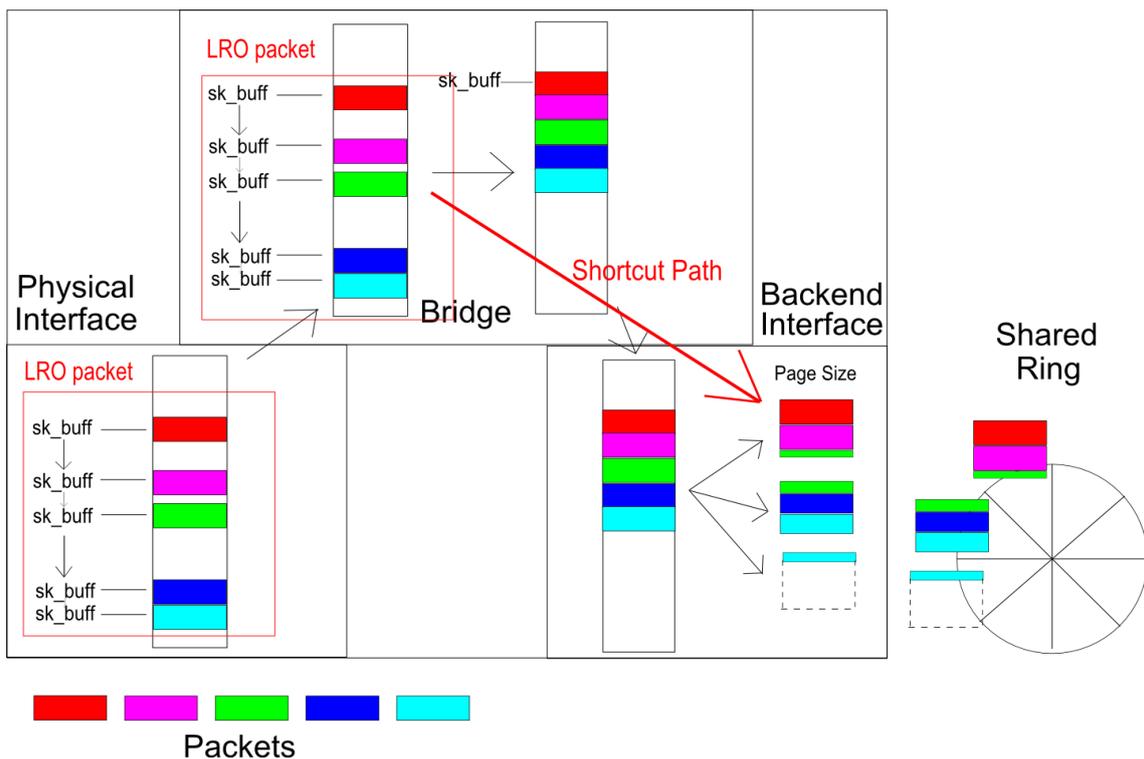


# Optimization and Performance Analysis of Large Receive Offload in the Xen Virtualized System

## 1. 目的

我々は以前の研究で[1]、Large Receive Offload というネットワーク性能を向上させる手法を Xen を用いた仮想化環境に適用した。その結果いくらかの性能の向上が見られたのだが、余計な処理も増えることになってしまった。それを省き、性能の向上を図るのが今回の SCCP の目的である。

## 2. 方法



上の図は Xen の Domain0 内で LRO 処理されたパケットが通る道筋を表したものである。パケットは Physical Interface (デバイスドライバ)によってカーネルメモリにコピーされ、sk\_buff 構造体と関連付けられる。その後 LRO モジュールがそれらをグループ化しブリッジへ送信する。ブリッジではどの仮想マシンに繋がっている Backend Interface に送るかを決め、そこへ送ろうとする。しかし送り先である Backend Interface は、このグループ化されたパケットを処理することができないので、ブリッジはパケットのデータを連続したメ

メモリ領域にコピーし、1つの大きなパケットを作り、Backend Interface に送る。Domain0 と DomainU でのデータのやりとりはページ単位で行われるため、Backend Interface は新しくページを割り当ててそこにパケットデータをコピーする。そのページは Shared Ring を通して DomainU に送られる。この一連の処理の中での、Backend Interface が LRO されたパケットを扱えないために生じるブリッジでのコピーは無駄な処理であるため、そこを省くように、つまり Backend Interface が LRO されたパケットを扱えるようにして連続してない領域にあるパケット郡をうまくページにコピーするよう修正したのが今回の最適化である。

### 3. 実験

実験はデータを送信するマシンと Xen で仮想化した環境でデータを受信するマシンを GbE 対応のクロスケーブルで繋ぎ、Netperf を使用してスループットを計測した。さらに Oprofile [2] を用いて各ドメインでのクロックサイクル数を測った。また、送信マシンには DELLPowerEdge430 を使用し、受信用マシンには DELLPowerEdge440 を使用した。

### 4. 結果

Configuration の Xen は仮想化しただけの環境、LRO はその Physical Interface で LRO をしたもの、Optimized はさらに Domain0 でのメモリコピーを減らしたものである。

Config.	Throughput	LRO rate
Xen	574.45	1.00
LRO	588.65	2.49
Optimized	588.47	2.43

Throughput (Mbps) and LRO rate (packets)

Config.	Domain0	DomainU
Xen	12.72	8.97
LRO	13.51	7.08
Optimized	12.64	7.78

Clock Cycles per byte in each domain

### 5. 考察

Xen と LRO での結果の差異については今回の論点ではないので割愛する。LRO と Optimized とを比べるとスループットと LROrate に関しては差が見受けられない。

Domain0でのクロックサイクル数が僅かに減少しているのが分かるが、これはメモリコピーを減らしたために得られた結果だと思われる。しかし DomainUでのクロックサイクル数が増えてしまっている。これは想定しない結果である。なぜなら、変更前と変更後での DomainUが受け取るページの内容は同じになるように Backend Interfaceに変更を加えたからだ。DomainUには変更を加えてないので、それらが全く同じページにはなっていなかったという可能性がある。Backend Interfaceでのページ内の sk\_buff 構造体の中身は同じであることは確認済みであるので、何か他の部分が違うのかもしれない。考えられる可能性としては、Physical Interfaceからブリッジを通して Backend Interfaceにつくまでに LRO パケットがどう処理されるのかを大雑把にしか見てないので、その間に LRO されたパケットにいくらか変更が加えられたのかもしれないということだ。DomainUでのクロックサイクル数を元に戻すことができればスループットは向上すると思われる。

## 6.参考文献等

[1] Takayuki Hatori and Hitoshi Oi, "Implementation and Analysis of Large Receive Offload in a Virtualized System" in Proceedings of the Virtualization Performance: Analysis, Characterization and Tools (VPAT'08), Austin, TX April 2008

[2] Oprofile, website, retrieved from <http://oprofile.sourceforge.net/> on Jun. 1, 2007.