Threaded Network Interrupts

Steven Rostedt
srostedt@redhat.com
<rostedt@goodmis.org>
http://people.redhat.com/srostedt
network-thread-irq.odp

Disclaimer

I'm not a network guru

I had to fight for the hardware I have

Threaded Interrupts

- Similar to NAPI code
- Interrupt comes in, disable the device
- Thread is awoken
- Thread polls on device, processing packets
- Finished and enables device

Benchmarking

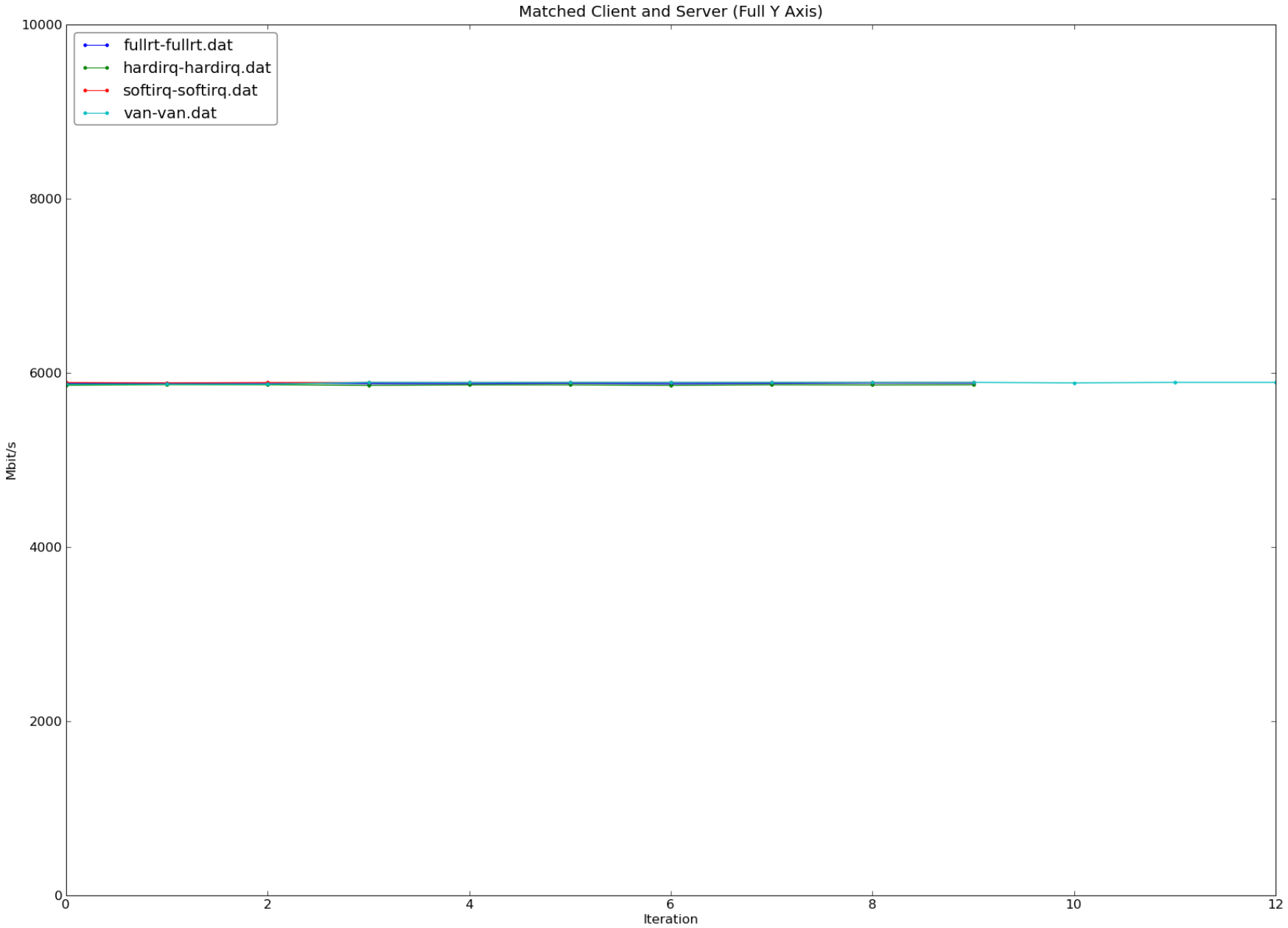
- On my 1Gb NICs, all kernels had same result in netperf
  - Saturated network
- 10 Gig NIC
  - Could not get working with the threaded infrastructure
    - I'm not a network device expert
  - Decided to just give up and test with the RT patch
    - This is far from optimizing the design

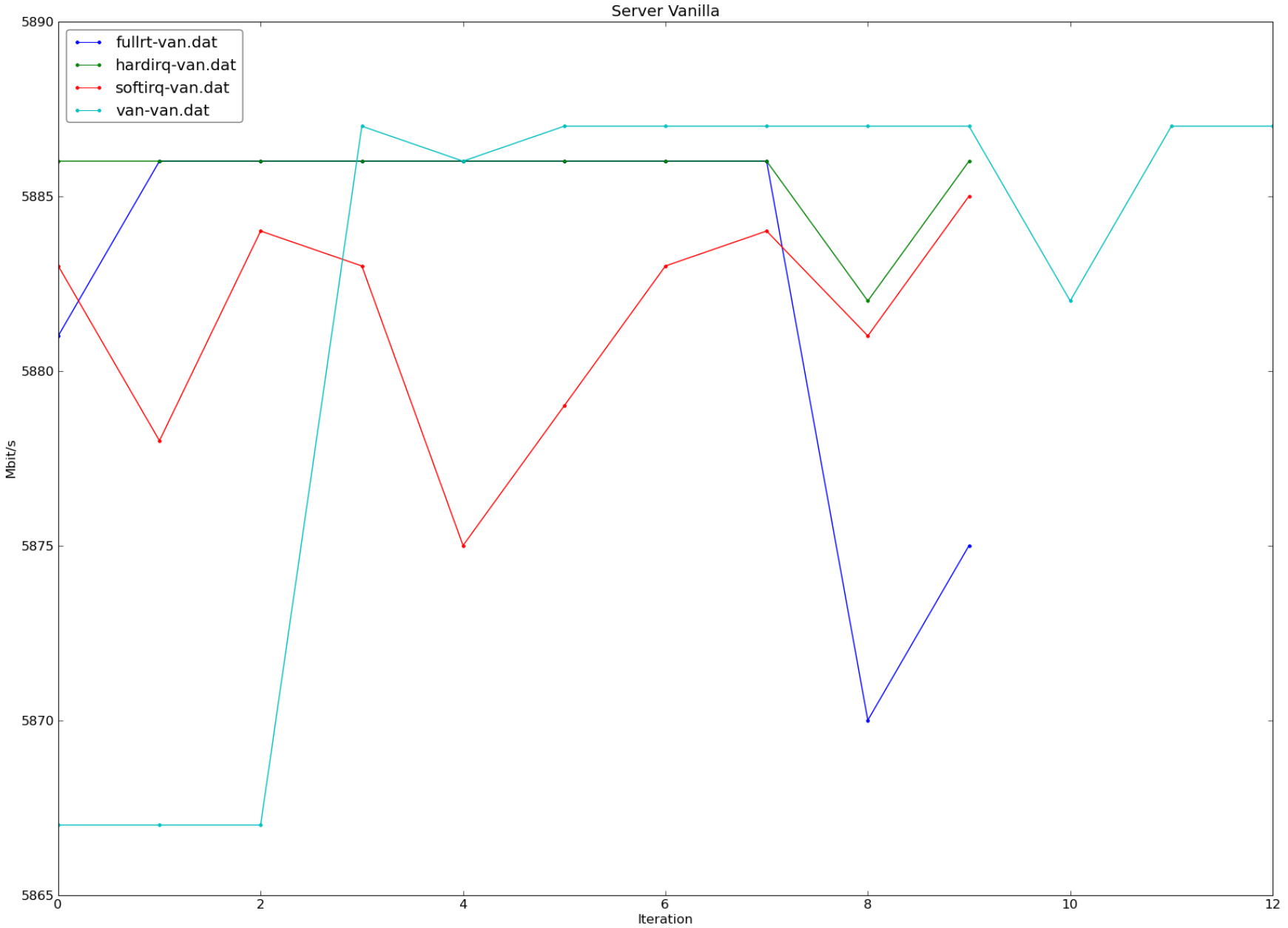
Benchmark Machine

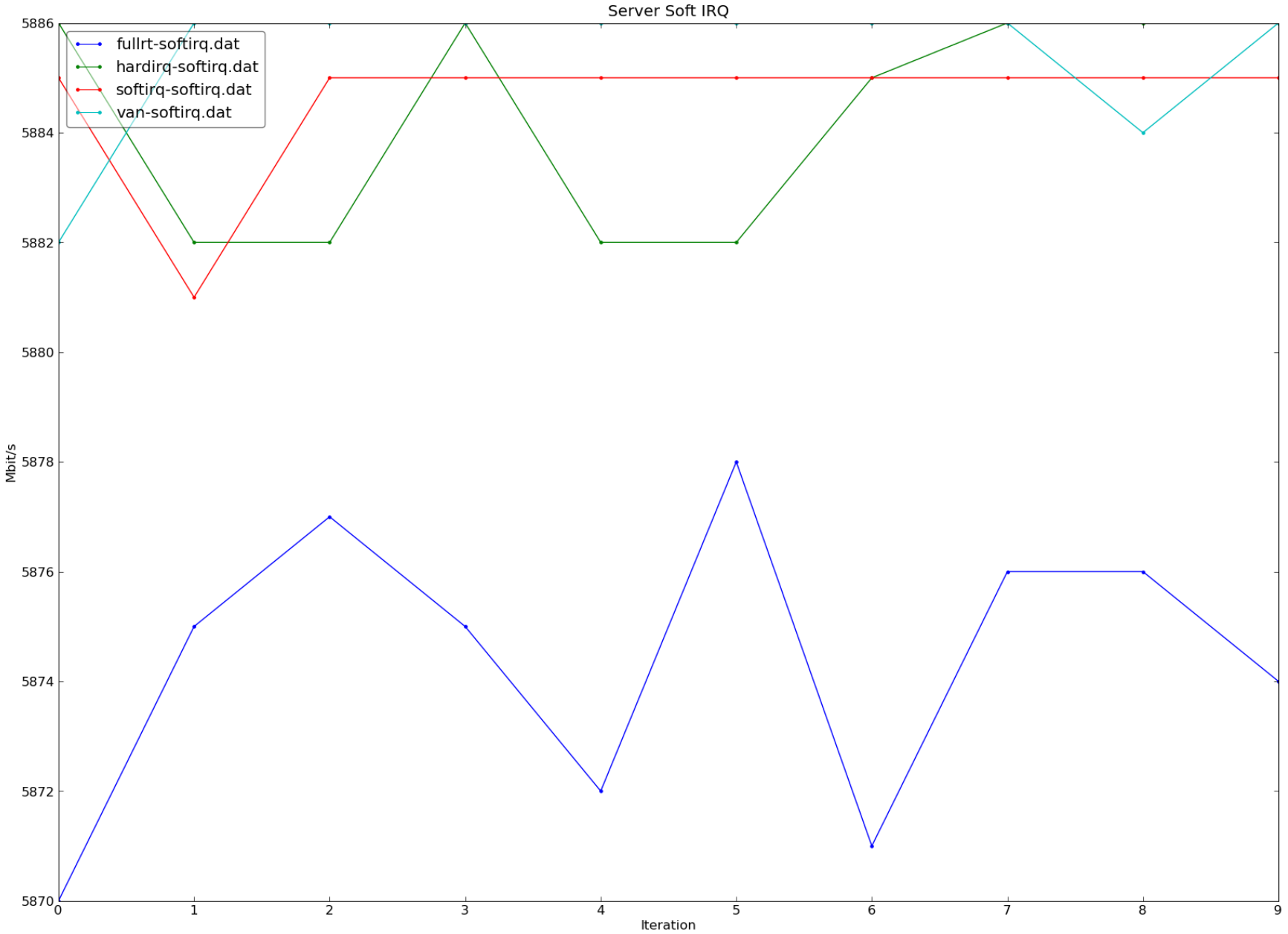
- Two Nehalem machines
- 2 x 4 core Intel Xeons
- Chelsio Communications Inc T310 10GbE Single Port Adapter

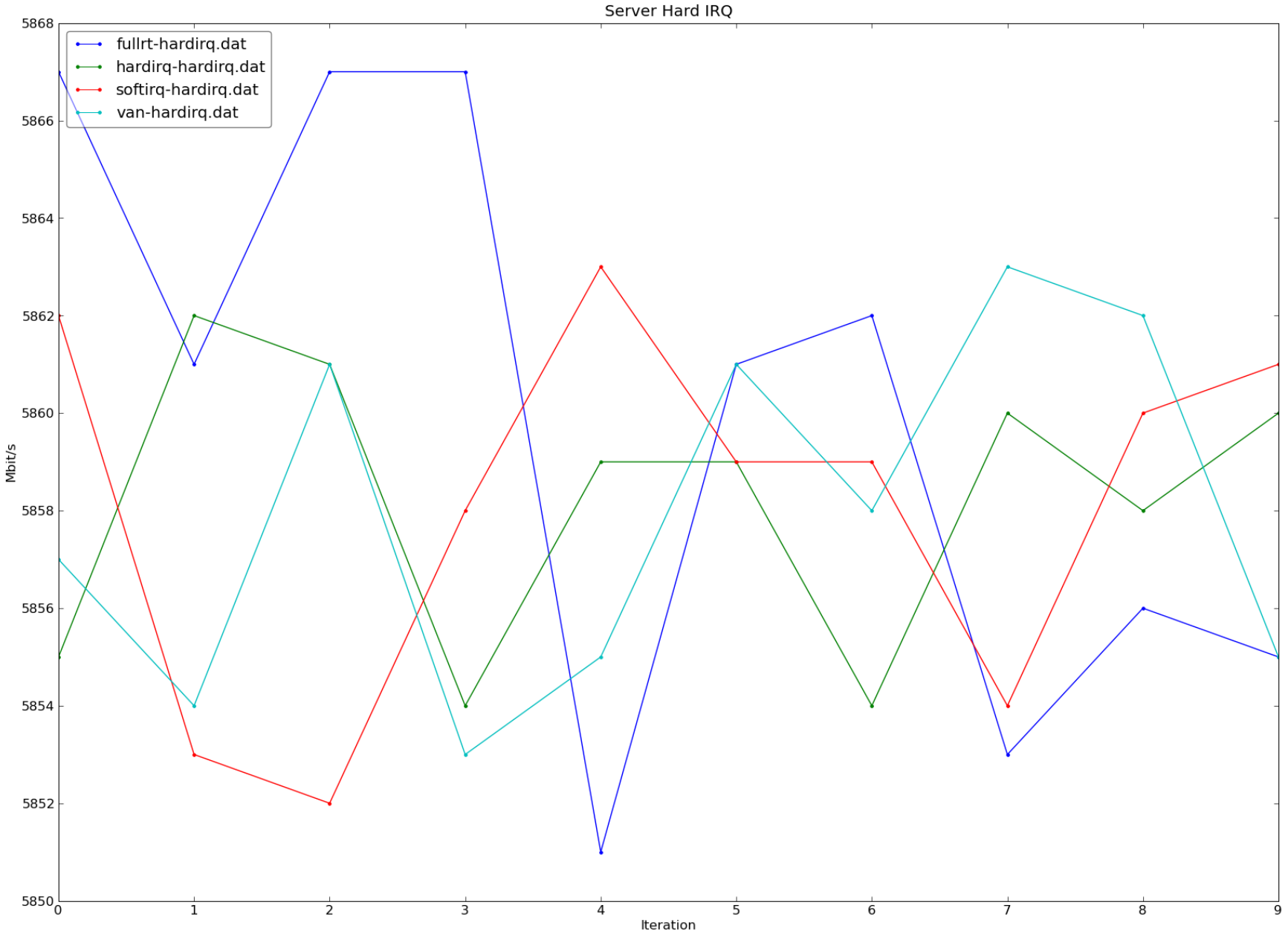
Benchmark

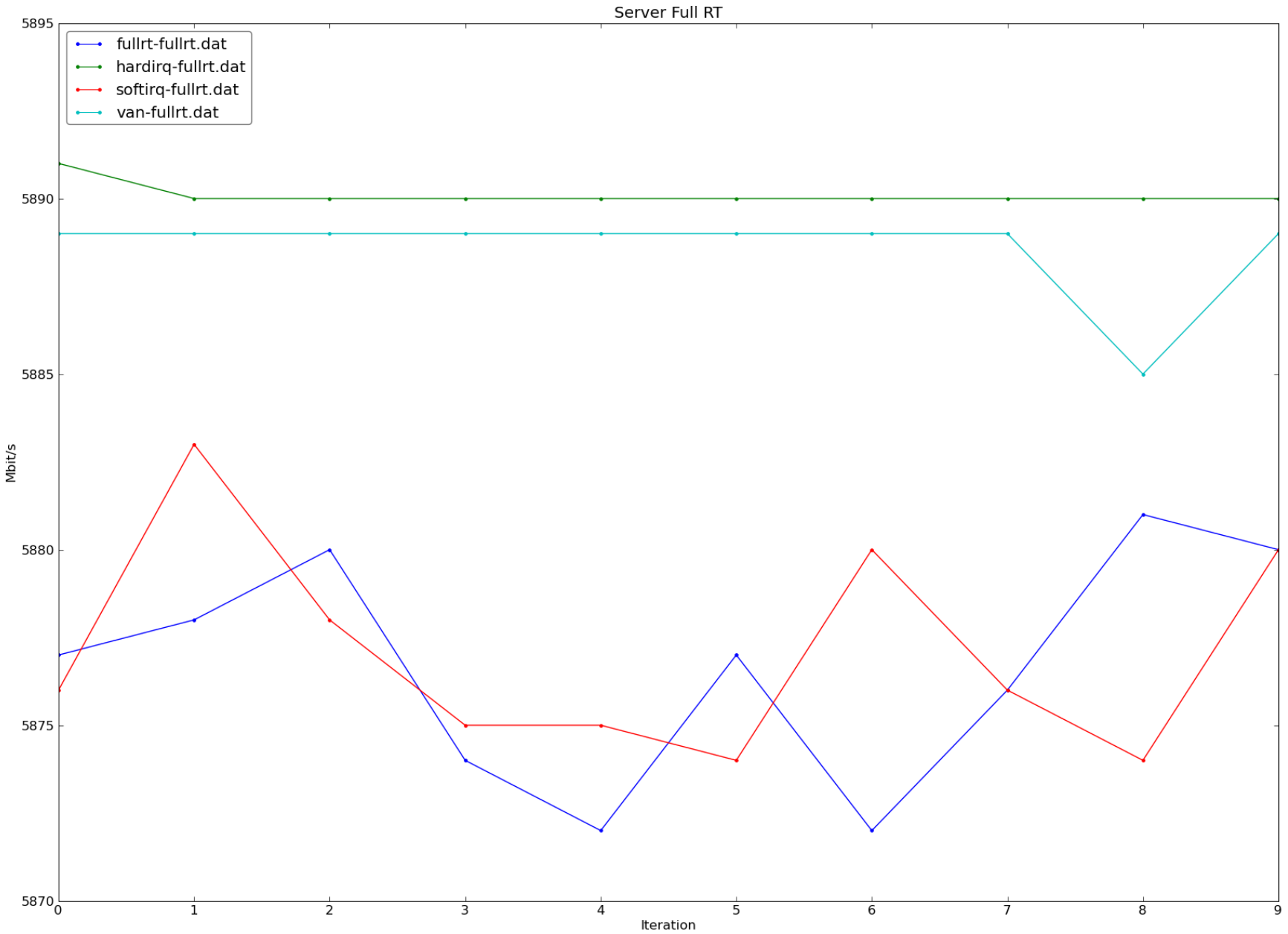
- Ran netperf -H host
- Red Hat Enterprise Linux 5.4
- Full Preempt Real Time Patch (PREEMPT_RT)
  - 2.6.31-rt10
- Hard and Soft interrupts as threads
  - RT patch without PREEMPT_RT enabled
- Soft interrupts as threads
  - RT patch with only PREEMPT_SOFTIRQ set
- Vanilla Linus Kernel
  - 2.6.31 downloaded from kernel.org

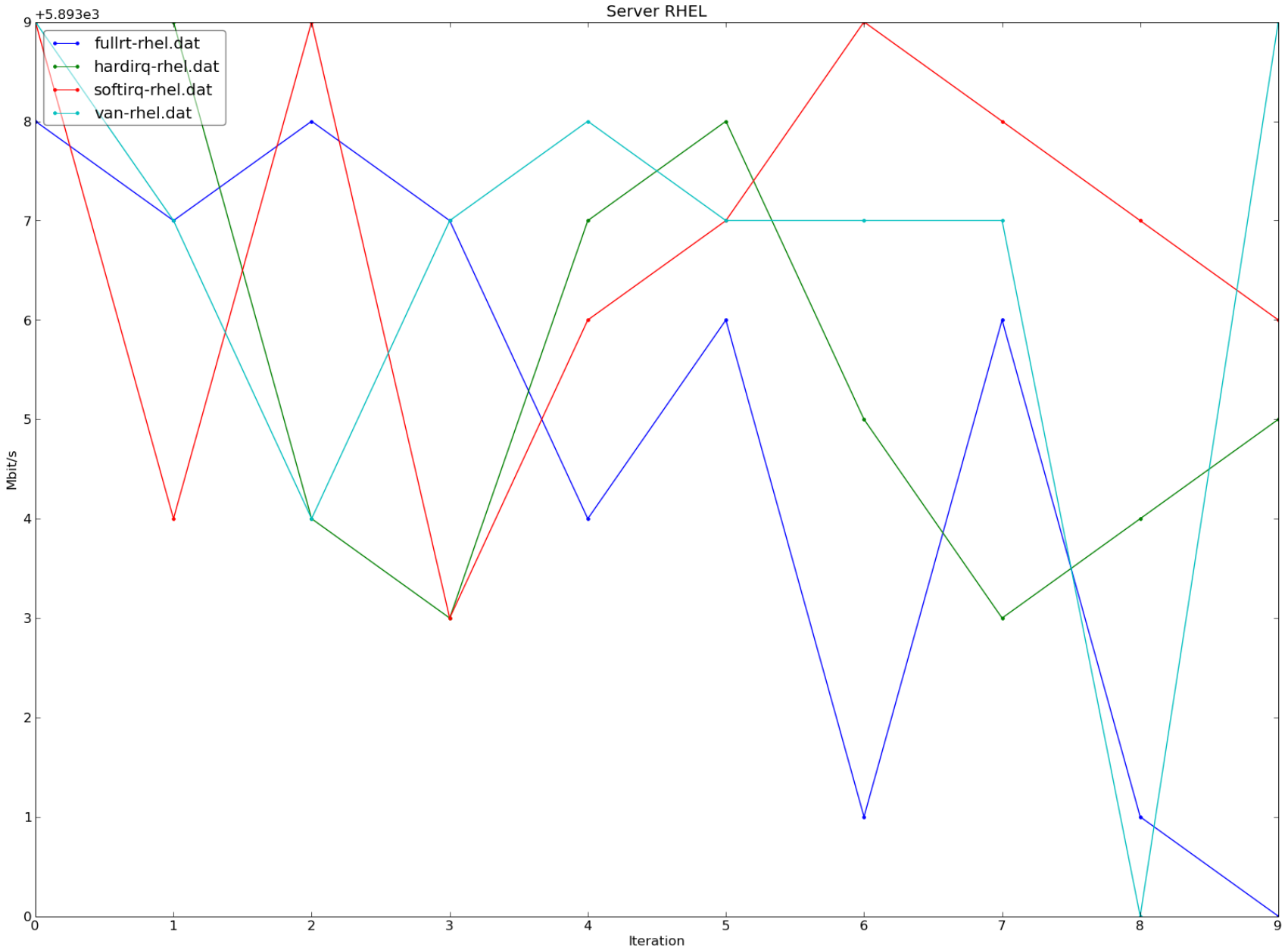

Matched Client and Server (Full Y Axis)
fullrt-fullrt.dat
hardirq-hardirq.dat
softirq-softirq.dat
van-van.dat
Mbit/s
Iteration
0
2000
4000
6000
8000
10000
0
2
4
6
8
10
12

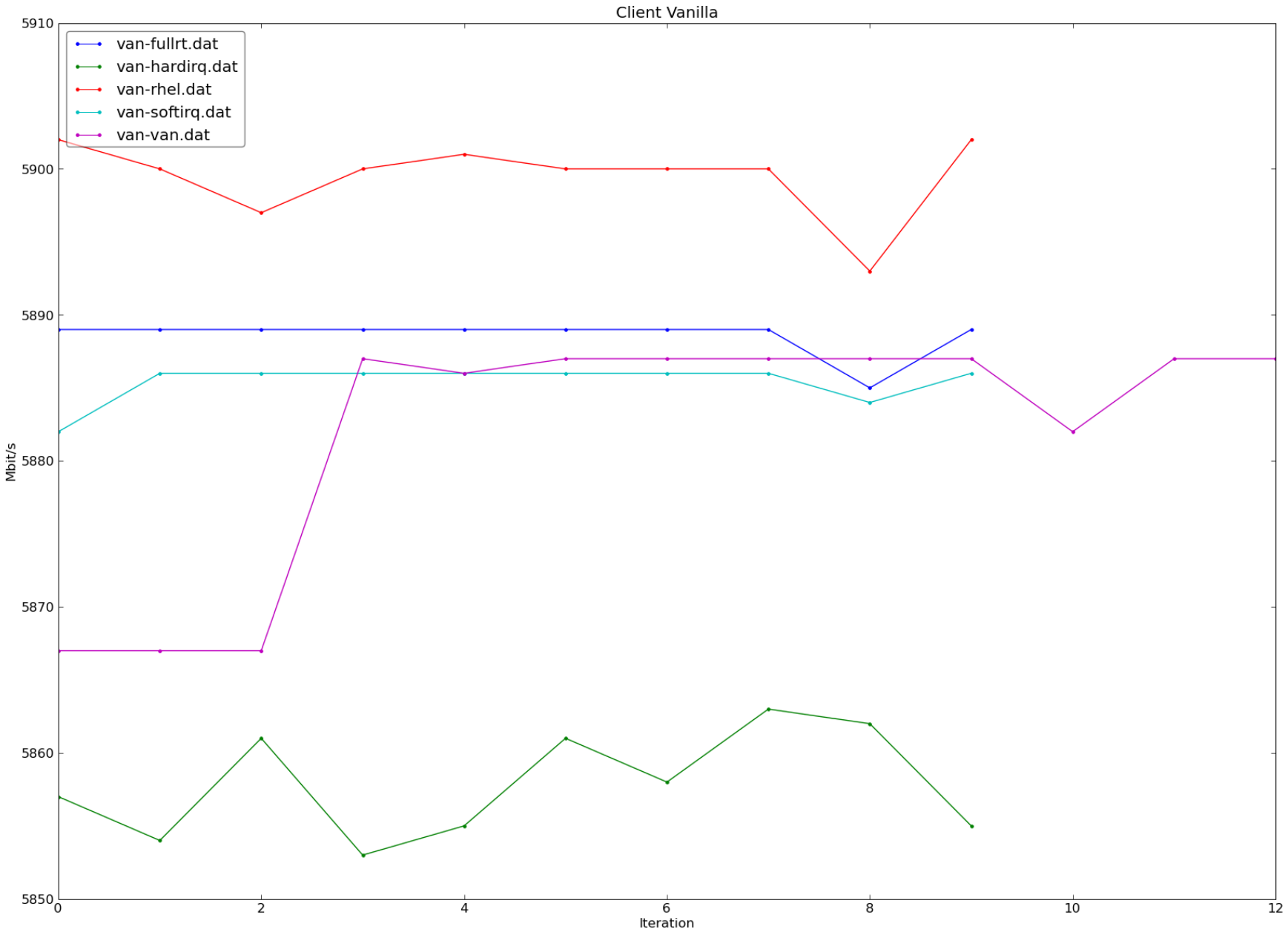

Server Vanilla
fullrt-van.dat
hardirq-van.dat
softirq-van.dat
van-van.dat
Mbit/s
Iteration
5890
5885
5880
5875
5870
5865
0
2
4
6
8
10
12

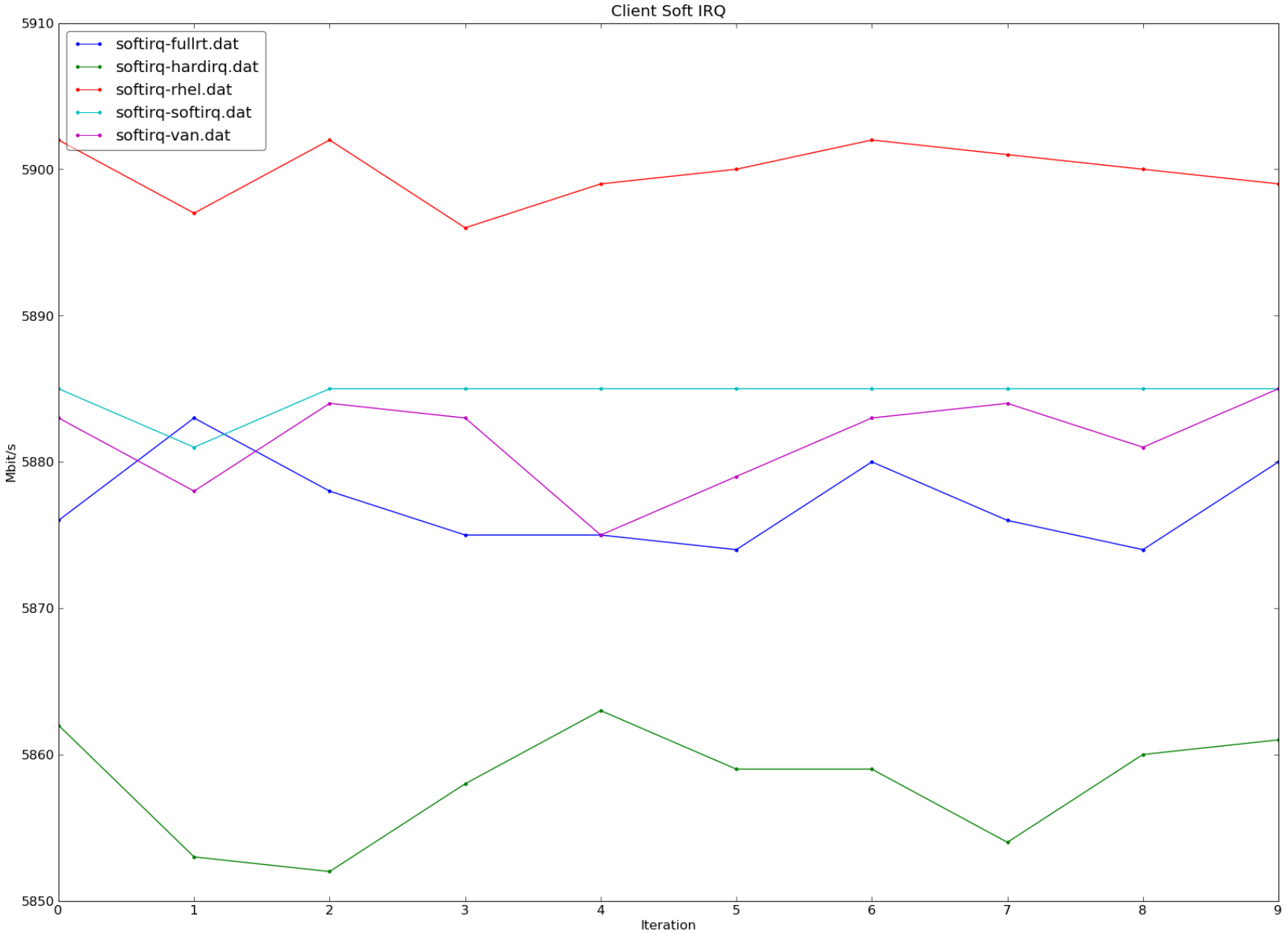

Server Soft IRQ
fullrt-softirq.dat
hardirq-softirq.dat
softirq-softirq.dat
van-softirq.dat
Mbit/s
Iteration
5886
5884
5882
5880
5878
5876
5874
5872
5870
0
1
2
3
4
5
6
7
8
9

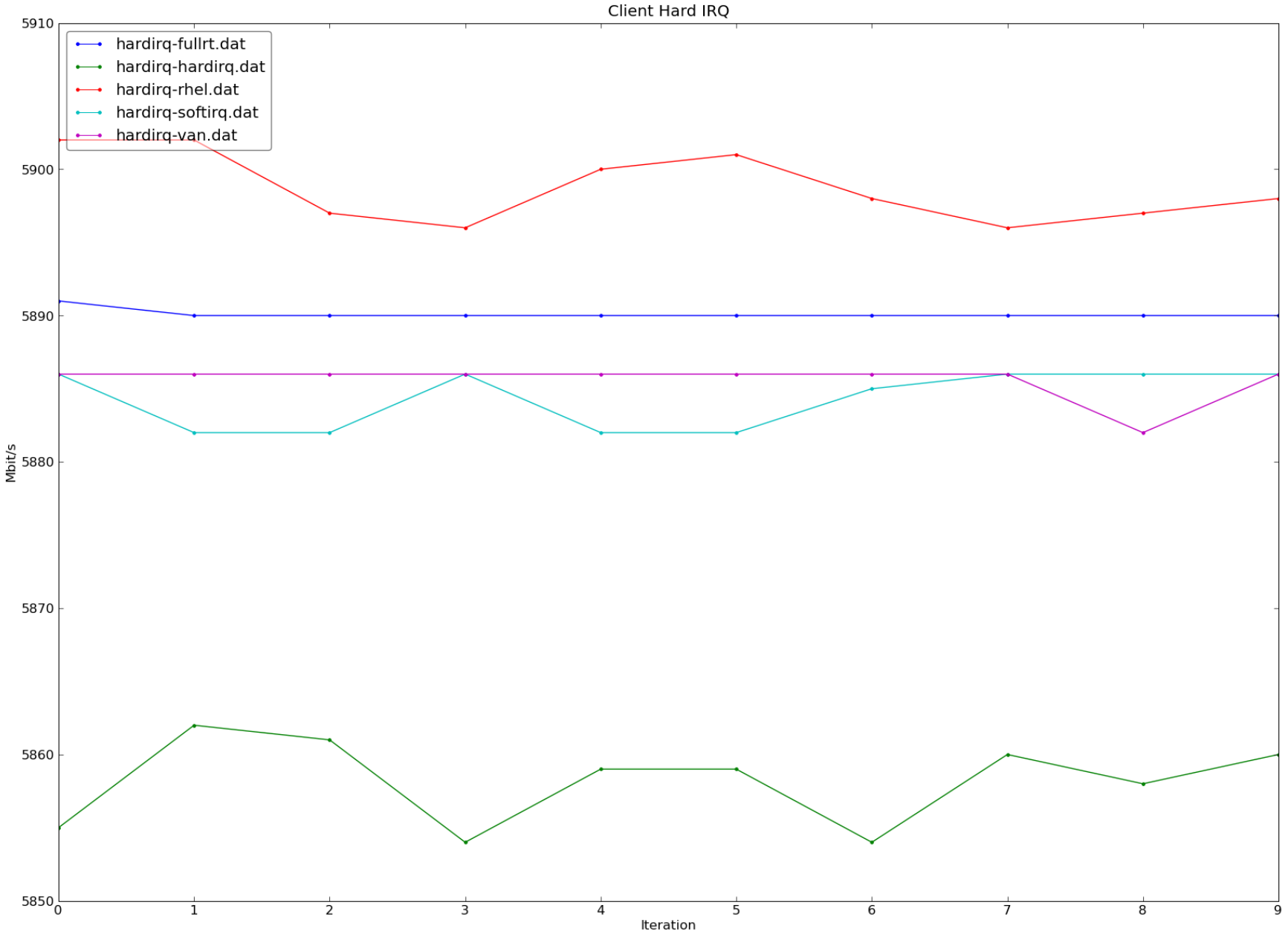

Server Hard IRQ
fullrt-hardirq.dat
hardirq-hardirq.dat
softirq-hardirq.dat
van-hardirq.dat
Mbit/s
Iteration
5868
5866
5864
5862
5860
5858
5856
5854
5852
5850
0
1
2
3
4
5
6
7
8
9

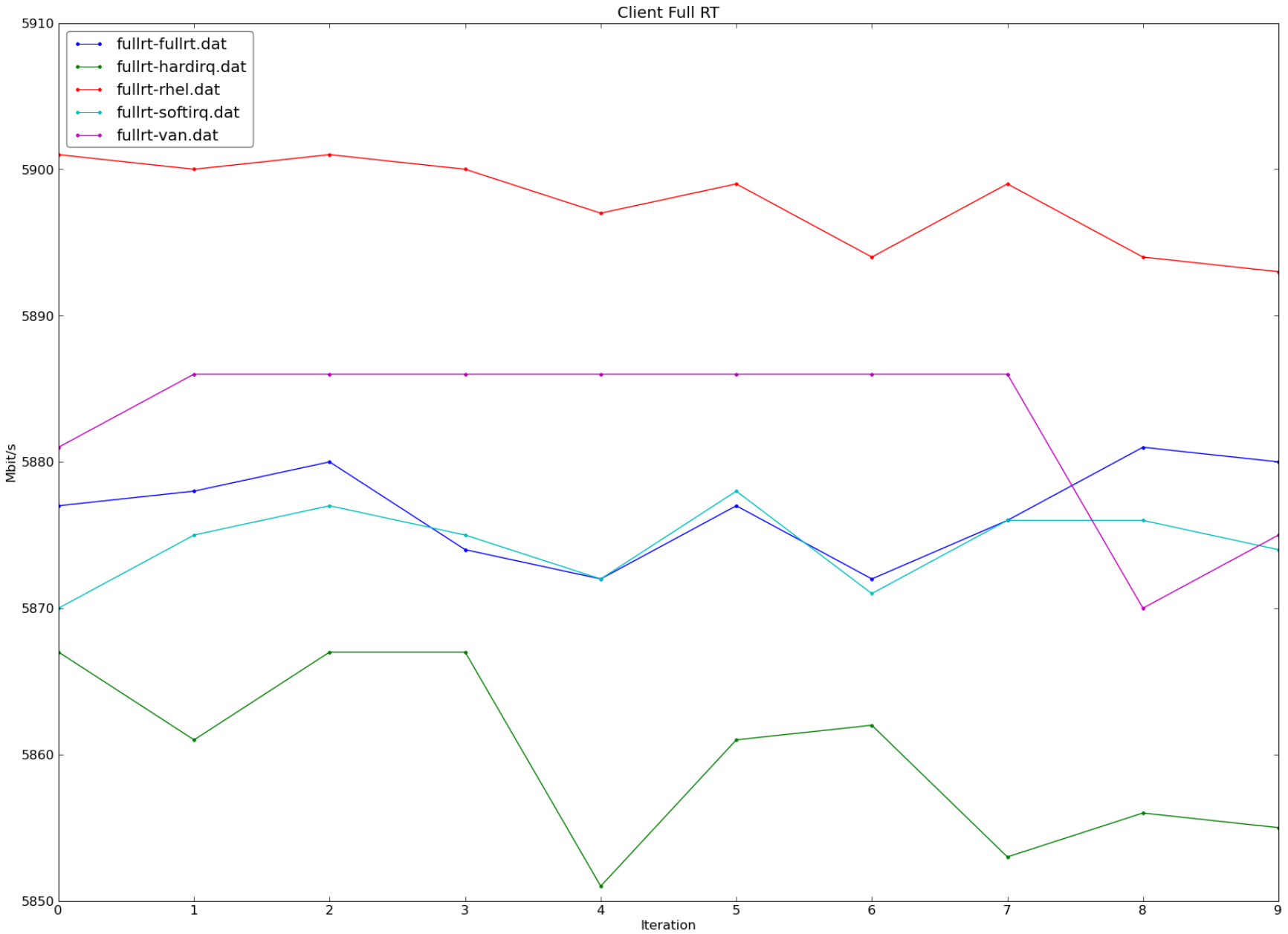

Server Full RT
fullrt-fullrt.dat
hardirq-fullrt.dat
softirq-fullrt.dat
van-fullrt.dat
5895
5890
5885
5880
5875
5870
Mbit/s
0
1
2
3
4
5
6
7
8
9
Iteration

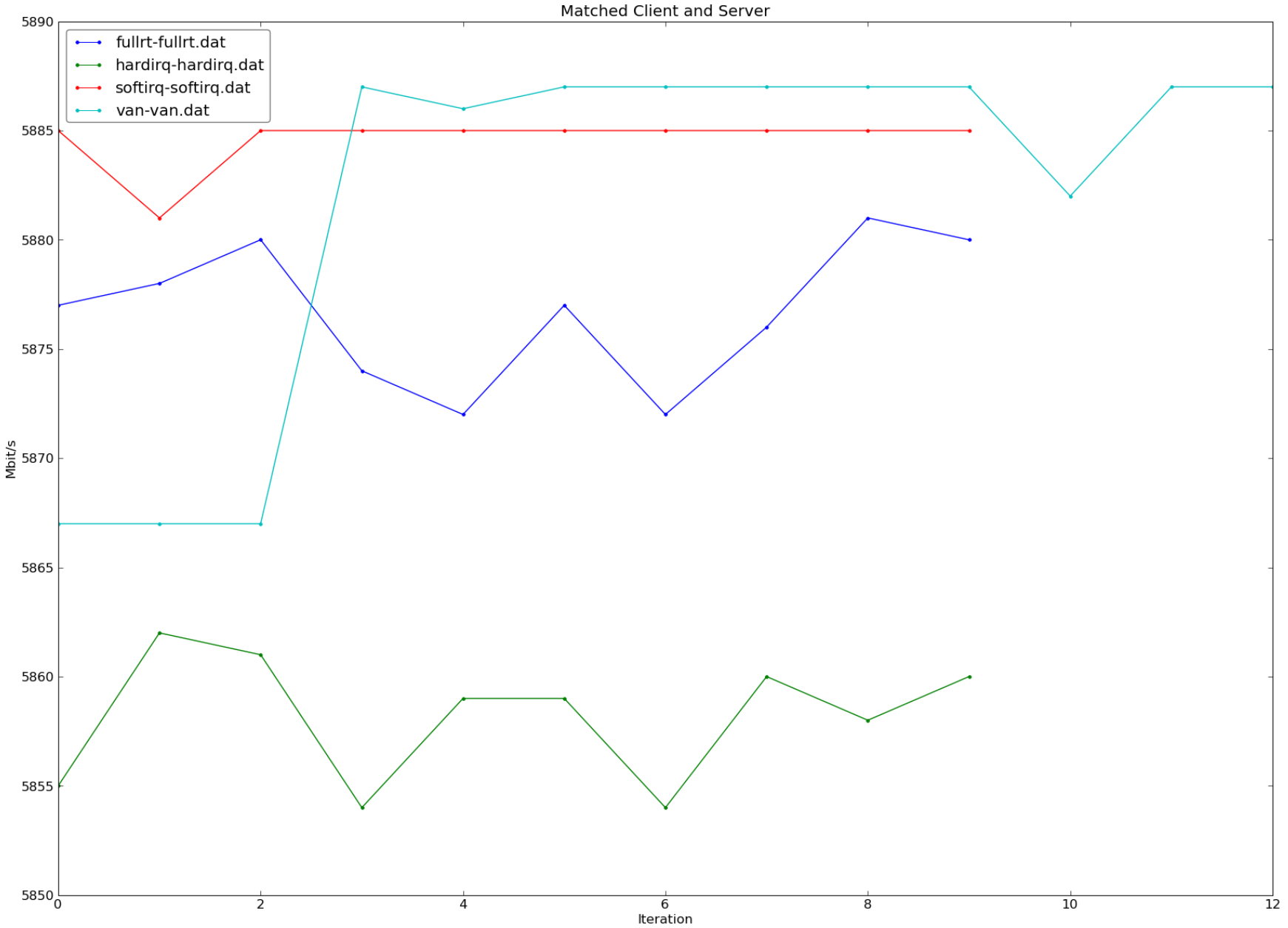

Server RHEL
+5.893e3
fullrt-rhel.dat
hardirq-rhel.dat
softirq-rhel.dat
van-rhel.dat
Mbit/s
Iteration
0
1
2
3
4
5
6
7
8
9


Client Vanilla
van-fullrt.dat
van-hardirq.dat
van-rhel.dat
van-softirq.dat
van-van.dat
Mbit/s
Iteration
5910
5900
5890
5880
5870
5860
5850
0
2
4
6
8
10
12


Client Soft IRQ
softirq-fullrt.dat
softirq-hardirq.dat
softirq-rhel.dat
softirq-softirq.dat
softirq-van.dat
Mbit/s
Iteration
5910
5900
5890
5880
5870
5860
5850
0
1
2
3
4
5
6
7
8
9


Client Hard IRQ
hardirq-fullrt.dat
hardirq-hardirq.dat
hardirq-rhel.dat
hardirq-softirq.dat
hardirq-van.dat
Mbit/s
Iteration
5910
5900
5890
5880
5870
5860
5850
0
1
2
3
4
5
6
7
8
9


Client Full RT
fullrt-fullrt.dat
fullrt-hardirq.dat
fullrt-rhel.dat
fullrt-softirq.dat
fullrt-van.dat
Mbit/s
Iteration
5910
5900
5890
5880
5870
5860
5850
0
1
2
3
4
5
6
7
8
9


Matched Client and Server
fullrt-fullrt.dat
hardirq-hardirq.dat
softirq-softirq.dat
van-van.dat
Mbit/s
Iteration
5890
5885
5880
5875
5870
5865
5860
5855
5850
0
2
4
6
8
10
12

Cons

- Horrible latency on server config (no kernel preemption)
- Redesign of network infrastructure

Pros

- Can schedule in handler
- Easier locking algorithms
- Prioritize interrupts
- Smoother desktop experience

Discussion

Threaded Network Interrupts

Steven Rostedt
srostedt@redhat.com
<rostedt@goodmis.org>
http://people.redhat.com/srostedt
network-thread-irq.odp

Disclaimer

I'm not a network guru

I had to fight for the hardware I have

Threaded Interrupts

- Similar to NAPI code
- Interrupt comes in, disable the device
- Thread is awoken
- Thread polls on device, processing packets
- Finished and enables device

Benchmarking

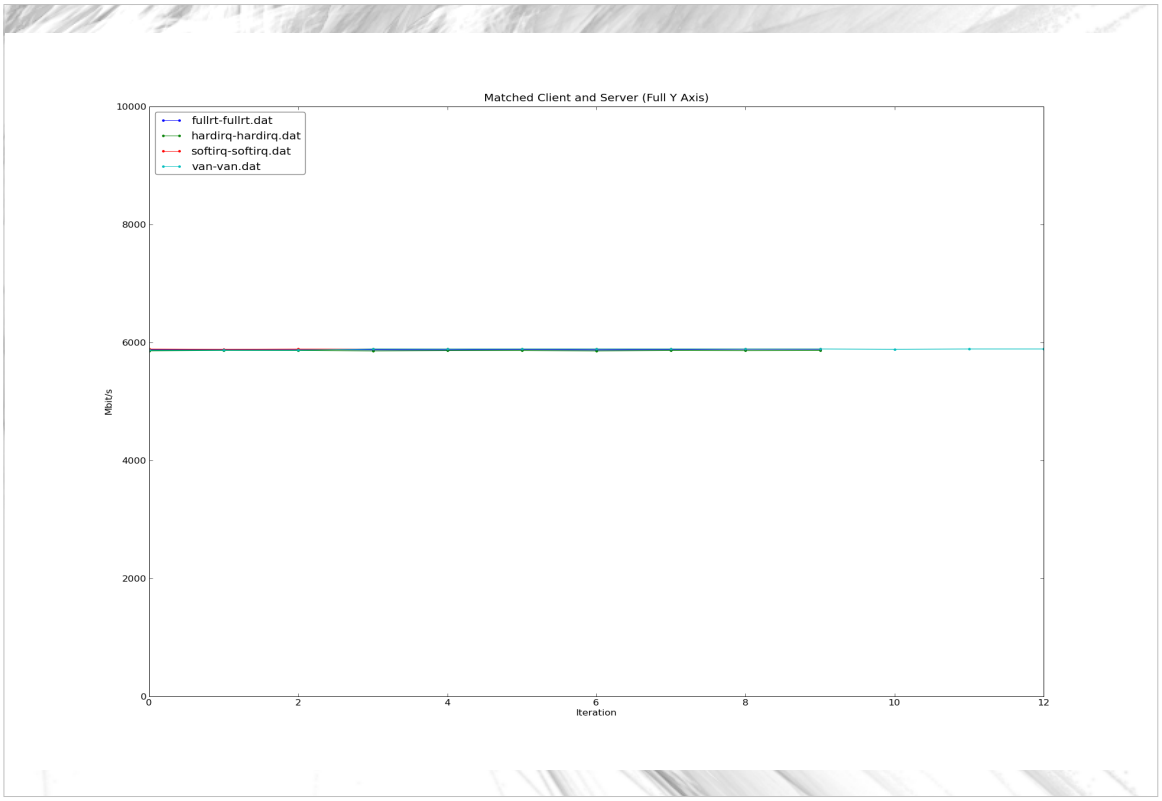
- On my 1Gb NICs, all kernels had same result in netperf
  - Saturated network
- 10 Gig NIC
  - Could not get working with the threaded infrastructure
    - I'm not a network device expert
  - Decided to just give up and test with the RT patch
    - This is far from optimizing the design

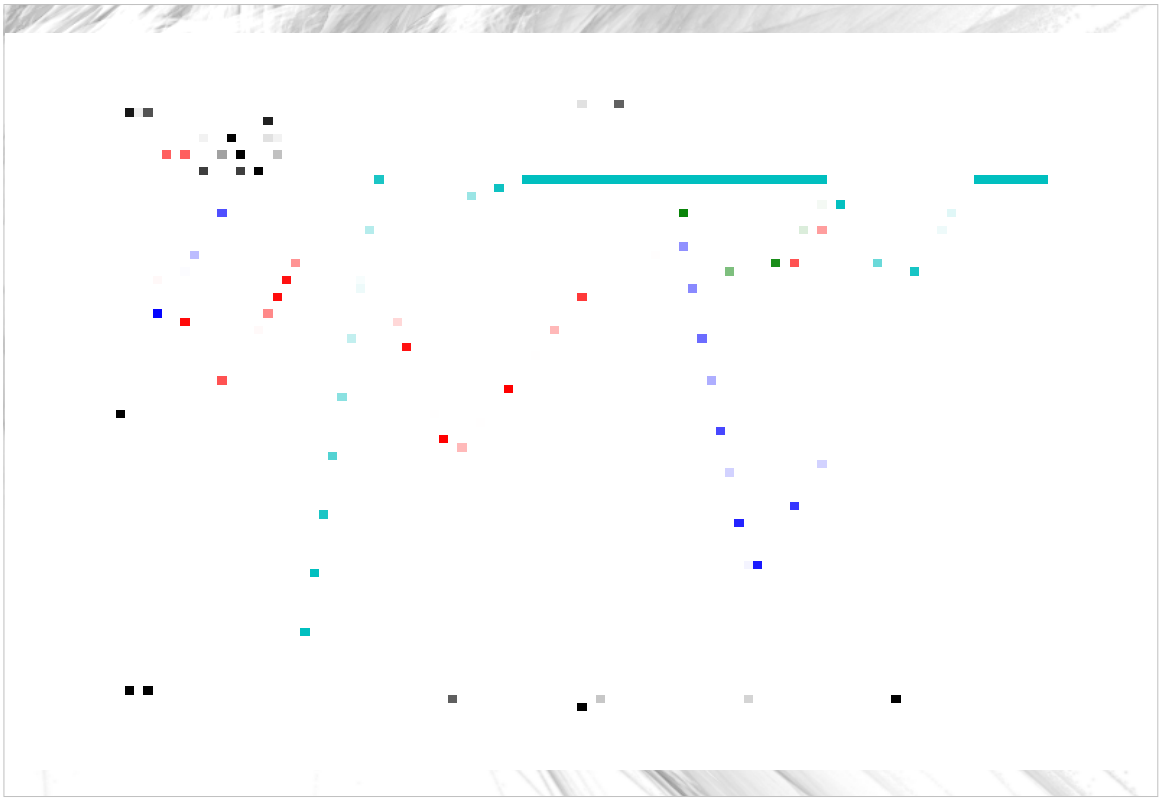
Benchmark Machine

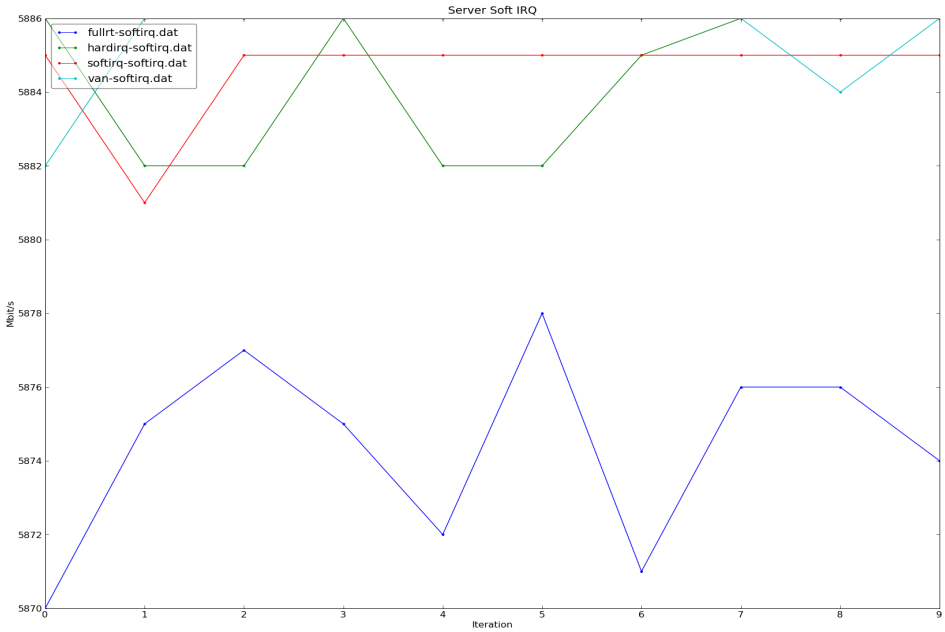
- Two Nehalem machines
- 2 x 4 core Intel Xeons
- Chelsio Communications Inc T310 10GbE Single Port Adapter

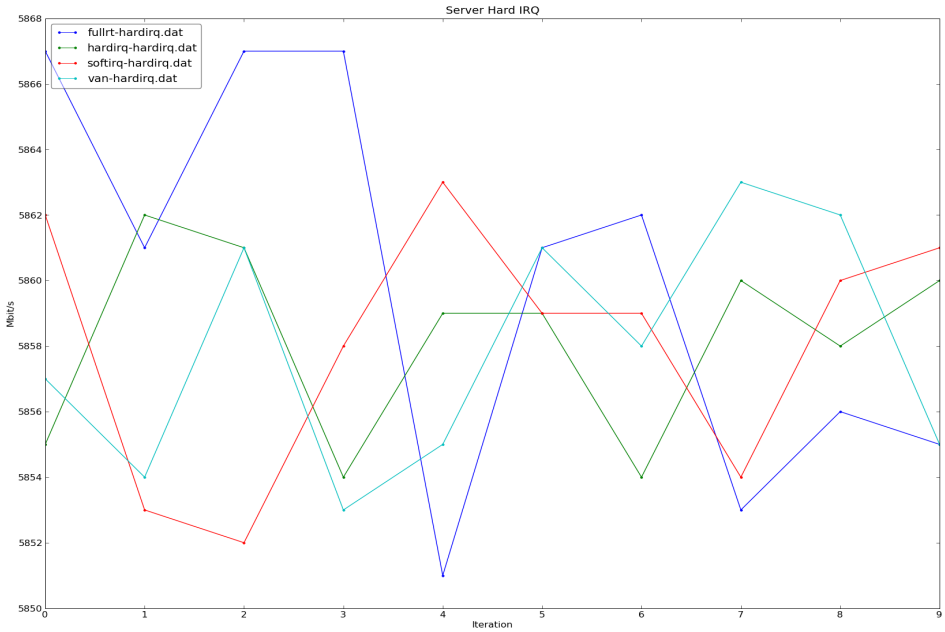
Benchmark

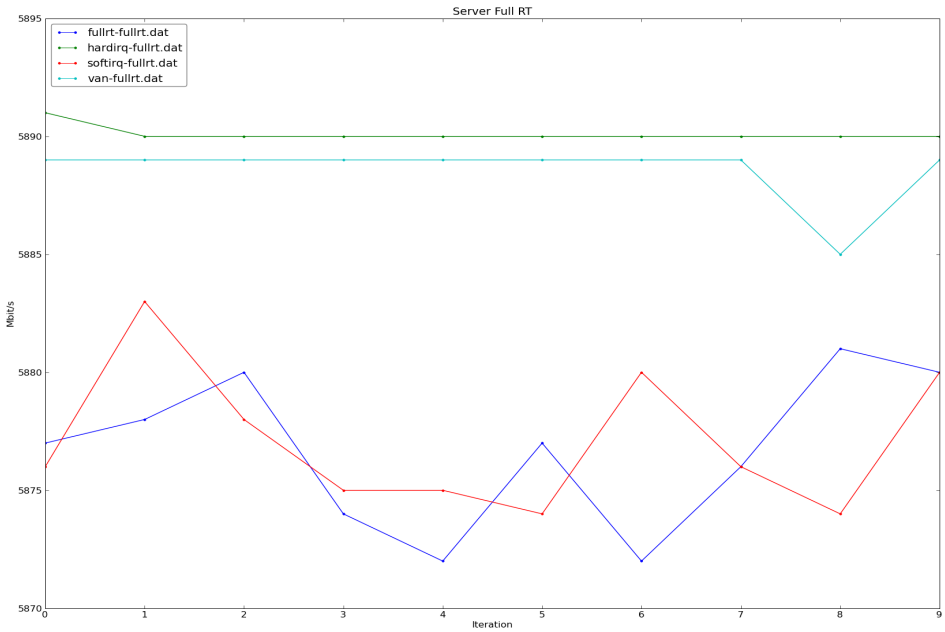
- Ran netperf -H host
- Red Hat Enterprise Linux 5.4
- Full Preempt Real Time Patch (PREEMPT_RT)
  - 2.6.31-rt10
- Hard and Soft interrupts as threads
  - RT patch without PREEMPT_RT enabled
- Soft interrupts as threads
  - RT patch with only PREEMPT_SOFTIRQ set
- Vanilla Linus Kernel
  - 2.6.31 downloaded from kernel.org

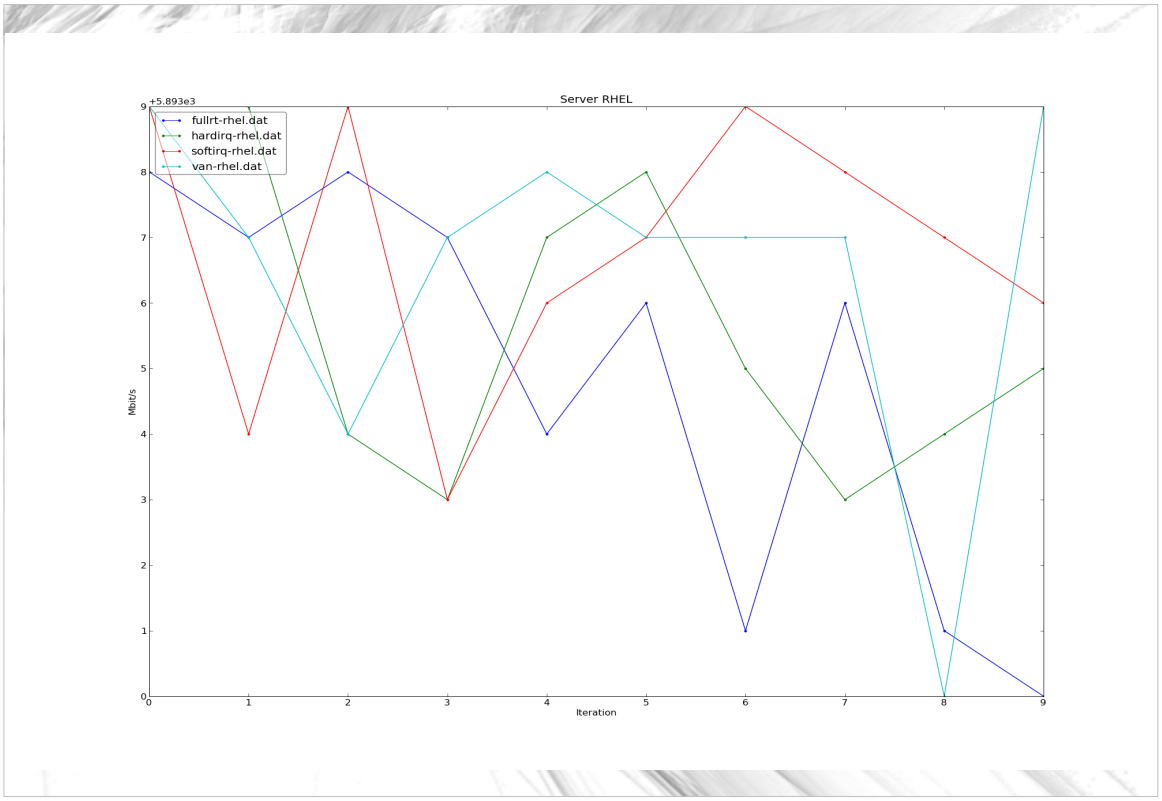

Matched Client and Server (Full Y Axis)
fullrt-fullrt.dat
hardirq-hardirq.dat
softirq-softirq.dat
van-van.dat
Mbit/s
Iteration
0
2000
4000
6000
8000
10000
0
2
4
6
8
10
12

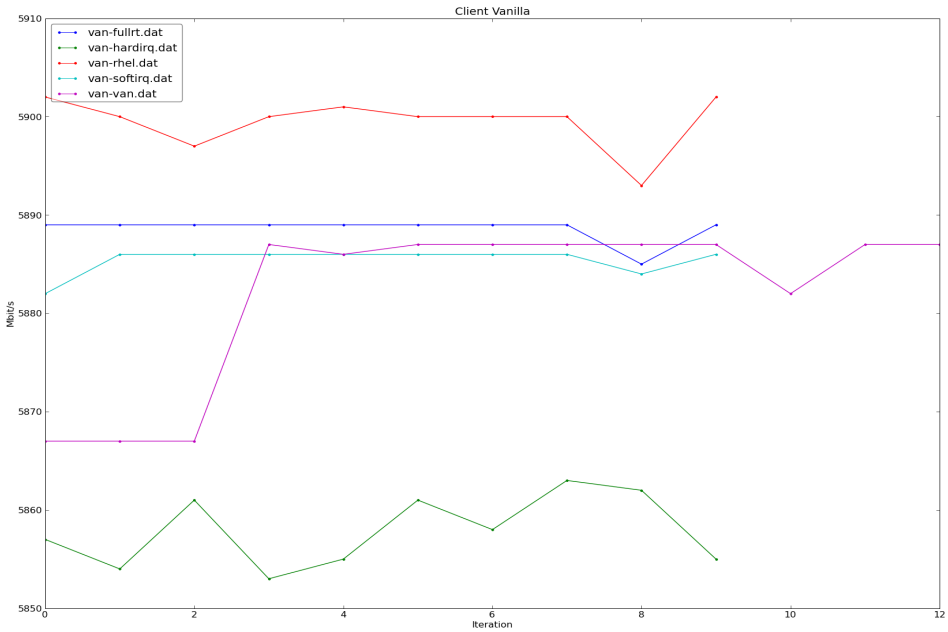


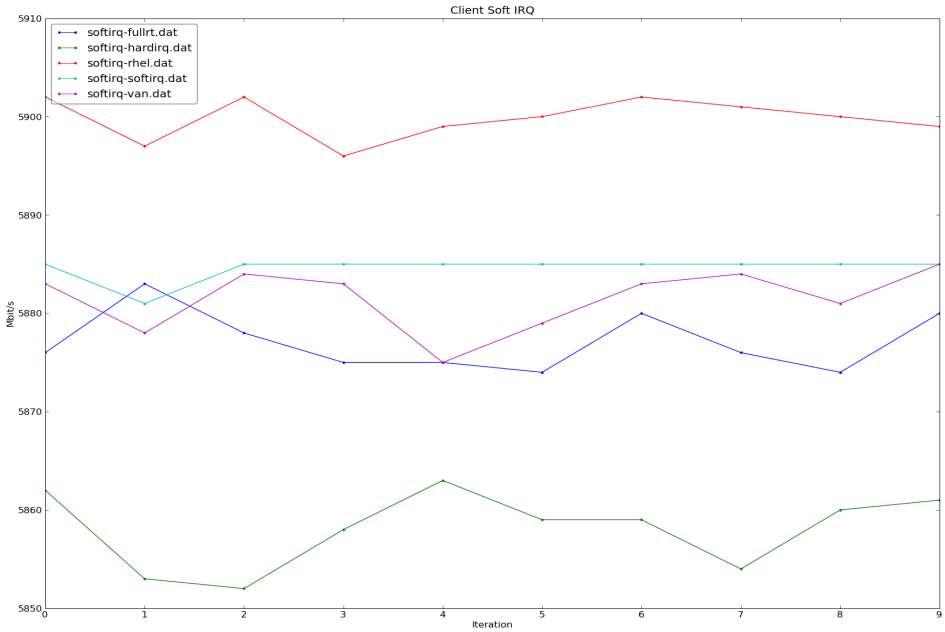

Server Soft IRQ
fullrt-softirq.dat
hardirq-softirq.dat
softirq-softirq.dat
van-softirq.dat
Mbit/s
Iteration
5886
5884
5882
5880
5878
5876
5874
5872
5870
0
1
2
3
4
5
6
7
8
9

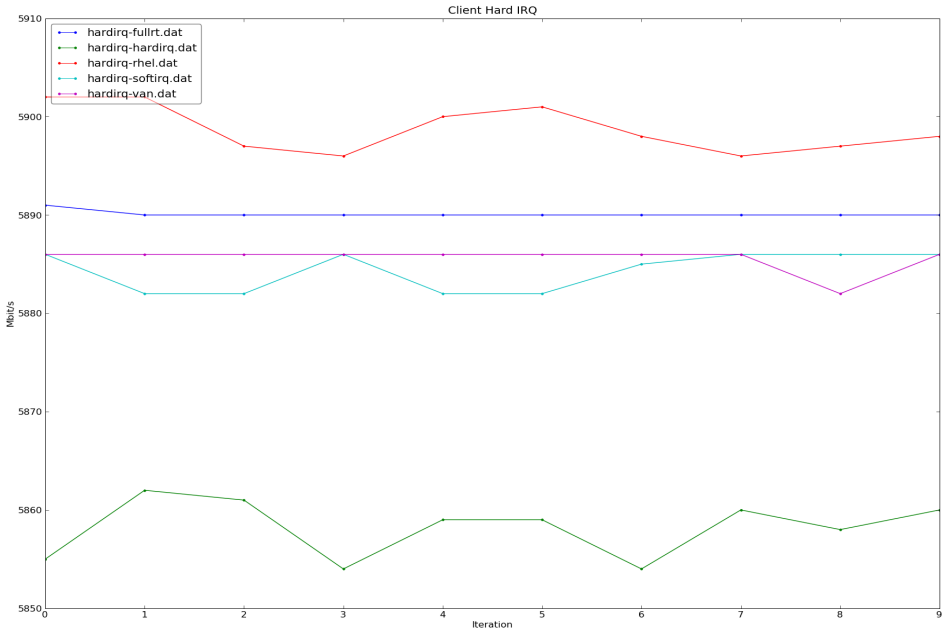

Server Hard IRQ
fullrt-hardirq.dat
hardirq-hardirq.dat
softirq-hardirq.dat
van-hardirq.dat
Mbit/s
Iteration
5868
5866
5864
5862
5860
5858
5856
5854
5852
5850
0
1
2
3
4
5
6
7
8
9

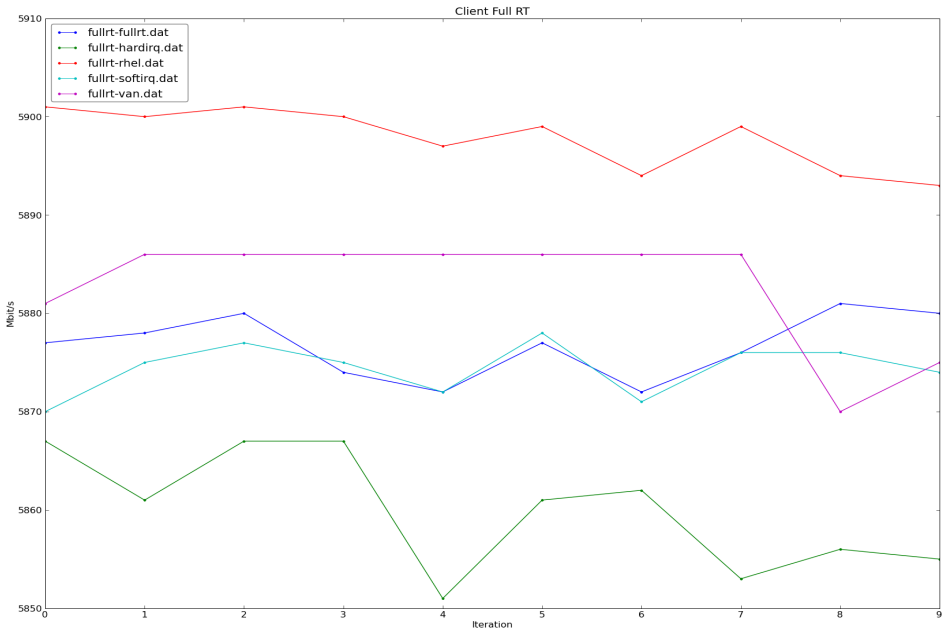

Server Full RT
fullrt-fullrt.dat
hardirq-fullrt.dat
softirq-fullrt.dat
van-fullrt.dat
Mbit/s
Iteration
5895
5890
5885
5880
5875
5870
0
1
2
3
4
5
6
7
8
9

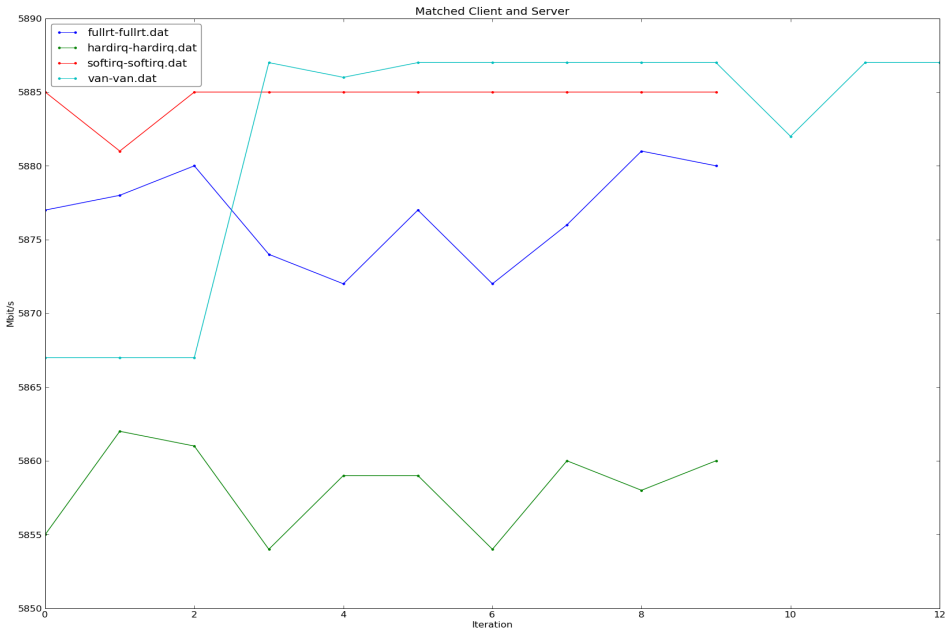

Server RHEL
+5.893e3
fullrt-rhel.dat
hardirq-rhel.dat
softirq-rhel.dat
van-rhel.dat
Mbit/s
Iteration
0
1
2
3
4
5
6
7
8
9


Client Vanilla
van-fullrt.dat
van-hardirq.dat
van-rhel.dat
van-softirq.dat
van-van.dat
Mbit/s
Iteration
5910
5900
5890
5880
5870
5860
5850
0
2
4
6
8
10
12


Client Soft IRQ
softirq-fullrt.dat
softirq-hardirq.dat
softirq-rhel.dat
softirq-softirq.dat
softirq-van.dat
Mbit/s
Iteration
5910
5900
5890
5880
5870
5860
5850
0
1
2
3
4
5
6
7
8
9


Client Hard IRQ
hardirq-fullrt.dat
hardirq-hardirq.dat
hardirq-rhel.dat
hardirq-softirq.dat
hardirq-van.dat
Mbit/s
Iteration
5910
5900
5890
5880
5870
5860
5850
0
1
2
3
4
5
6
7
8
9


Client Full RT
fullrt-fullrt.dat
fullrt-hardirq.dat
fullrt-rhel.dat
fullrt-softirq.dat
fullrt-van.dat
Mbit/s
Iteration
5910
5900
5890
5880
5870
5860
5850
0
1
2
3
4
5
6
7
8
9


Matched Client and Server
fullrt-fullrt.dat
hardirq-hardirq.dat
softirq-softirq.dat
van-van.dat
Mbit/s
Iteration
5890
5885
5880
5875
5870
5865
5860
5855
5850
0
2
4
6
8
10
12

Cons

- Horrible latency on server config (no kernel preemption)
- Redesign of network infrastructure

Pros

- Can schedule in handler
- Easier locking algorithms
- Prioritize interrupts
- Smoother desktop experience

Discussion