desc = desc;
101         vq->vq.num_free -= descs_used;
102     /*desc链中第一个desc放入vring_avail中的ring[]*/
103         /* Put entry in available array (but don't update avail->idx until they
104          * do sync). */
105         avail = virtio16_to_cpu(_vq->vdev, vq->vring.avail->idx) & (vq->vring.num - 1);
106         vq->vring.avail->ring[avail] = cpu_to_virtio16(_vq->vdev, head);
107
108         /* Descriptors and available array need to be set before we expose the
109          * new available array entries. */
110         virtio_wmb(vq->weak_barriers);
111     /*更新vring_avail中idx*/
112         vq->vring.avail->idx = cpu_to_virtio16(_vq->vdev, virtio16_to_cpu(_vq->vdev, vq->vring.avail->i
```

90%

```
113         vq->num_added++;
114     ......
115     }
```

回到**xmit_skb**之后调用**virtqueue_kick**通知Host(具体的通知过程会另写一篇文章分析),上述过程总结如下图。



Backend

Backend收到通知后会执行**virtio_queue_host_notifier_read**函数，我们从这里开始分析

```
1   void virtio_queue_host_notifier_read(EventNotifier *n)
2   {
3       VirtQueue *vq = container_of(n, VirtQueue, host_notifier);
4       if (event_notifier_test_and_clear(n)) {
5           virtio_queue_notify_vq(vq);
6       }
7   }
```

virtio_queue_notify_vq中调用了**virtio_net_device_realize**中**virtio_net_add_queue**时绑定的处理函数**handle_output**，该函数根据不同的设备有不同的实现，以网卡为例看看创建VirtQueue的时候给绑定的是哪个函数。

```
1   static void virtio_queue_notify_vq(VirtQueue *vq)
2   {
3       if (vq->vring.desc && vq->handle_output) {
4           VirtIODevice *vdev = vq->vdev;
5
6           if (unlikely(vdev->broken)) {
7               return;
8           }
9
```

```
        trace_virtio_queue_notify(vdev, vq - vdev->vq, vq);
        vq->handle_output(vdev, vq);
    }
}
```

可以看到这里给rx_vq绑定的是**virtio_net_handle_rx**，tx_vq绑定的**virtio_net_handle_tx_timer**或**virtio_net_handle_tx_bh**，下面以**virtio_net_handle_tx_bh**为例看一下。

```
static void virtio_net_add_queue(VirtIONet *n, int index)
{
    VirtIODevice *vdev = VIRTIO_DEVICE(n);

    n->vqs[index].rx_vq = virtio_add_queue(vdev, n->net_conf.rx_queue_size,
                                           virtio_net_handle_rx);

    if (n->net_conf.tx && !strcmp(n->net_conf.tx, "timer")) {
        n->vqs[index].tx_vq =
            virtio_add_queue(vdev, n->net_conf.tx_queue_size,
                             virtio_net_handle_tx_timer);
        n->vqs[index].tx_timer = timer_new_ns(QEMU_CLOCK_VIRTUAL,
                                              virtio_net_tx_timer,
                                              &n->vqs[index]);
    } else {
        n->vqs[index].tx_vq =
            virtio_add_queue(vdev, n->net_conf.tx_queue_size,
                             virtio_net_handle_tx_bh);
        n->vqs[index].tx_bh = qemu_bh_new(virtio_net_tx_bh, &n->vqs[index]);
    }

    n->vqs[index].tx_waiting = 0;
    n->vqs[index].n = n;
}
```

virtio_net_handle_tx_bh中调用了**qemu_bh_schedule**(q->tx_bh)运行参数指定的函数，而参数tx_bh在**virtio_net_add_queue**中调用**qemu_bh_new**(virtio_net_tx_bh, &n->vqs[index])绑定为**virtio_net_tx_bh**

```
static void virtio_net_handle_tx_bh(VirtIODevice *vdev, VirtQueue *vq)
{
    VirtIONet *n = VIRTIO_NET(vdev);
    VirtIONetQueue *q = &n->vqs[vq2q(virtio_get_queue_index(vq))];

    if (unlikely((n->status & VIRTIO_NET_S_LINK_UP) == 0)) {
```

```
        virtio_net_drop_tx_queue_data(vdev, vq);
        return;
    }

    if (unlikely(q->tx_waiting)) {
        return;
    }
    q->tx_waiting = 1;
    /* This happens when device was stopped but VCPU wasn't. */
    if (!vdev->vm_running) {
        return;
    }
    virtio_queue_set_notification(vq, 0);
    qemu_bh_schedule(q->tx_bh);
}
```

virtio_net_tx_bh函数调用发送函数**virtio_net_flush_tx**

```
static void virtio_net_tx_bh(void *opaque)
{
    ……
    ret = virtio_net_flush_tx(q);
    ……
}
```

virtio_net_flush_tx函数调用**virtqueue_pop**从avail ring中取数据buffer的head desc，然后调用**qemu_sendv_packet_async**发送packet，qemu_sendv_packet_async函数本文不做详细的分析。

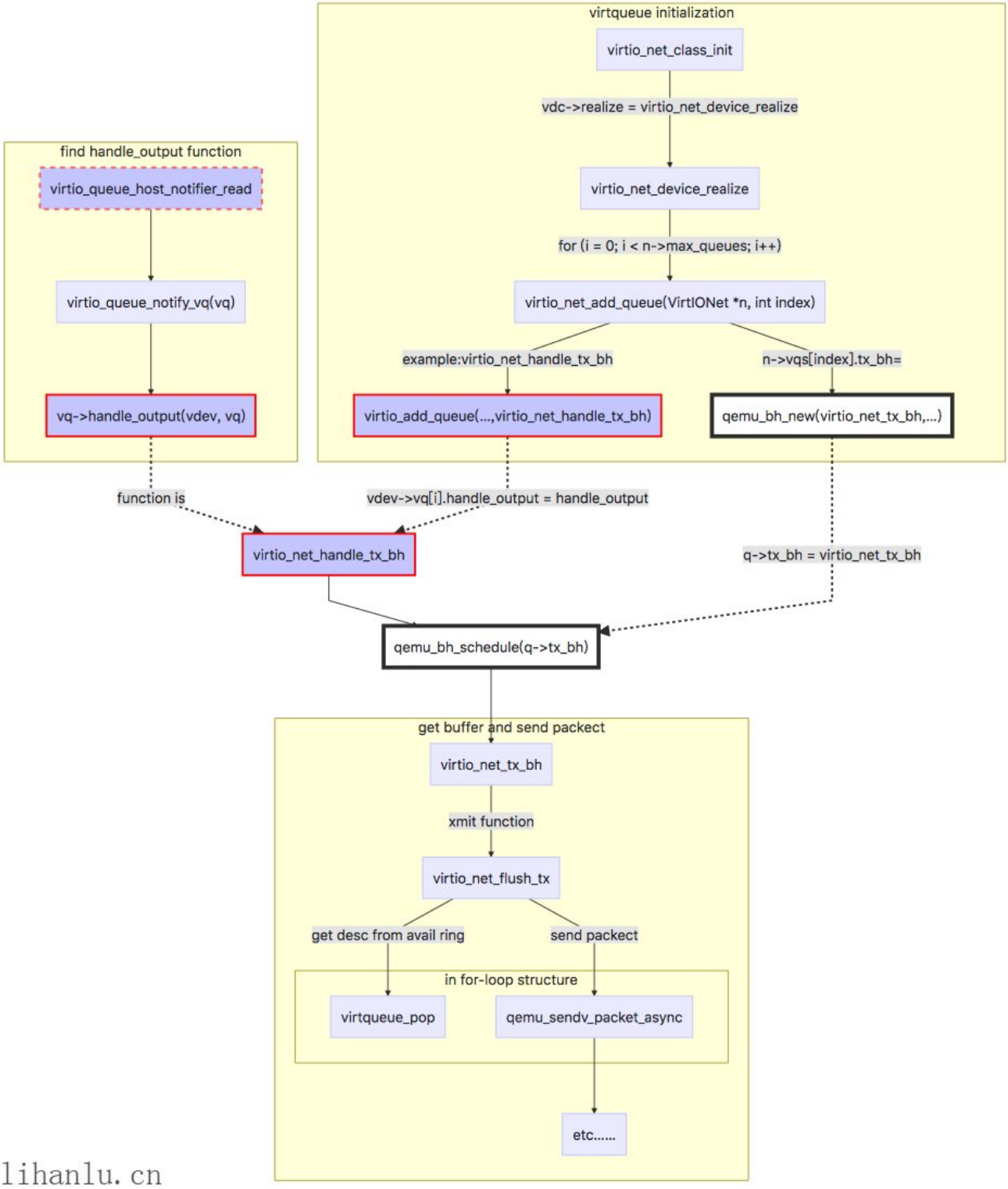
```
static int32_t virtio_net_flush_tx(VirtIONetQueue *q)
{
    VirtIONet *n = q->n;
    VirtIODevice *vdev = VIRTIO_DEVICE(n);
    VirtQueueElement *elem;
    int32_t num_packets = 0;
    int queue_index = vq2q(virtio_get_queue_index(q->tx_vq));
    if (!(vdev->status & VIRTIO_CONFIG_S_DRIVER_OK)) {
        return num_packets;
    }

    if (q->async_tx.elem) {
        virtio_queue_set_notification(q->tx_vq, 0);
        return num_packets;
    }
```

```
17      for (;;) {
18          ssize_t ret;
19          unsigned int out_num;
20          struct iovec sg[VIRTQUEUE_MAX_SIZE], sg2[VIRTQUEUE_MAX_SIZE + 1], *out_sg;
21          struct virtio_net_hdr_mrg_rxbuf mhdr;
22      //从avail ring中取数据buffer的head desc
23          elem = virtqueue_pop(q->tx_vq, sizeof(VirtQueueElement));
24          if (!elem) {
25              break;
26          }
27
28          out_num = elem->out_num;
29          out_sg = elem->out_sg;
30          if (out_num < 1) {
31              virtio_error(vdev, "virtio-net header not in first element");
32              virtqueue_detach_element(q->tx_vq, elem, 0);
33              g_free(elem);
34              return -EINVAL;
35          }
36
37          if (n->has_vnet_hdr) {
38              if (iov_to_buf(out_sg, out_num, 0, &mhdr, n->guest_hdr_len) <
39                  n->guest_hdr_len) {
40                  virtio_error(vdev, "virtio-net header incorrect");
41                  virtqueue_detach_element(q->tx_vq, elem, 0);
42                  g_free(elem);
43                  return -EINVAL;
44              }
45              if (n->needs_vnet_hdr_swap) {
46                  virtio_net_hdr_swap(vdev, (void *) &mhdr);
47                  sg2[0].iov_base = &mhdr;
48                  sg2[0].iov_len = n->guest_hdr_len;
49                  out_num = iov_copy(&sg2[1], ARRAY_SIZE(sg2) - 1,
50                                     out_sg, out_num,
51                                     n->guest_hdr_len, -1);
52                  if (out_num == VIRTQUEUE_MAX_SIZE) {
53                      goto drop;
54                  }
55                  out_num += 1;
56                  out_sg = sg2;
57              }
58          }
59          /*
60           * If host wants to see the guest header as is, we can
61           * pass it on unchanged. Otherwise, copy just the parts
62           * that host is interested in.
63           */
```

```
64          assert(n->host_hdr_len <= n->guest_hdr_len);
65          if (n->host_hdr_len != n->guest_hdr_len) {
66            unsigned sg_num = iov_copy(sg, ARRAY_SIZE(sg),
67                                       out_sg, out_num,
68                                       0, n->host_hdr_len);
69            sg_num += iov_copy(sg + sg_num, ARRAY_SIZE(sg) - sg_num,
70                               out_sg, out_num,
71                               n->guest_hdr_len, -1);
72            out_num = sg_num;
73            out_sg = sg;
74          }
75    /*调用qemu发包函数，通过qemu nic发送*/
76          ret = qemu_sendv_packet_async(qemu_get_subqueue(n->nic, queue_index),
77                                        out_sg, out_num, virtio_net_tx_complete);
78          if (ret == 0) {
79            virtio_queue_set_notification(q->tx_vq, 0);
80            q->async_tx.elem = elem;
81            return -EBUSY;
82          }
83
84    drop:
85    /*取消内存映射，更新used vring信息*/
86          virtqueue_push(q->tx_vq, elem, 0);
87          virtio_notify(vdev, q->tx_vq);
88          g_free(elem);
89
90          if (++num_packets >= n->tx_burst) {
91            break;
92          }
93        }
94        return num_packets;
95    }
```

Host收到通知后的发包流程总结如下图。



lihanlu.cn

联系我

你可以直接在下方留言，也可以E-Mail联系我。

打赏

本文作者： Lauren

90%

virtio

Virtio原理简介

virtio前端通知机制分析

昵称 邮箱 网址(http://)

Just go go

提交

1 评论

Anonymous Chrome 80.0.3987.132 Windows 10.0

2020-03-19

回复

先读一下,有问题再请教