

TCP Behavior of BGP

Cisco NAG

2012.10.10

Randy Bush <randy@psg.com>

Mark Allman <mallman@icir.org>

Keyur Patel <keyupate@cisco.com>

Balaji Pitta Venkatachalapathy <bvenkata@cisco.com>

Manoj Pandey <mpandey@cisco.com>

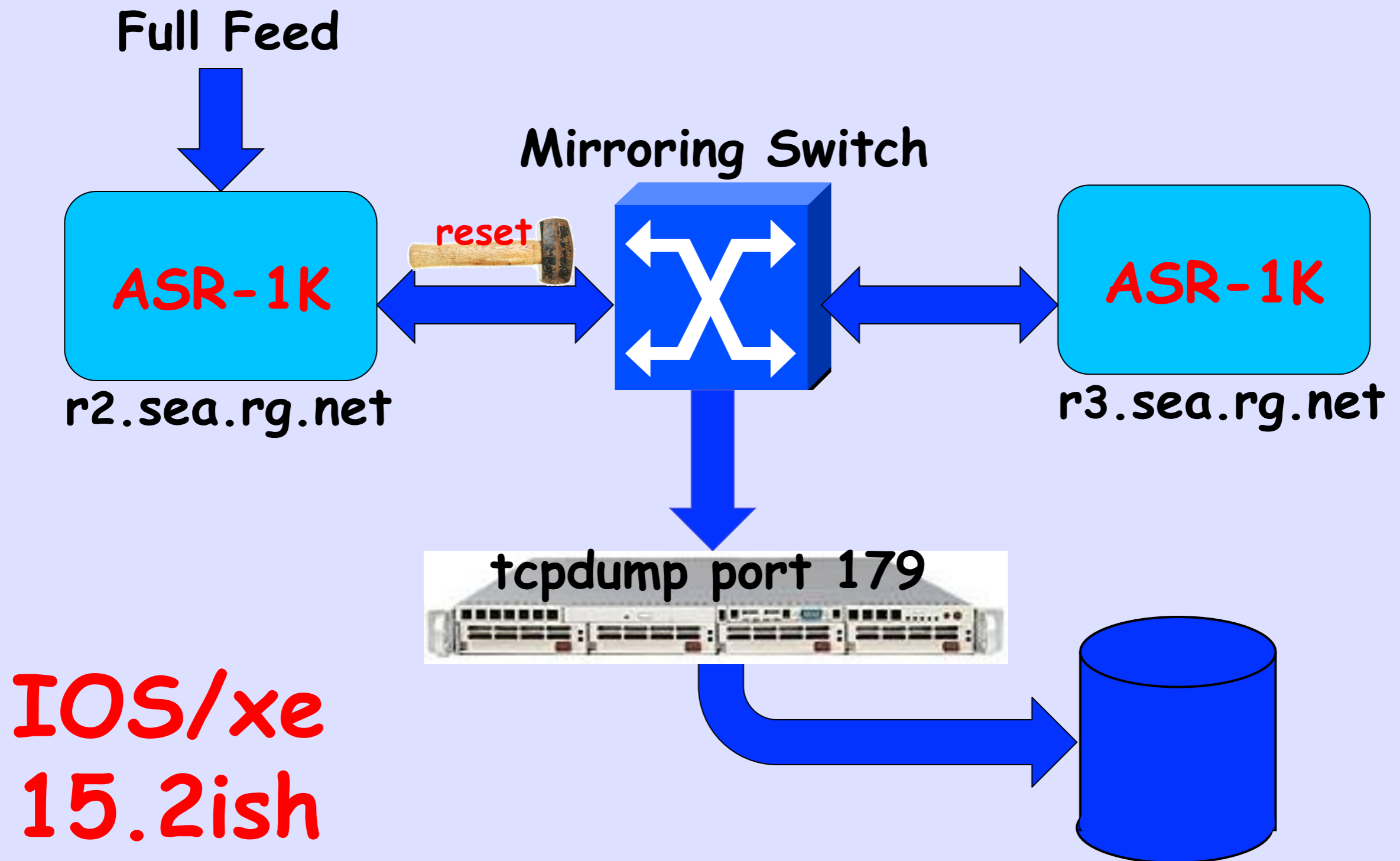
Vern Paxson <vern@icir.org>

Cristel Pelsser <cristel@iij.ad.jp>

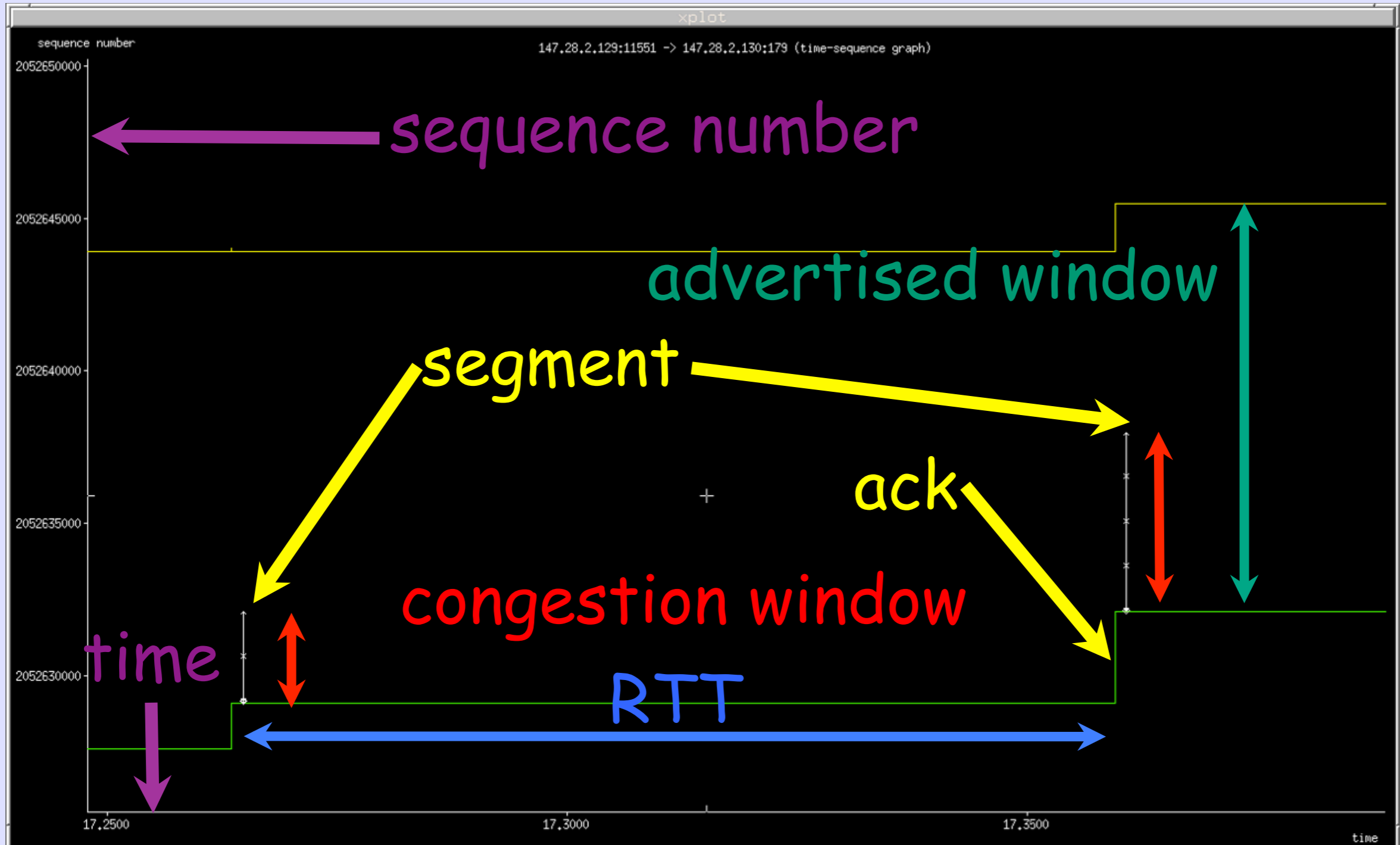
Ed Kern <ejk@cisco.com>

Goal:
Can We Make
BGP Convergence
Even Faster?

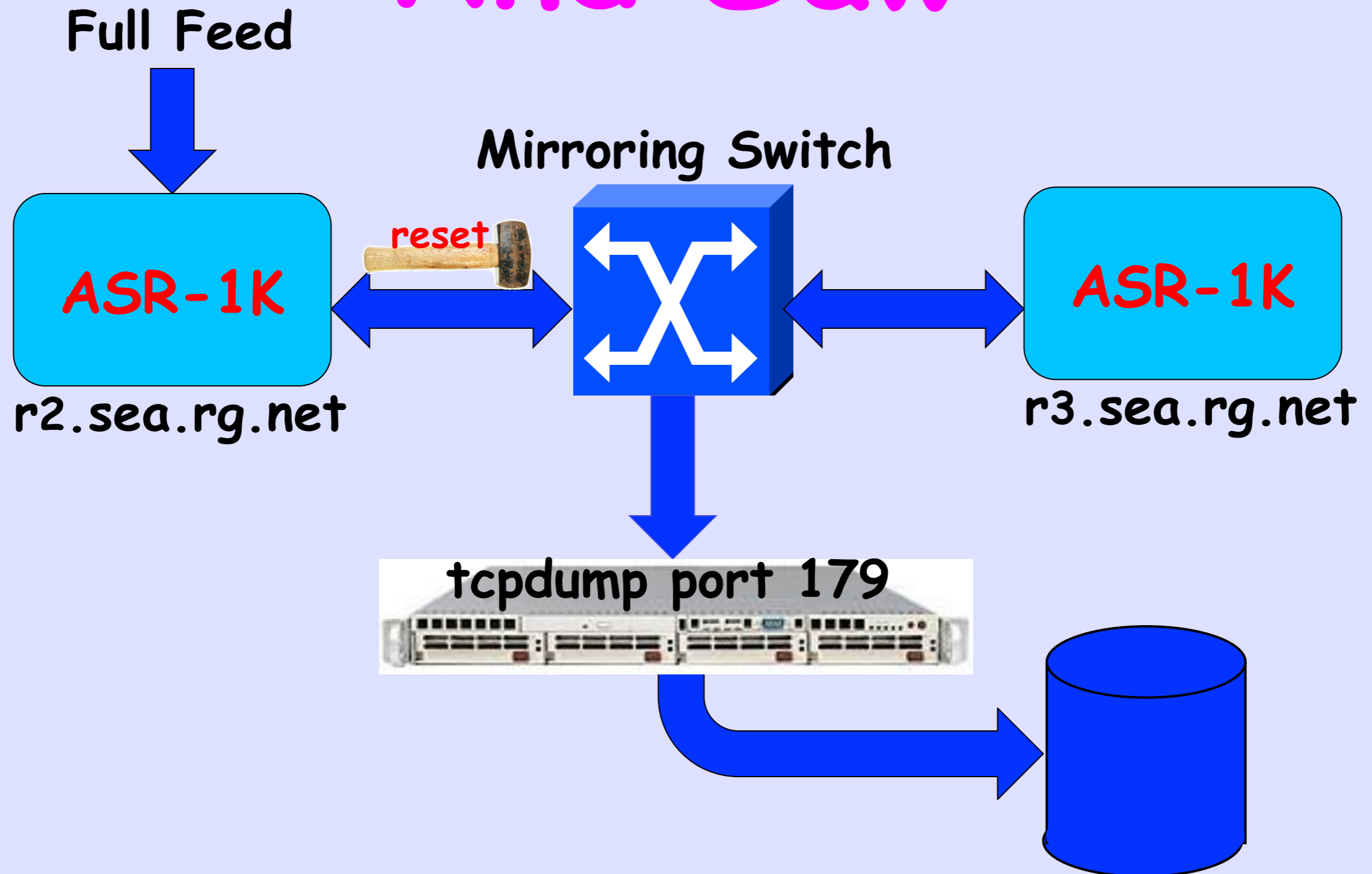
Experimental Set-Up



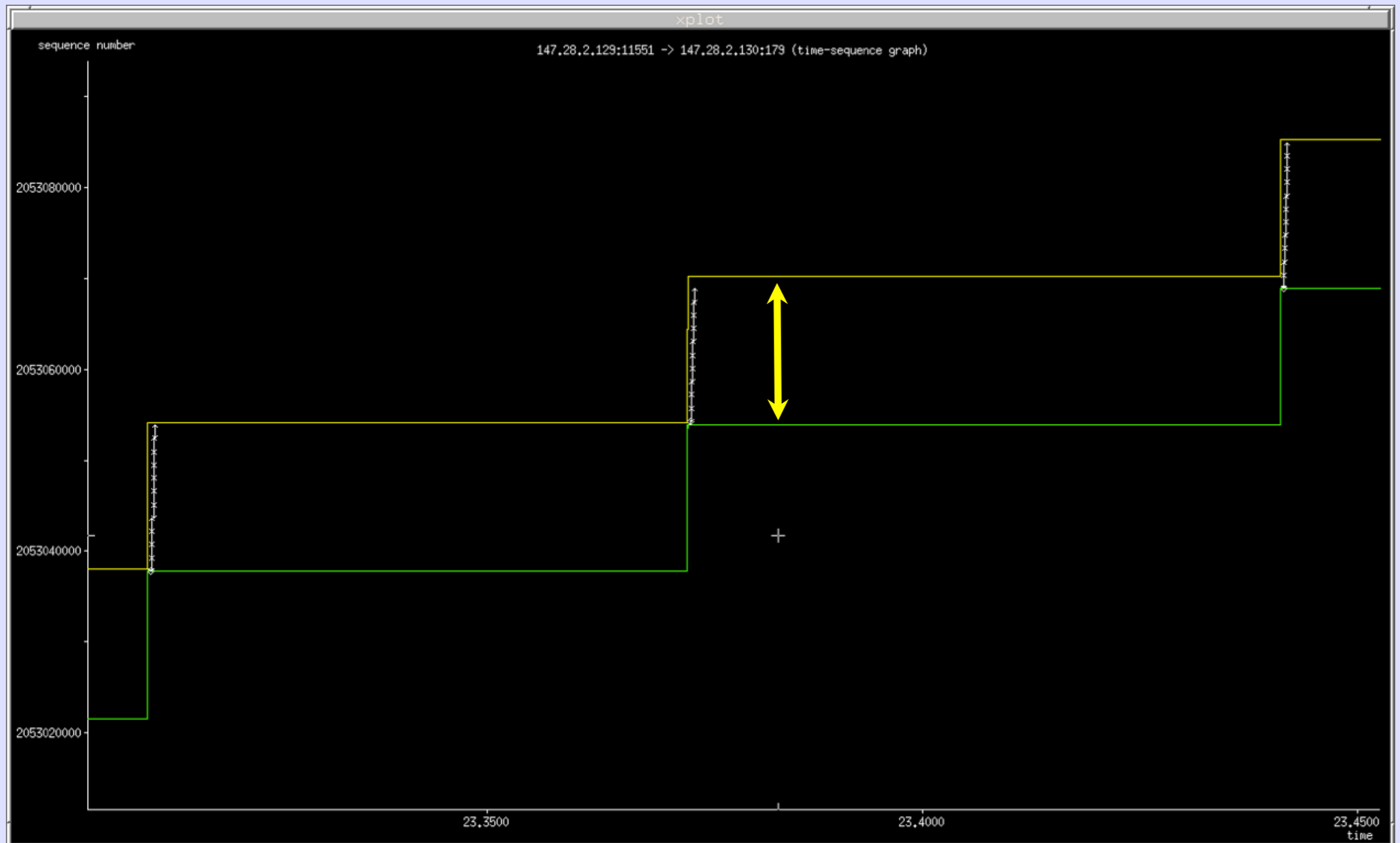
Time Sequence Plots



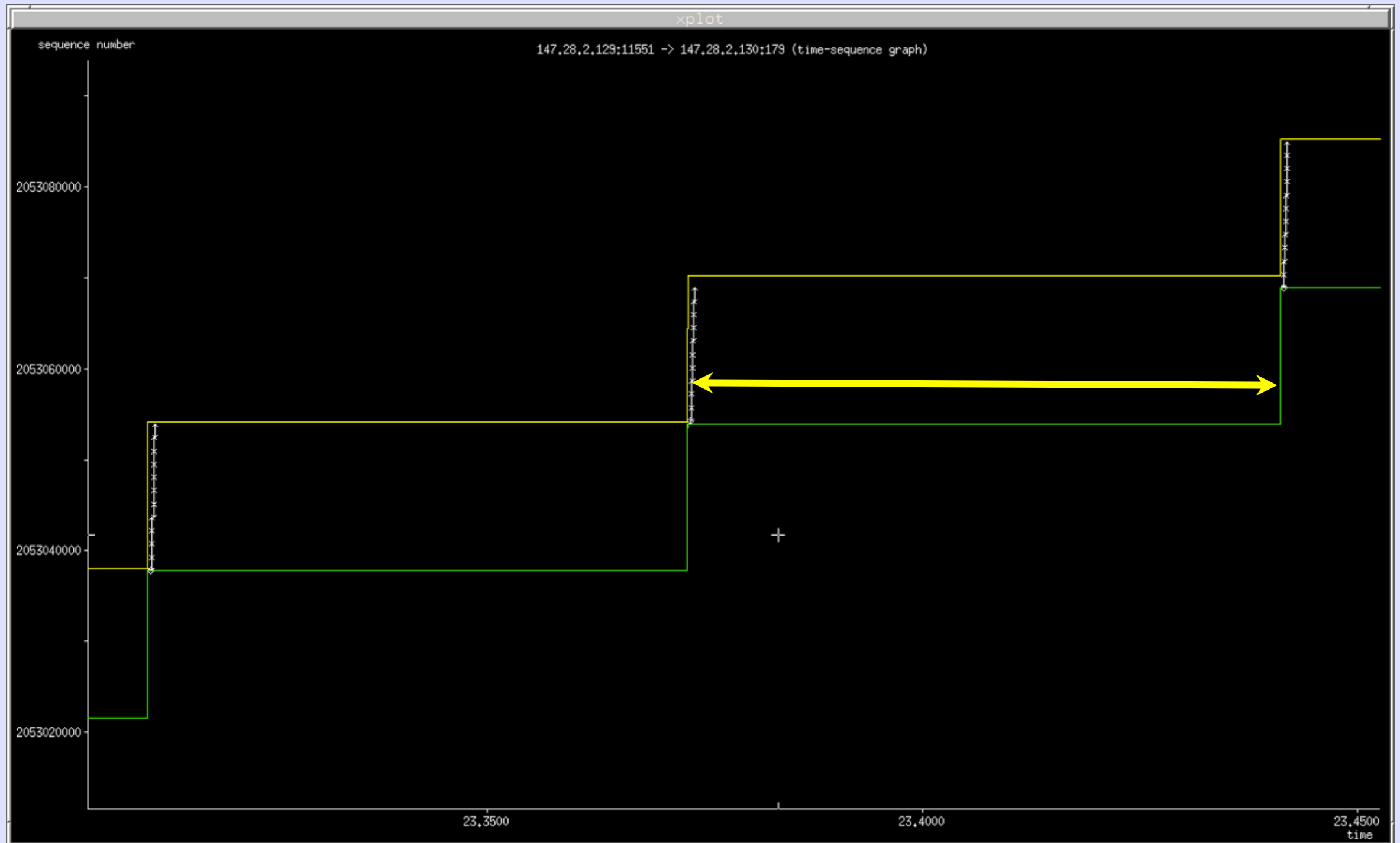
So We Measured And Saw



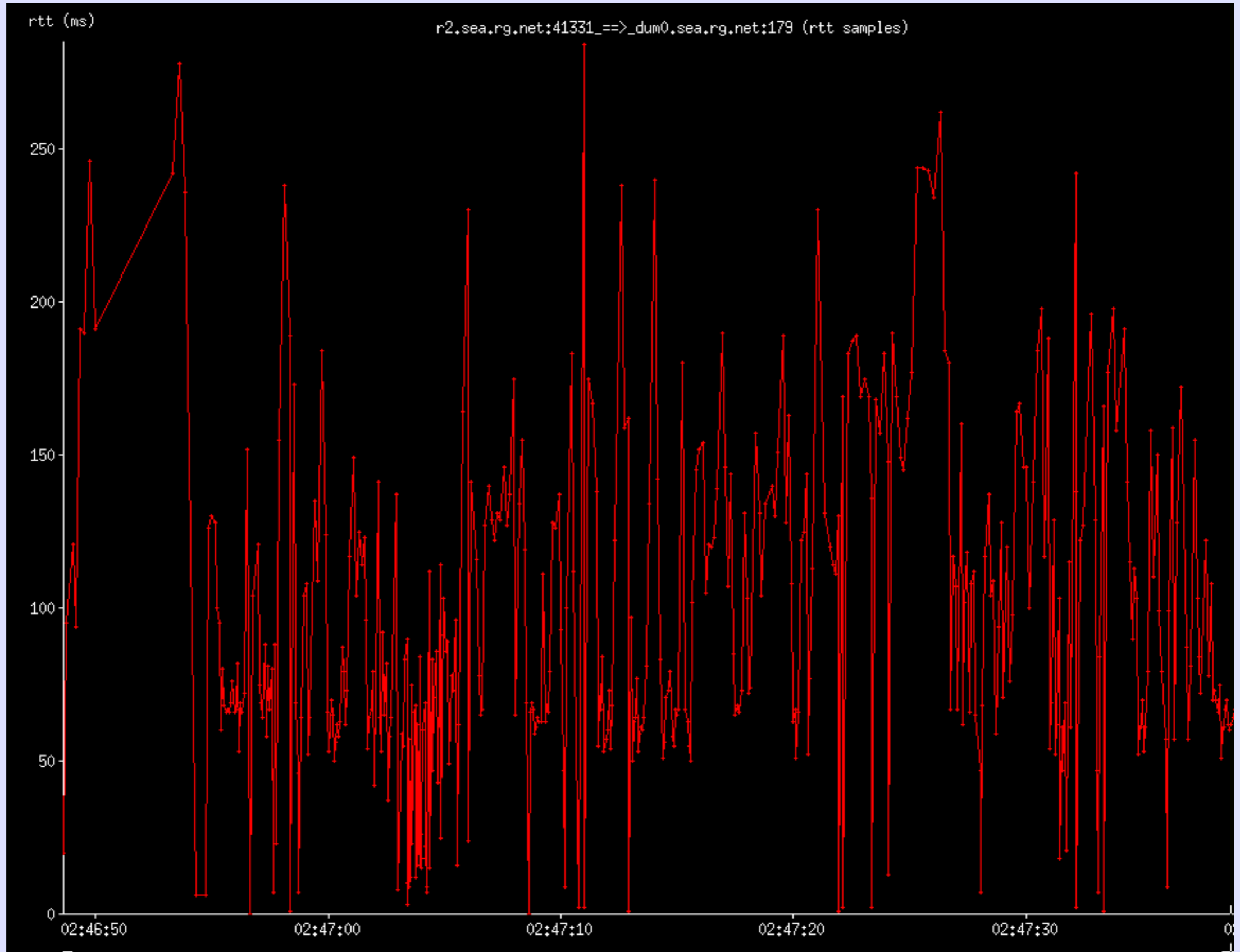
cwnd limited by adv.win



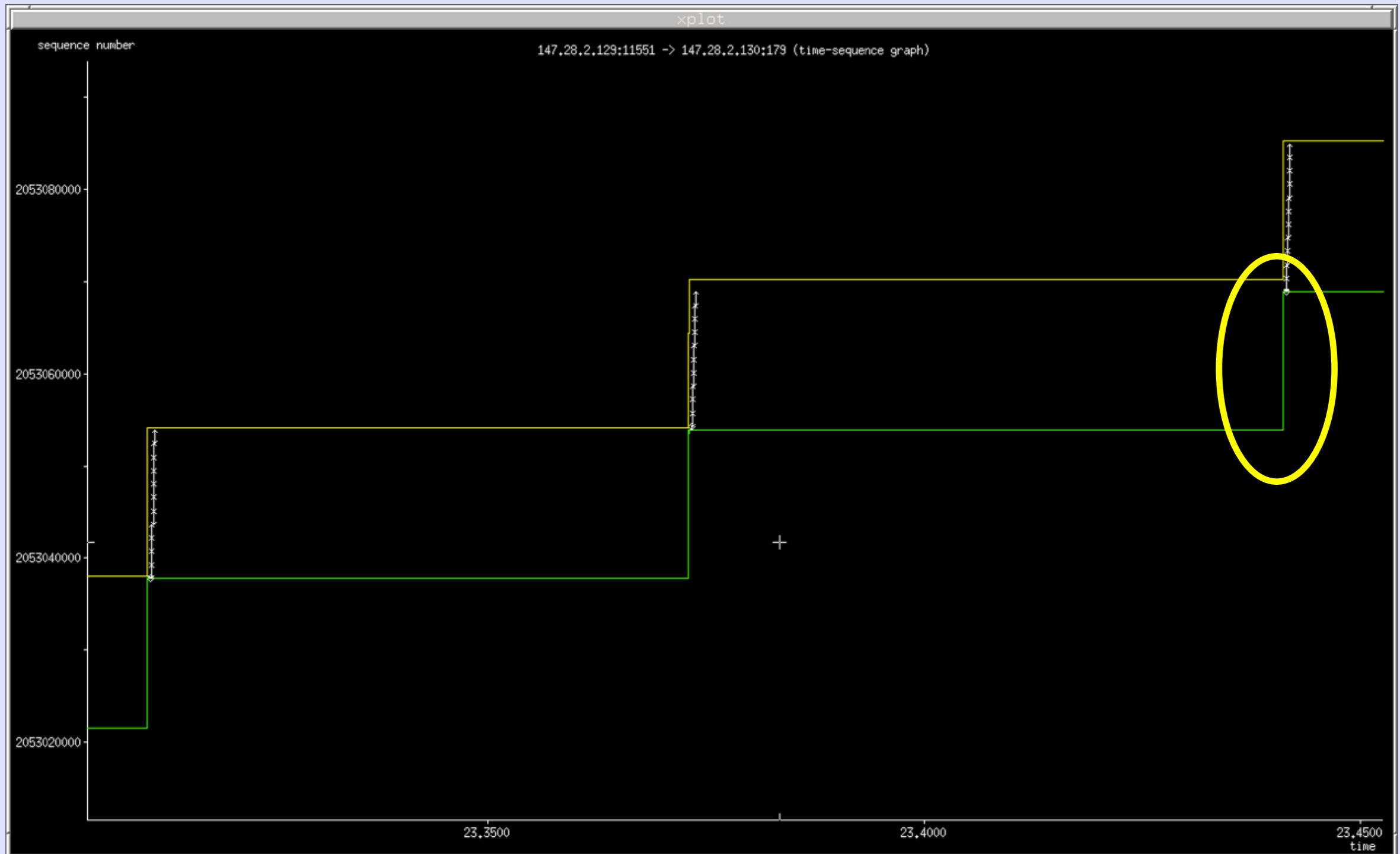
RTT ~ 125ms - On a LAN!



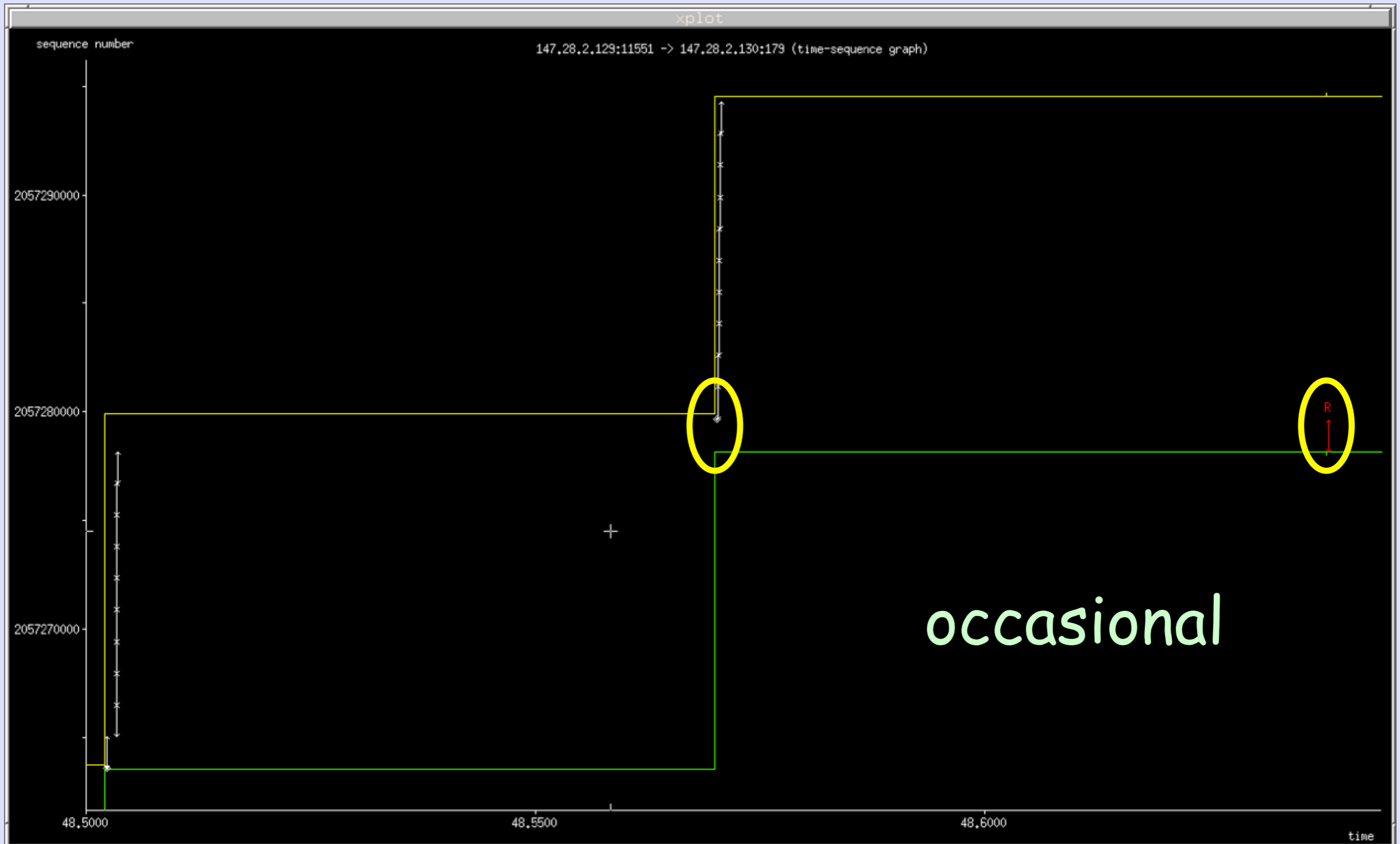
RTT over Full Session



One ACK for Entire Window (AKA 'Stretch' ACK)



Loss and Retransmit



Stretch ACK Not Good

- Stretch ACK for the entire window
 - May contribute to long RTTs, as we wait to coalesce ACKs
 - Very Bursty, as big ACKs cause large window shifts
 - Loss, as bursts overwhelm a buffer (maybe NIC?)

Small Advertised Window

- Would like at least $RTT * \text{Bandwidth}$ for TCP to fully utilize the path capacity
- Window size issue is exacerbated by artificially long RTTs
- We increased the advertised window size but it had no impact, as the stretch ACKs were hindering

Removed
Stretch ACK
from Code

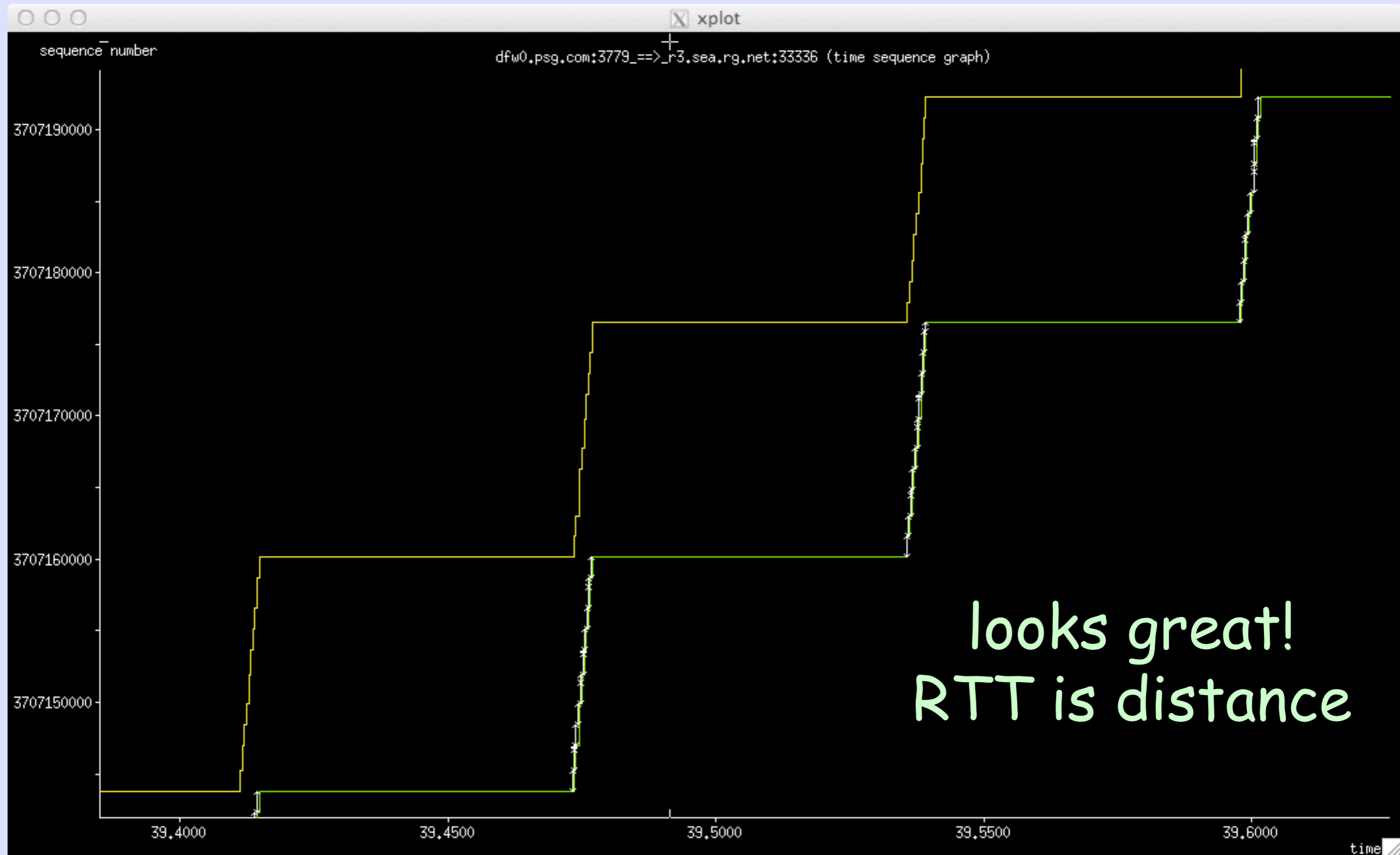
Stretch ACK Removed



What if it is the
TCP Stack?

So Let's Measure a
Non-BGP Protocol

RPKI-Rtr (SEA-DFW)



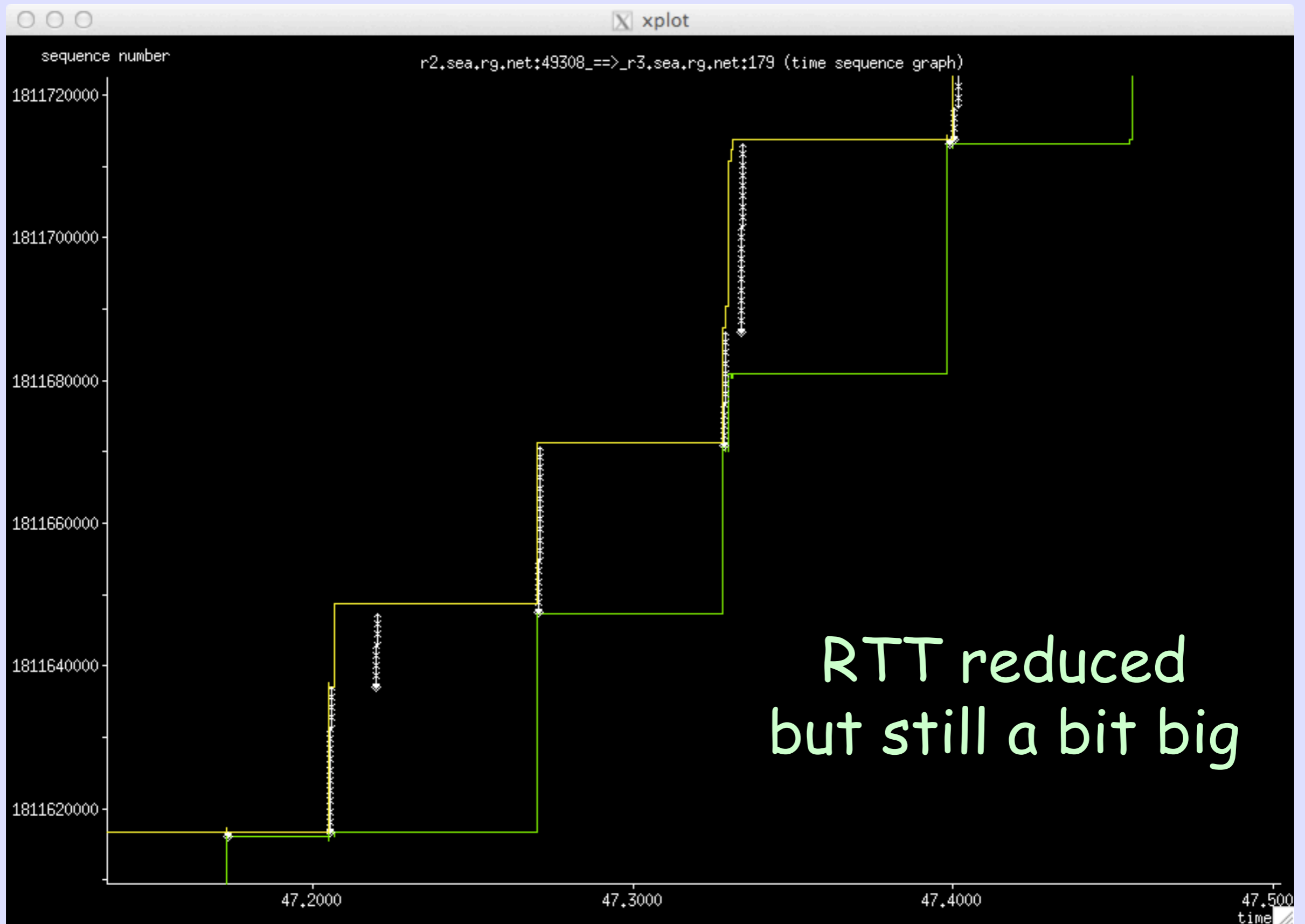
Stack Looks Good!

- Looks like what one would really expect TCP to look like
- ACKs are generated correctly
- Sender fills the window
- RTT looks to be roughly right for the underlying path, Dallas to Seattle
- Window is too small for the path, but buffer small as they're saving RAM

Stack is OK

So is it BGP
or Could it Be
RIB - > FIB?

BGP w/o RIB -> FIB



Better, Not Yet Beautiful

- Get rid of Stretch ACK
- Open the Window to $\geq 32K$
- RIB→FIB is a known 'opportunity'
- Is Run To Completion keeping the RTT high?
- The stack is not so bad. Yay!
- We really want to measure XR!!!

We got the RTTs Down

They are Still Too Long

We are Still

Chasing This

And
We Saw Something
Very Strange
in Dallas

RPKI (DFW-DFW) but XR

