NIC ドライバの送受信ルーチンの調査

2017/7/5

小倉 伊織

1 はじめに

共有メモリを介した OS ノード間通信を実現する際, OS ノード間通信用のネットワークインタフェースを新たに実装すると，コード改変量が膨大になると予想される．そこで，既存の NIC ドライバを改変することでコード改変量を抑える．具体的には，NIC ドライバにおけるパケットの送受信ルーチンを改変し，共有メモリを介して通信できるようにすることで，コードの改変量を抑える．このため，既存の NIC ドライバにおける送受信ルーチンについて調査した．具体的には，NIC ドライバのソースコードと書籍 [1] を読んで調査した．本資料では，既存の NIC ドライバにおける送受信ルーチンについて述べる．

2 調査環境

NIC ドライバの調査に使用した計算機の環境を表 1 に示す.

表 1　調査環境

項目名	環境
OS	Debian7.10
カーネル	Mint カーネル 3.15.0
CPU	Intel Core i7-860 Processor
メモリ	2.0GB
NIC ハードウェア	RTL8111/8168B PCI Express Gigabit Ethernet controller

3 調査環境における NIC ドライバ

OS には複数の NIC ドライバがあり，計算機に搭載されている NIC ハードウェアによって使用されるものが異なる．調査で使用した NIC ハードウェアに対応する NIC ドライバは `r8169` であり，ソースコードは `drivers/net/ethernet/realtek/r8169.c` に書かれている．

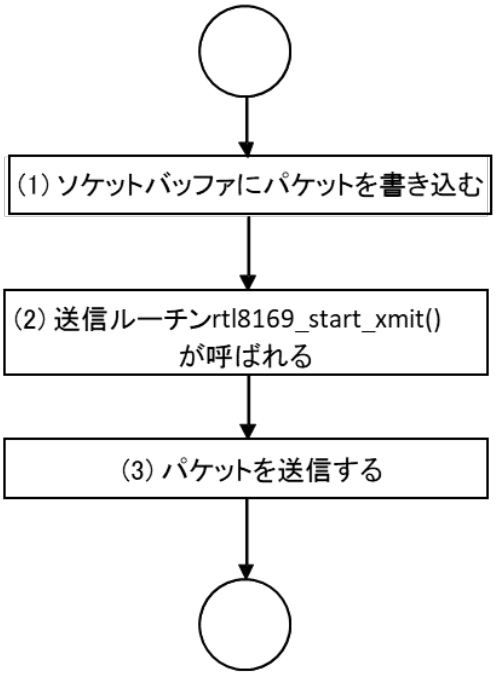

図 1　r8169 におけるパケットの送信ルーチンの処理流れ

4　NIC ドライバの送受信ルーチン

4.1　送信ルーチン

送信ルーチンは，アプリケーションがソケットバッファにデータを書き込むと呼ばれる．関数 `rtl8169_start_xmit()` は，ソケットバッファに書き込まれるたびに呼ばれている．また，この関数内でソケットバッファ中のパケットを DMA バッファに転送している．このため，`rtl8169_start_xmit()` は送信ルーチンである．

送信ルーチンの処理について図 1 に示し，以下で説明する．

(1) アプリケーションがソケットバッファにパケットを書き込む．
(2) 送信ルーチン `rtl8169_start_xmit` が呼ばれる．この送信ルーチン内で，パケットを DMA バッファにコピーする．
(3) パケットが送信される．

送信ルーチン `rtl8169_start_xmit` について以下に示す．

【形式】`static netdev_tx_t rtl8169_start_xmit(struct sk_buff *skb, strcut net_device *dev)`
【引数】struck sk_buff *skb: ソケットバッファ
struct net_device *dev: net_device 構造体
【戻り値】成功: `NETDEV_TX_OK`
失敗: `NETDEV_TX_BUSY`
【機能】ソケットバッファとして渡されたパケットを，DMA バッファへ書き込む．その後，ソケットバッファを解放する．この関数では，パケットの送信時の時刻を記録しており，この時間を基準

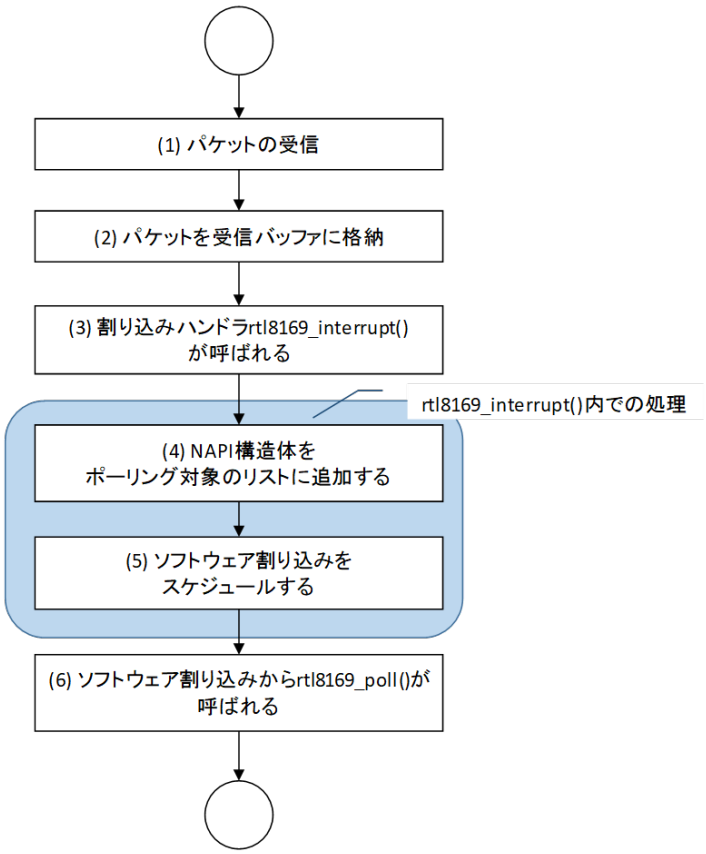

図 2　r8169 におけるパケットの受信ルーチンの処理流れ

に一定時間が経過すると送信をタイムアウトする．また，DMA バッファがいっぱいになった場合には，上位レイヤからの送信を一時停止させる．

4.2　受信ルーチン

受信ルーチンは，パケットの受信タイミングに関係せず，ネットワークへの接続が確立でき次第，非同期で呼ばれる．割り込みハンドラ `rtl8169_interrupt()` とポーリング関数 `rtl8169_poll()` は，ネットワークに接続されている場合，繰り返し呼ばれている．また，`rtl8169_poll()` は，`rtl8169_interrupt()` によってスケジュールされるソフトウェア割り込みで呼ばれ，ポーリングで受信パケットを上位レイヤに転送している．このため，`rtl8169_poll()`，`rtl8169_interrupt()`，およびこれによってスケジュールされるソフトウェア割り込みが受信ルーチンである．

受信ルーチンでの処理について図 2 に示し，以下で説明する．

(1) ネットワークデバイスが外部からパケットを受信する．
(2) パケットをデバイス内の受信バッファに格納する．
(3) ハードウェア割り込みによって，割り込みハンドラ `rtl8169_interrupt()` が呼ばれる．
(4) net_device 構造体中の NAPI 構造体をポーリング対象のリストに追加する．NAPI(New API) とは，カーネルが提供する NIC ドライバフレームワークである．

(5) ソフトウェア割り込みをスケジュールする．

(6) ソフトウェア割り込みからポーリング関数 rtl8169_poll() が呼ばれる．この関数によって受信処理が行われる．

以上のようにして受信処理が行われる．割り込みハンドラ rtl8169_interrupt() とポーリング関数 rtl8169_poll() について以下に示す．

(1) 割り込みハンドラ rtl8169_interrupt()

【形式】static irqreturn_t rtl8169_interrupt(int irq, void *dev_instance)

【引数】int irq: irq 番号

void *dev_instance: net_device 構造体

【戻り値】成功: IRQ_HANDLED

失敗: IRQ_NONE

【機能】NAPI 構造体をポーリング対象のリストに追加する．その後，ソフトウェア割り込みをスケジュールする．

(2) ポーリング関数 rtl8169_poll()

【形式】static int rtl8169_poll(struct napi_struct *napi, int budget)

【引数】struck napi_struct *napi: NAPI 構造体

int budget: 受信処理の上限数

【戻り値】受信パケットを上位レイヤに渡した回数

【機能】パケットの受信処理をポーリングで行う．受信したパケットは，上位レイヤに渡し，渡した回数を記録する．記録した回数と受信処理の上限数を比較することで受信バッファが空か否かを判別する．受信バッファが空になる，または，受信処理を上限回数分行うことで処理を終了する．

5 おわりに

本資料では，NIC ドライバ r8169 での送受信ルーチンについて述べた．本資料から，パケットの送信に関しては rtl8169_start_xmit() を，受信に関しては rtl8169_poll() を改変することで，共有メモリを介した OS ノード間通信が実現できると考えられる．

参考文献

[1] Jonathan Corbet, Alessandro Rubini, Greg Kroah-Hartman : Linux デバイスドライバ 第 3 版 (2005).