Following is the data that I was able to collect last week regarding our 64-bit/NUMA agenda.

Observe the trending in the last two lines of the table: unless we apply all the three optimization viz – NUMA aware buffer allocation, buffer alignment (to cache lines) and removing of all references to shared variables in the adapter/netdev objects, we do not reach the peak of 5.7 Mpps. This is somewhat puzzling – considering that throughput is not affected by odd sized/ misaligned packets. Yet, we needed buffer alignment to go from 4bMpps to 5.7 Mpps.

Where do we go from here? Remove 3 instances of spinlocks forcefully from the xmit path in the kernel J, or add a buffer recycling patch (as we had done before) in order to cut down the overall cost of processing packet (which is about 3000 cycles presently).

I have cleaned the driver code and added conditional flags to compile each of these options independently.

- Hari

_____

Performance Measurement and Analysis

Intel/ DEG/ ECG/ Chandler, x552-3163

Platform

Green City, NHM 2x, 2.53 GHz, NO SMT

DDR3, 1066MHz, 2 x 3 x 2GB per DIMM = 12 GB

RHEL 5.1, Linux 2.6.30.1 SMP+NUMA+(SLUB)

IxGbE Driver Version: 2.0.34.3 + ECG Optimizations

1x Dual Port Niantic, 8 queue pairs per port enabled

L3 Forwarding with 8 Flows (1 flow per queue and per core)

LEGEND

- STOCK = Original 2.0.34.3 driver

- NUMA = Node aware descriptor and packet buffer allocation

- ALIGN = packet buffer is aligned to start at a Cache line boundary

- NO_SHARED_REFs = we have removed references to some variables (like device stats) that are shared by different HW threads.

Ethernet Frame Size (Bytes) → Driver Optimizations ↓	64	65	68	72	128	129	132
STOCK	1,534,512						
STOCK + NUMA	1,589,069	1,590,253	1,485,897	1,593,753	1,596,406	1,591,232	1,399,414
STOCK + ALIGN	1,544,697	1,438,494	1,522,180	1,522,180	1,485,238	1,550,954	1,548,117
STOCK + NO_SHARED_REFS	1,589,069	1,438,494	1,522,180	1,522,180	1,485,238	1,550,954	1,548,117
STOCK + ALIGN + NO_SHARED_REFS	1,542,680						
STOCK + NUMA + ALIGN	2,974,131						
STOCK + NUMA + NO_SHARED_REFS	4,020,881						
STOCK + NUMA + NO_SHARED_REFS + ALIGN	5,741,256	5,660,142	5,690,721	5,725,247	5,653,075	5,718,067	5,678,655

Small Packet Forwarding on Nehalem/Ninatic in 2-port, 8-Q Configuration
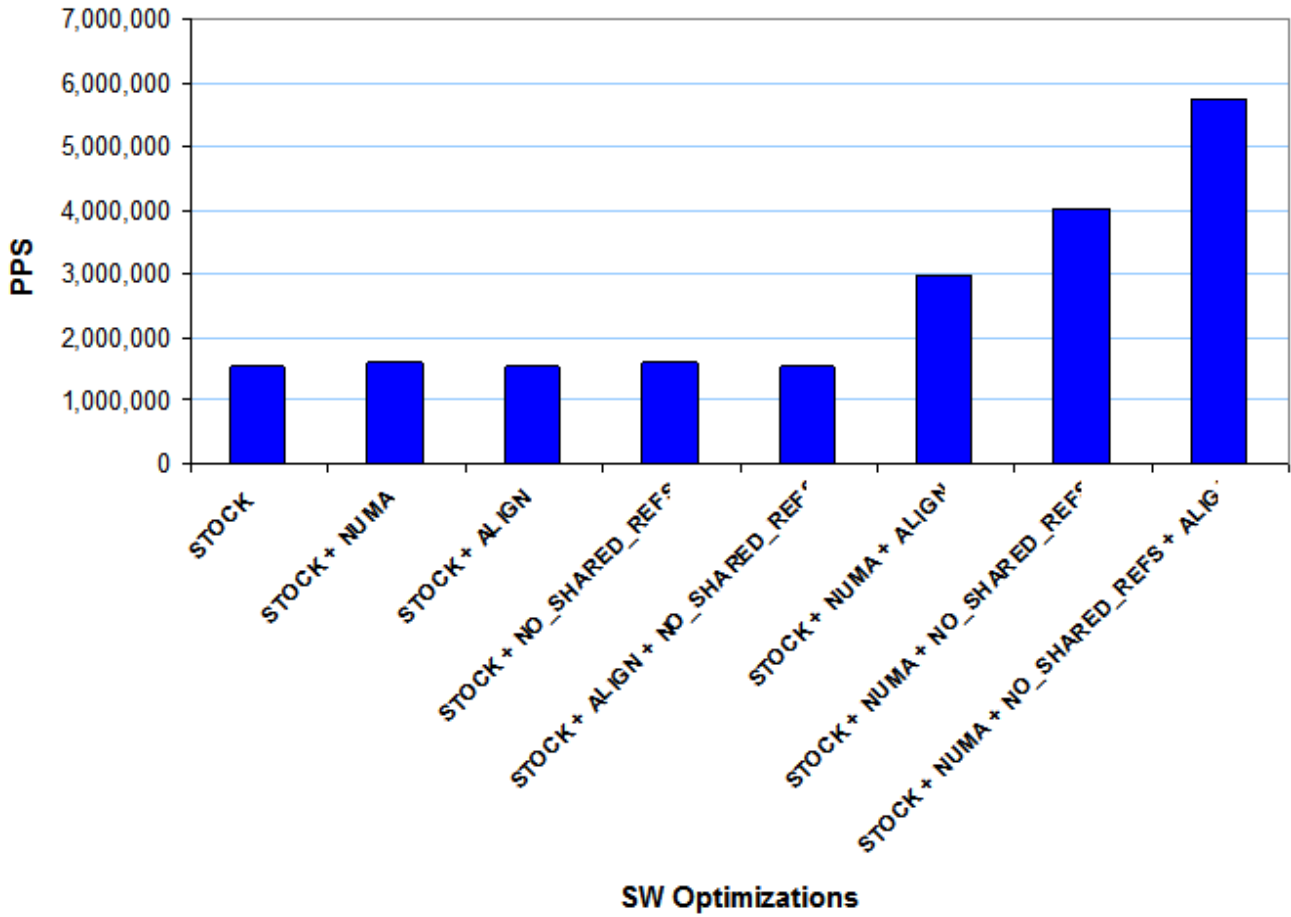
"It is better to do the right problem the wrong

way than the wrong problem the right way"

—— Richard Hamming