

Mintオペレーティングシステムにおける 割り込みルーティングの変更による NICの移譲方式

増田 陽介

岡山大学 工学部 情報工学科

平成26年2月14日

研究背景

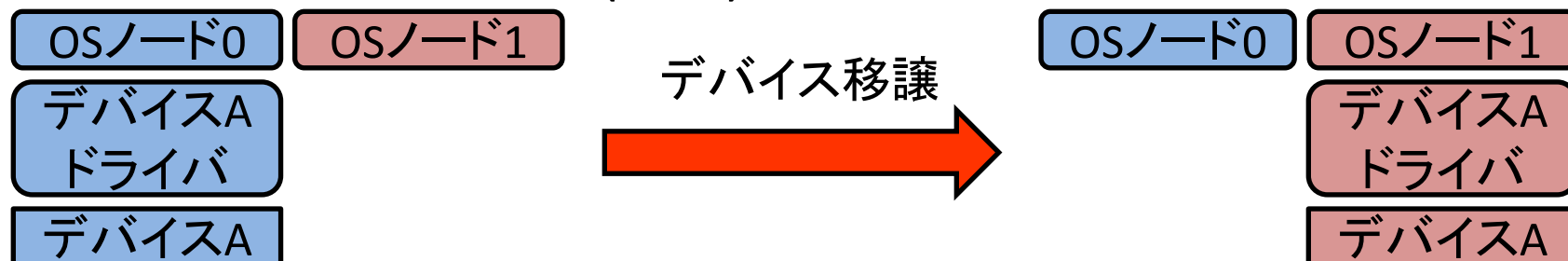
計算機資源を効率的に利用するためにMintを開発

<Mint>

(1) 複数のLinuxが同時に走行

(2) 計算機資源を分割し占有

Loadable Kernel Module(LKM)を利用したデバイス移譲が可能



LKMを利用した場合、移譲時間が長い

➡ デバイス移譲にかかる時間を最小限にしたい



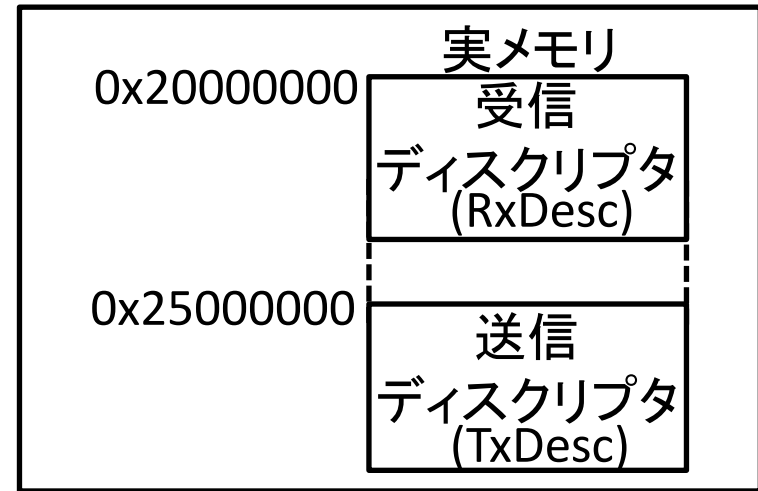
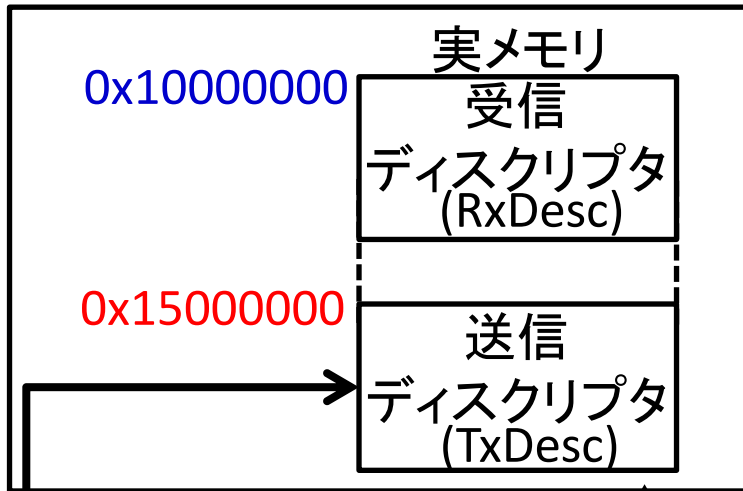
割り込みルーティングの変更による高速なNIC移譲を提案

最小限のレジスタの初期化と変数の初期化により実現

NIC機能の概要

OSノード0

OSノード1



(1) パケットの情報を取得

(2) パケット送信

NIC

(3) 送信完了割り込み通知

S/W
H/W

割り込み通知先	OSノード0
RxDesc	0x10000000
TxDesc	0x15000000

NICはパケットの送受信時にディスクリプタを参照

∴ 割り込みルーティングに加えてNIC固有の変更が必要

割り込みルーティングの変更による NICの移譲

NICを対象とした高速なデバイス移譲を実現

NICの移譲のために2つを変更

<割り込みルーティングの変更>

(1) レジスタの書き換えにより割り込みルーティングを変更

<NIC固有の変更>

(1) NICが参照するディスクリプタの変更

(2) NICとNICドライバが参照するエントリのずれの修正

割り込みルーティングの変更

<Message Signaled Interrupt(MSI)>

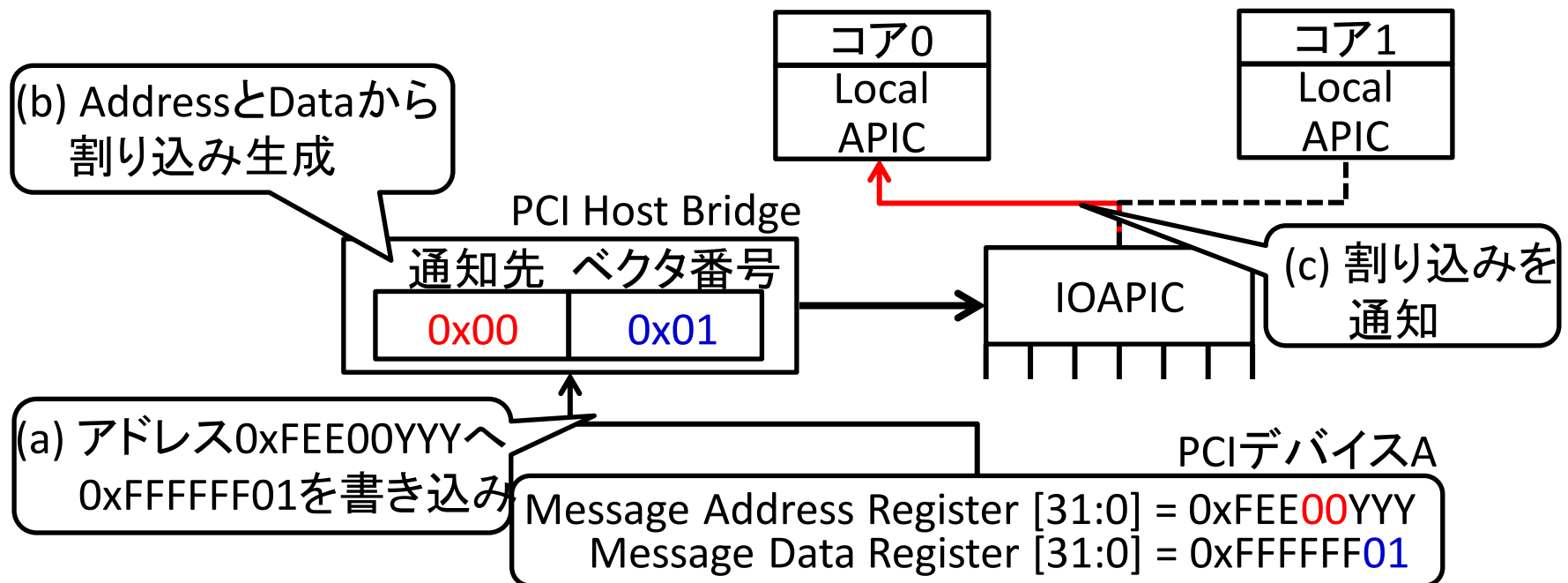
(1) PCIデバイスがメモリアクセスにより割り込みを通知する手法

(2) PCIデバイスはMSIに使用するレジスタを2つ所持

(A) Message Address Register


(B) Message Data Register

2つのレジスタの書き換えで、割り込みルーティングを変更




NIC固有の変更

<NICが参照するディスクリプタの変更>

- (1) ディスクリプタの先頭アドレスを持つレジスタを変更
 - (2) NICはカウンタを利用して参照するエントリを決定
 - (3) レジスタの変更の反映にはカウンタを初期化する必要有
-  NICの送受信処理を再起動し, レジスタの変更を反映

<NICとNICドライバが参照するエントリのずれの修正>

- (1) NICはレジスタに格納されているアドレスから処理を再開
 - (2) NICドライバは変数を利用して参照するエントリを決定
 - (3) NICとNICドライバが参照するエントリにずれが発生
-  変数を初期化しエントリのずれを修正

評価

<評価概要>

- (1) 提案手法によるNIC移譲機能の性能を評価
- (2) LKMを利用した場合と移譲時間を比較
- (3) LKMはロード, アンロードの実行時間の合計を計測
- (4) 提案手法はシステムコールの実行時間を計測

NIC	RTL8111/8168B PCI Express Gigabit Ethernet controller
NICドライバ	RTL8169

<評価結果>

LKMを利用したNIC移譲

処理手順	処理時間(ms)
ロード	4.79
アンロード	3.40
合計	8.19

提案手法によるNIC移譲

処理手順	処理時間(ms)
システムコール呼出し	1.02
移譲処理	0.02
合計	1.04

- (1) LKMを利用した場合と比較して移譲時間を87%削減した
- (2) 処理時間の大半をシステムコール呼出しが占める

まとめ

<実績>

- (1) 2つの変更により高速なNIC移譲を実現
- (2) 割り込みルーティングの変更
 - (A) MSIを利用する割り込みはメモリアクセスで割り込み生成
 - (B) アドレスとデータを変更し、割り込みルーティングを変更
- (3) NIC固有の変更
 - (A) 変更の反映のために、送受信処理を再起動
 - (B) NICとNICドライバが参照するエントリにずれが発生
 - (C) エントリを決定する変数を初期化してずれを修正
- (4) 評価
 - (A) 移譲時間を87%削減
 - (B) 処理時間の大半をシステムコール呼出しが占める