Do the Time Warp – the Rocky Horror PTP Show

Verification of Network Time Synchronization in the Real World

Johannes Zink – j.zink@pengutronix.de

$whoami

- Johannes Zink, a.k.a. dickenhobelix
- Former Systems Engineer, mostly worked on Realtime Networking Systems for ~10 years
- Senior Kernel Developer at Pengutronix since 2022

- Embedded Linux consulting & support since 2001
- ~ 8700 patches in Linux kernel

j.zink@pengutronix.de

Disclaimer

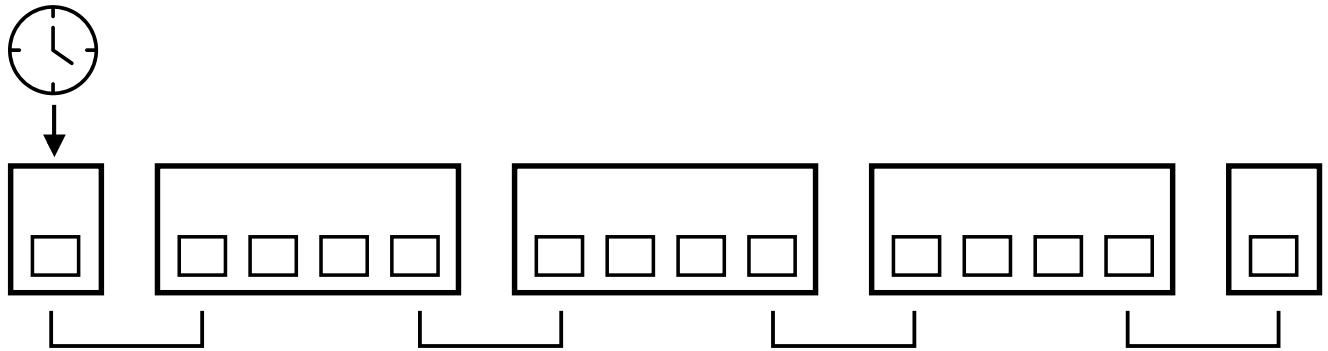
- This talk contains references to the Rocky Horror Show and therfore some profanity. Viewers descretion is advised.
- Rocky Horror is about Interaction, so let's get interactive!
- This talk represents *my* personal experience
- YMMV, especially on other hardware
- This talk and its examples were prepared with linuxptp v3.1, in the meantime v4 has been released, stability has improved a lot

Agenda

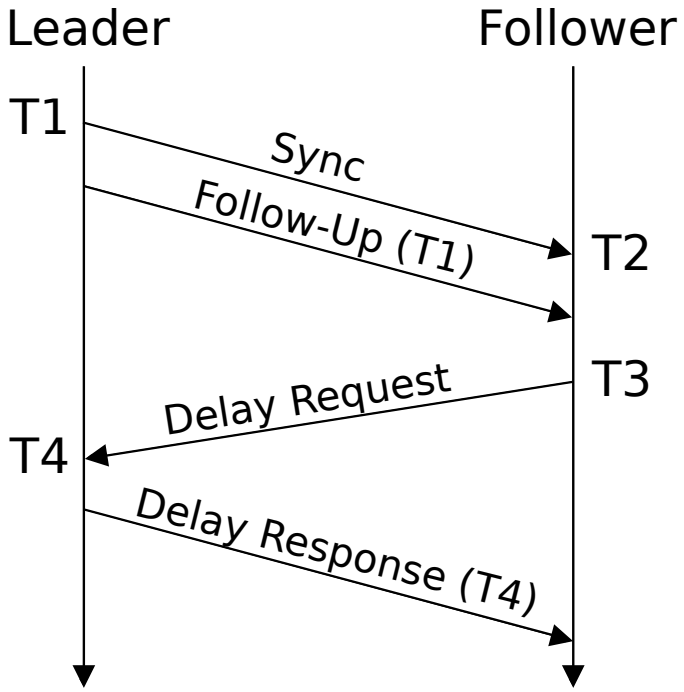
- A brief Introduction to the Precision Time Protocol
- Linux Kernel Components
- Userspace Components
- Measurement Methods
- Examples
- Common Pitfalls
- Best Practices
- Recap and Q&A

Science Fiction – Double Feature

- Synchronize Multiple Clocks over a Network
- Automagically select the best reference clock possible
- Compensate for Network Path delays
  - Unknown path delay
  - Queuing



A brief Introduction to PTP – 2step sync



Timepoints known to Follower

T2

T1, T2

T1, T2, T3

T1, T2, T3, T4

delay = ((T2 – T1) + (T4 – T3)) / 2

offset = (T2 -T1) - delay

A brief Introduction to PTP – BMCA

- Every Clock Listens, may announce its Capabilities, if it thinks its better than what is already announced in the network
- Decision Tree
  - User configurable Fields (priority1, priority2)
  - Clock Quality (clock class, clock accuracy, clock variance)
  - Tie braker (clock identity, port number)
- TL;DR: Autoselect the best possible clock, unless user configures otherwise

Just a sweet trans...port layer – PTP profiles

- Clock Types in Bridges
  - Boundary Clock: Run Leader/Follower per Port
  - Transparent Clock: Adjust Messages for Queuing Delay
  - Ordinary Clock: Just pass Messages without Correction
- Sync Type (1step vs 2step)
- Transport Layer
- Delay Measurement (P2P vs E2E)
- Protocol Extensions (Redundancy)

PTP in Linux - Kernel

- PTP Hardware Clock
  - See Documentation/driver-api/ptp.rst
- Packet Timestamping Support
  - Software, or – better – Hardware offloaded
  - Software Timestamping: generic in network stack
  - HW Offloading usually done in the MAC/NIC drivers, sometimes in the PHY drivers
  - See Documentation/networking/timestamping.rst

PTP in Linux - Userspace

- linuxptp
  - well-established and maintained codebase
  - supports many different profiles
  - tricky to configure
  - ~~only seldom releases new versions - use master~~ quarterly relase schedule as of version 4.0
- some other projects exist
  - often only cover smaller subsets of profiles
  - often „industry code quality“

PTP in Linux – Measurement Methods

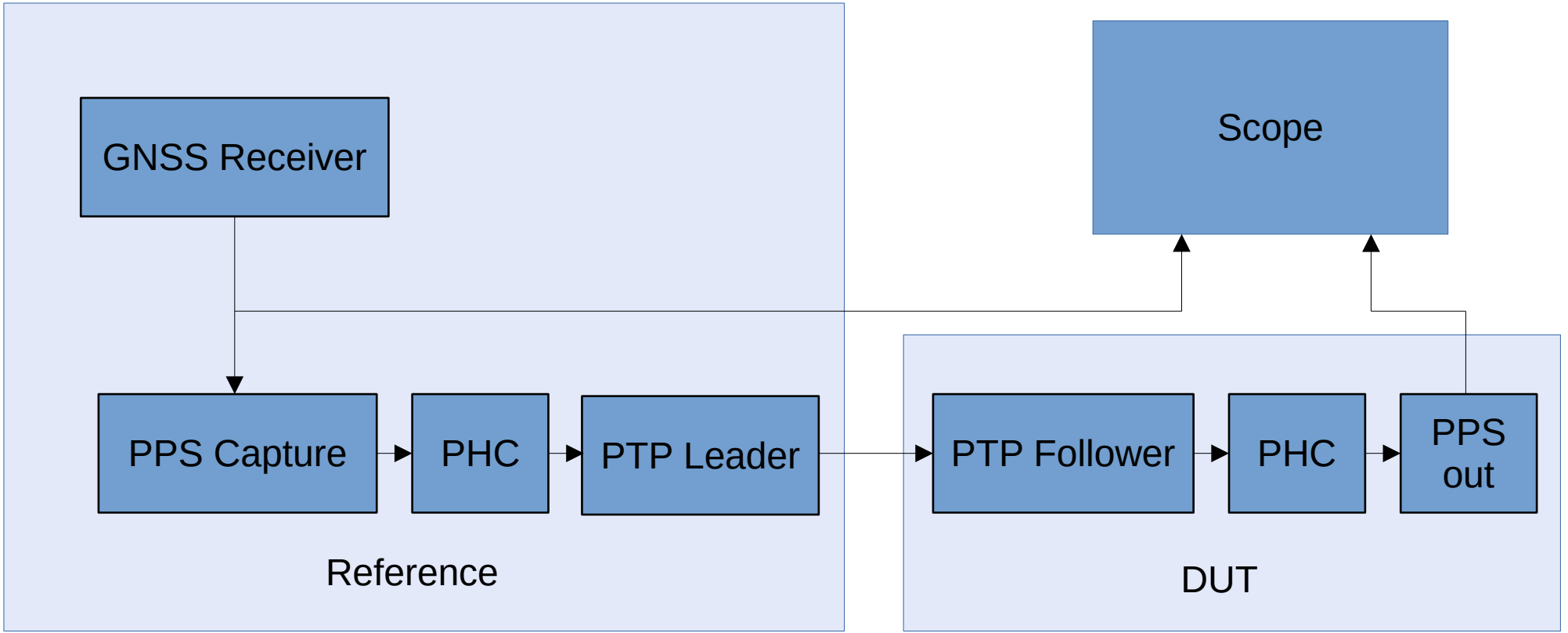
- compare 2 Systems' outputs directly against each other
  - Scope, Time Domain Analyzer
- Reverse-Sync Method
  - Send additional Sync Messages from Follower (DUT) to Leader
- Ingress Measurement Method
  - Follower (DUT) reports incoming sync errors against own reference
- Egress Measurement Method
  - Follower provides timestamps to its messages sent to Leader

Dammit, Janet

- Murphy's law is very strong when setting up PTP
- Settings offer lots of possible permutations
- The list of possible misconfig given in this talk is not exhaustive

- Verify your measurement Setup
- Add plausability checks
- Always check your assumptions!

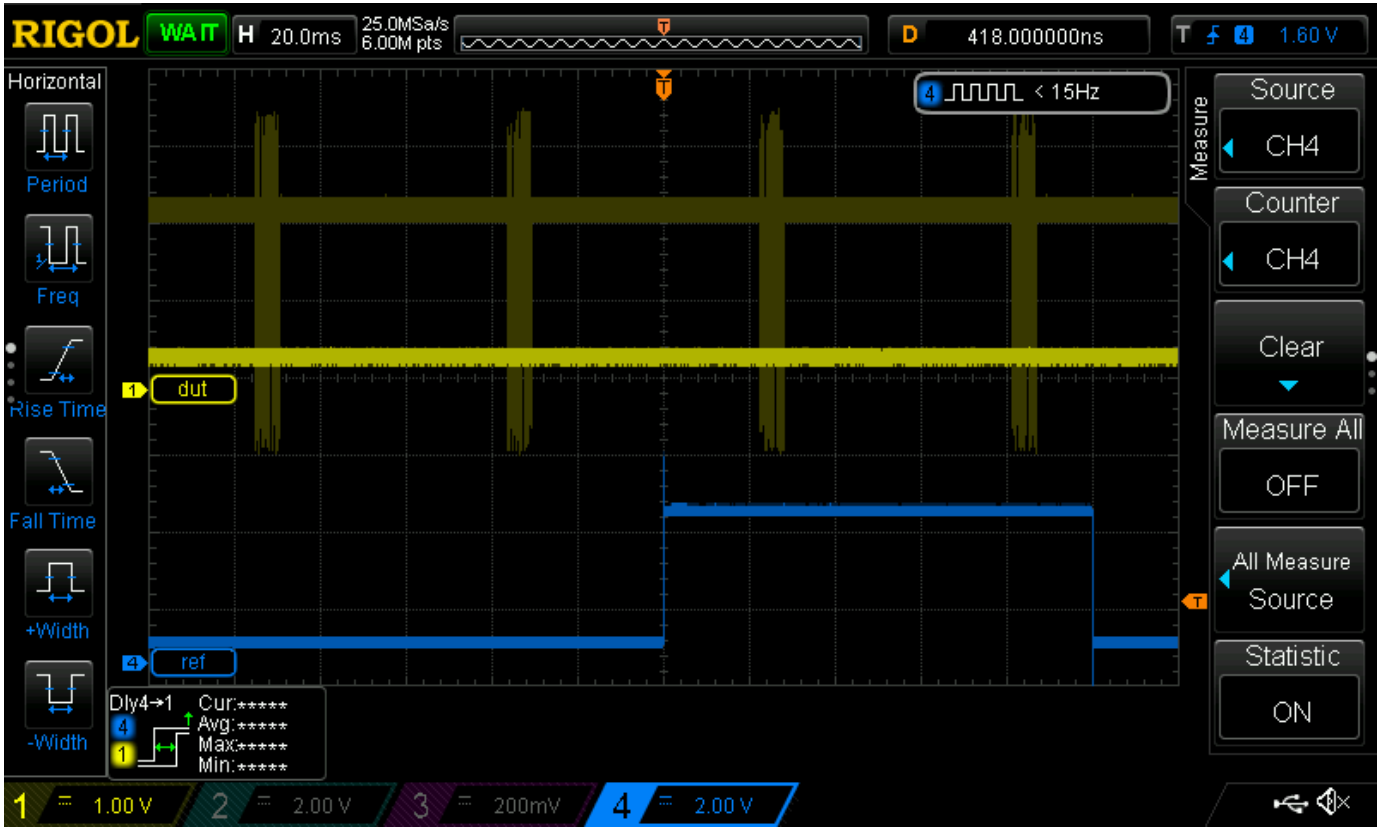
Demo Setup


GNSS Receiver
Scope
PPS Capture
PHC
PTP Leader
PTP Follower
PHC
PPS out
Reference
DUT

Reference – Good Measurement

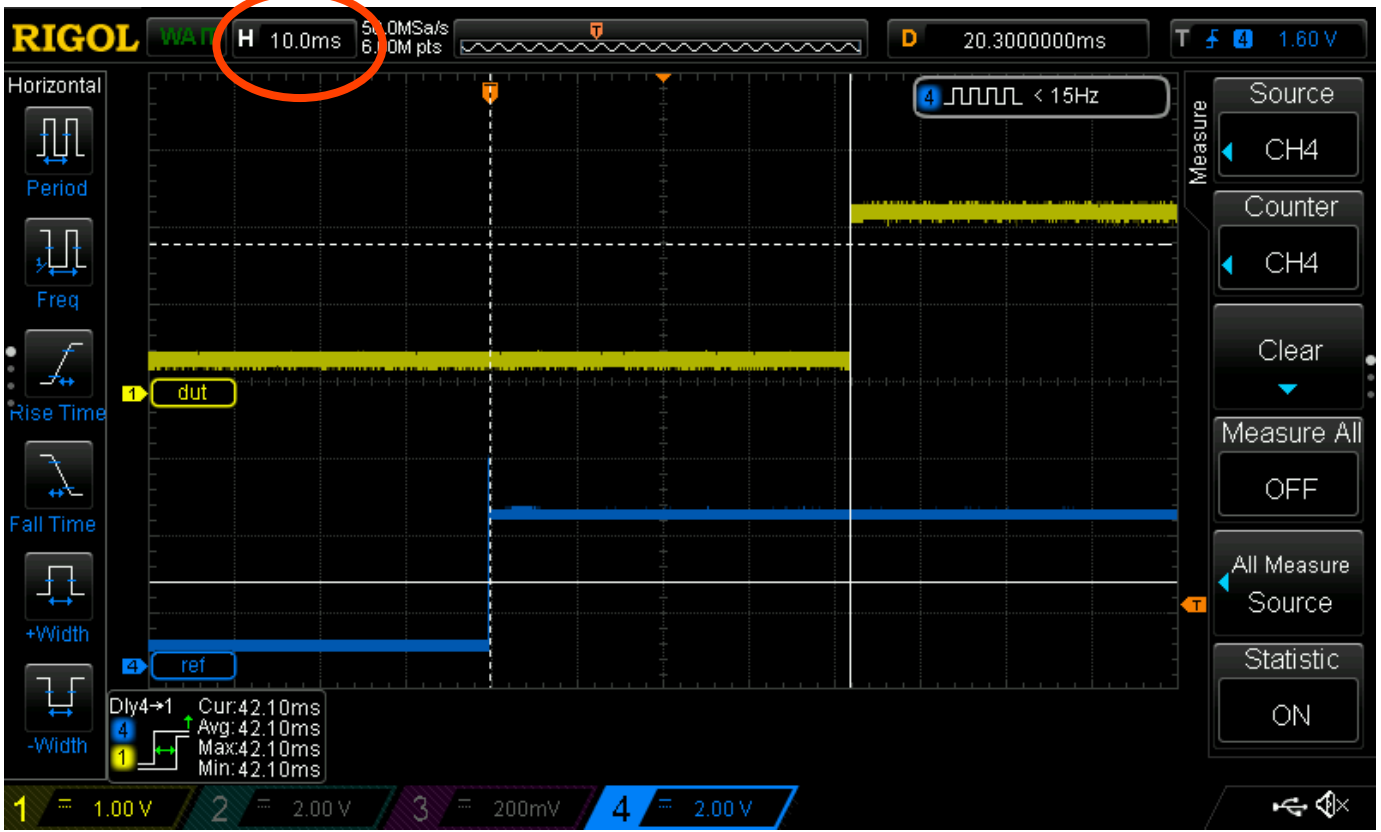
- Synchronized
- Stable over >3h
  - Persistency → ∞
- Excursions normally distributed

Touch-A-Touch-A-Touch-A-Touchy... (1)



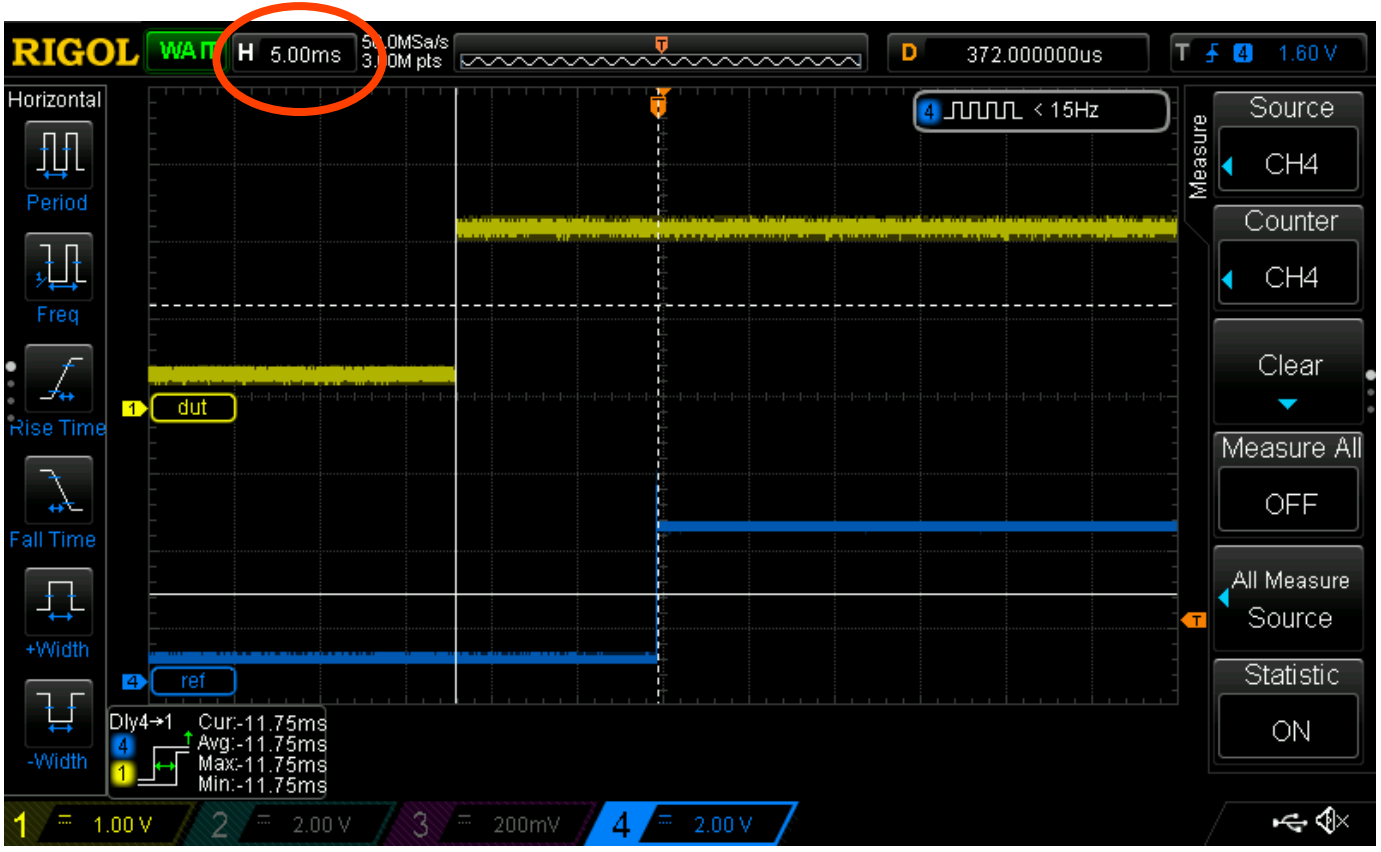
- Unsynchronized
  - Reference Signal Capture failed
  - Link Issue
  - Incompatible Settings
- ...

Touch-A-Touch-A-Touch-A-Touchy… (2)



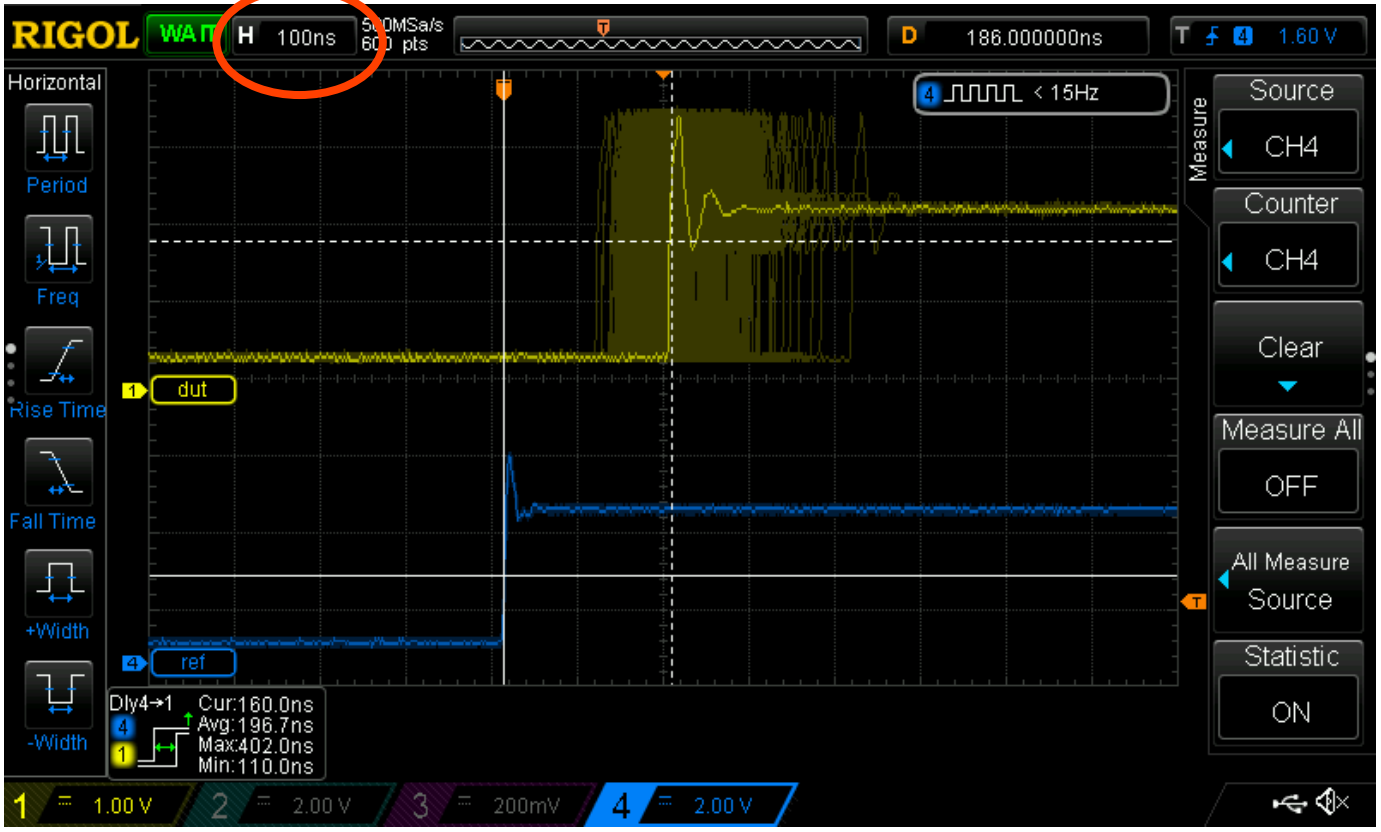
- Large time offset
- Only momentary snapshot
- cannot determine Jitter and Wander
- TL;DR: unknown state

Touch-A-Touch-A-Touch-A-Touchy… (3)



- Trailing edge
- Syntonous, but not synchronous
  - Initial offset at PPS generation
  - Timescale
  - Delay over-compensated
  - broken PHC

Touch-A-Touch-A-Touch-A-Touchy… (4)

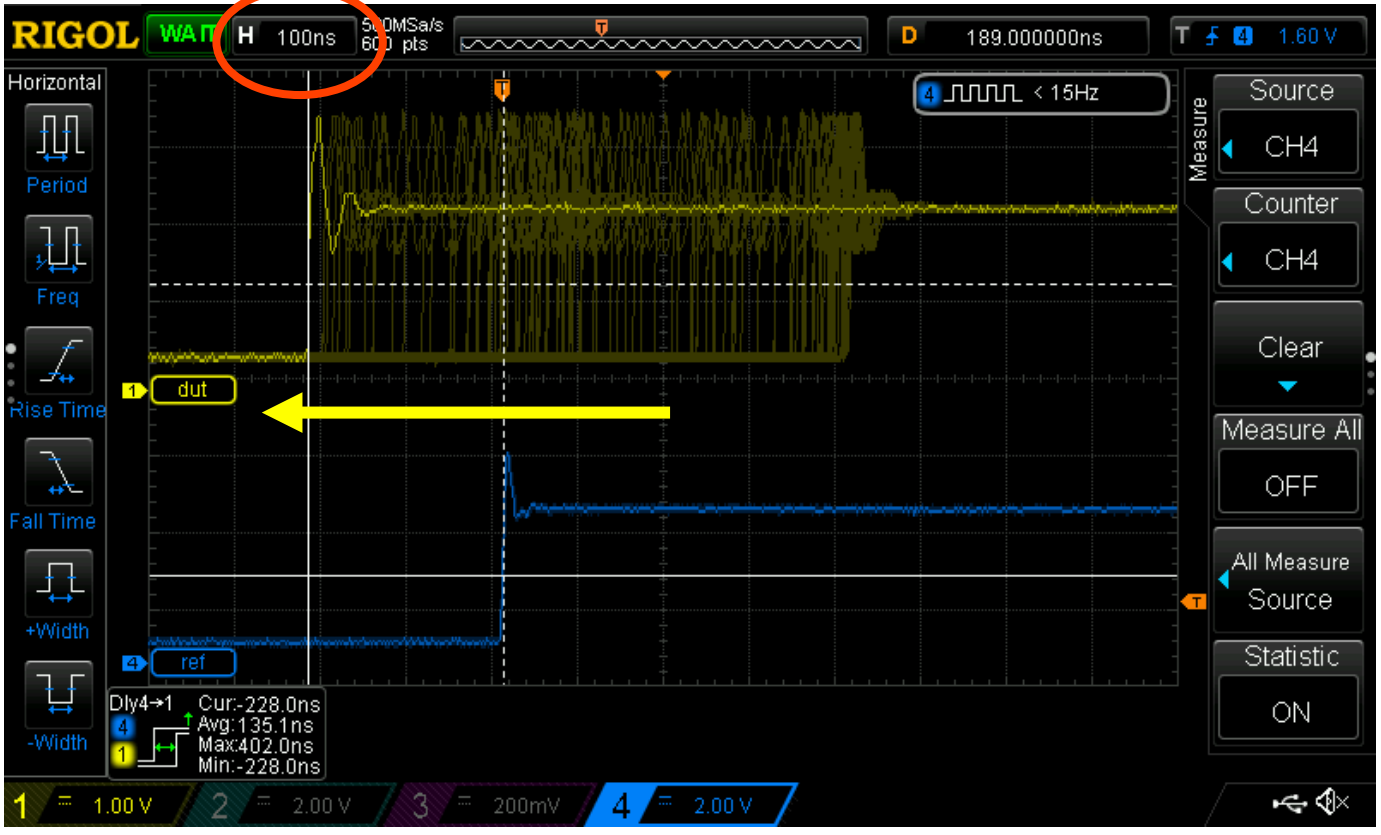


- EEE enabled
  - Influences Timestamping
  - Asymmetric distribution of error

Touch-A-Touch-A-Touch-A-Touchy… (5)

- temporary Sync loss
- Here: leader missed tx ts interrupt

Touch-A-Touch-A-Touch-A-Touchy... (6)



- Permanent Loss of leader
- Thermally stable system
- Too short measuremen (15min)

Reference – Startup

```
ptp4l[36.485]: delay   filtered        378   raw        374
ptp4l[36.819]: port 1: UNCALIBRATED to SLAVE on MASTER_CLOCK_SELECTED
ptp4l[37.445]: rms 17871811494171372 max 35743622988344432 freq -23588 +/- 8915 delay   378 +/-   0
ptp4l[37.485]: port 1: delay timeout
ptp4l[37.485]: delay   filtered        377   raw        374
ptp4l[38.446]: rms   68 max   78 freq -26851 +/-  27 delay   377 +/-   0
ptp4l[38.485]: port 1: delay timeout
ptp4l[38.485]: negative delay     -87918
ptp4l[38.485]: delay = (t2 - t3) * rr + (t4 - t1)
ptp4l[38.485]: t2 - t3 =         +0
ptp4l[38.485]: t4 - t1 =    -175836
ptp4l[38.485]: rr = 0.000000056
ptp4l[38.485]: delay   filtered        375   raw     -87918
ptp4l[39.447]: rms   58 max   79 freq -26797 +/-  11 delay   375 +/-   0
ptp4l[39.485]: port 1: delay timeout
ptp4l[39.485]: negative delay     -79398
ptp4l[39.485]: delay = (t2 - t3) * rr + (t4 - t1)
ptp4l[39.485]: t2 - t3 =         +0
ptp4l[39.486]: t4 - t1 =    -158796
ptp4l[39.486]: rr = 0.000000056
ptp4l[39.486]: delay   filtered        374   raw     -79398
ptp4l[40.448]: rms   22 max   42 freq -26819 +/-  21 delay   374 +/-   0
ptp4l[40.485]: port 1: delay timeout
ptp4l[40.486]: delay   filtered        374   raw        370
ptp4l[41.312]: clearing fault on port 1
ptp4l[41.449]: rms   22 max   28 freq -26874 +/-  12 delay   374 +/-   0
...
```

note: Output looks slightly different for L3 Transport

Reference – stable running Follower

```
...
ptp4l[4567.904]: port 1: delay timeout
ptp4l[4567.904]: delay   filtered        375   raw        375
ptp4l[4568.278]: rms    7 max   10 freq -27554 +/-   6 delay   375 +/-   0
ptp4l[4568.904]: port 1: delay timeout
ptp4l[4568.904]: delay   filtered        374   raw        372
ptp4l[4569.279]: rms   10 max   20 freq -27557 +/-  13 delay   374 +/-   0
ptp4l[4569.904]: port 1: delay timeout
ptp4l[4569.904]: delay   filtered        375   raw        382
ptp4l[4570.280]: rms   10 max   18 freq -27543 +/-  12 delay   375 +/-   0
ptp4l[4570.904]: port 1: delay timeout
ptp4l[4570.905]: delay   filtered        375   raw        381
ptp4l[4571.281]: rms   24 max   34 freq -27504 +/-  14 delay   375 +/-   0
ptp4l[4571.904]: port 1: delay timeout
ptp4l[4571.905]: delay   filtered        375   raw        375
ptp4l[4572.282]: rms   12 max   27 freq -27510 +/-  14 delay   375 +/-   0
ptp4l[4572.904]: port 1: delay timeout
ptp4l[4572.905]: delay   filtered        375   raw        370
ptp4l[4573.283]: rms   26 max   38 freq -27561 +/-  20 delay   375 +/-   0
ptp4l[4573.904]: port 1: delay timeout
ptp4l[4573.905]: delay   filtered        375   raw        378
ptp4l[4574.284]: rms   22 max   28 freq -27582 +/-  10 delay   375 +/-   0
ptp4l[4574.905]: port 1: delay timeout
ptp4l[4574.905]: delay   filtered        375   raw        379
ptp4l[4575.285]: rms   17 max   30 freq -27549 +/-  22 delay   375 +/-   0
ptp4l[4575.905]: port 1: delay timeout
ptp4l[4575.905]: delay   filtered        375   raw        375
...
```

- Output looks slightly different for L3 Transport

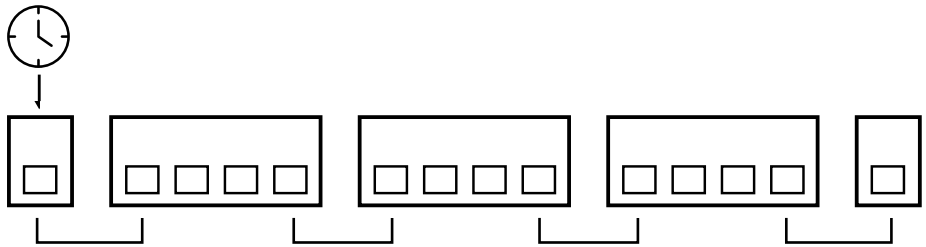
Touch-A-Touch-A-Touch-A-Touchy... (7)

```
ptp4l[5389.613]: port 1: delay timeout
ptp4l[5389.614]: delay   filtered        374   raw        375
ptp4l[5389.781]: rms    6 max    8 freq -27548 +/-   9 delay   374 +/-   0
ptp4l[5390.613]: port 1: delay timeout
ptp4l[5390.614]: delay   filtered        375   raw        385
ptp4l[5390.782]: rms   11 max   20 freq -27538 +/-  13 delay   375 +/-   0
ptp4l[5391.613]: port 1: delay timeout
ptp4l[5391.614]: delay   filtered        375   raw        369
ptp4l[5391.783]: rms   20 max   28 freq -27577 +/-  15 delay   375 +/-   0
ptp4l[5392.534]: port 1: rx sync timeout
ptp4l[5392.534]: port 1: SLAVE to MASTER on ANNOUNCE_RECEIPT_TIMEOUT_EXPIRES
ptp4l[5392.534]: selected local clock 00049f.fffe.079c42 as best master
ptp4l[5392.534]: port 1: assuming the grand master role
ptp4l[5392.535]: port 1: master tx announce timeout
ptp4l[5392.614]: port 1: delay timeout
ptp4l[5392.659]: port 1: master sync timeout
ptp4l[5392.784]: port 1: master sync timeout
ptp4l[5392.909]: port 1: master sync timeout
ptp4l[5393.034]: port 1: master sync timeout
ptp4l[5393.159]: port 1: master sync timeout
ptp4l[5393.284]: port 1: master sync timeout
ptp4l[5393.409]: port 1: master sync timeout
ptp4l[5393.535]: port 1: master sync timeout
ptp4l[5393.535]: port 1: master tx announce timeout
ptp4l[5393.614]: port 1: delay timeout
ptp4l[5393.660]: port 1: master sync timeout
...
```

- Follower looses Sync from Leader

Touch-A-Touch-A-Touch-A-Touchy... (8)

```
...
ptp4l[2216.058]: delay   filtered      16065   raw      16060
ptp4l[2216.106]: port 1: master sync timeout
ptp4l[2216.231]: port 1: master sync timeout
ptp4l[2216.356]: port 1: master sync timeout
ptp4l[2216.474]: port 1: master tx announce timeout
ptp4l[2216.481]: port 1: master sync timeout
ptp4l[2216.606]: port 1: master sync timeout
ptp4l[2216.731]: port 1: master sync timeout
ptp4l[2216.856]: port 1: master sync timeout
ptp4l[2216.981]: port 1: master sync timeout
ptp4l[2217.058]: port 1: delay timeout
ptp4l[2217.058]: delay   filtered      16065   raw      16080
ptp4l[2217.106]: port 1: master sync timeout
ptp4l[2217.231]: port 1: master sync timeout
ptp4l[2217.356]: port 1: master sync timeout
ptp4l[2217.474]: port 1: master tx announce timeout
ptp4l[2217.481]: port 1: master sync timeout
ptp4l[2217.606]: port 1: master sync timeout
ptp4l[2217.732]: port 1: master sync timeout
ptp4l[2217.857]: port 1: master sync timeout
ptp4l[2217.982]: port 1: master sync timeout
...
```



- One of the intermediate links falls back to half-duplex

Touch-A-Touch-A-Touch-A-Touchy... (9)

```
...
ptp4l[2840.920]: port 0: INITIALIZING to LISTENING on INIT_COMPLETE
ptp4l[2840.920]: port 1: received link status notification
ptp4l[2840.920]: interface index 3 is down
ptp4l[2840.920]: port 1: link down
ptp4l[2840.920]: port 1: LISTENING to FAULTY on FAULT_DETECTED (FT_UNSPECIFIED)
ptp4l[2840.952]: waiting 2^{4} seconds to clear fault on port 1
ptp4l[2840.952]: selected local clock 00049f.fffe.079c42 as best master
ptp4l[2840.952]: port 1: assuming the grand master role
ptp4l[2856.952]: clearing fault on port 1
...
```

- Start on half-duplex link

Touch-A-Touch-A-Touch-A-Touchy... (10)

```
...
ptp4l[3742.603]: delay   filtered      17175   raw      17175
ptp4l[3742.637]: port 1: master sync timeout
ptp4l[3742.688]: port 1: master tx announce timeout
ptp4l[3742.762]: port 1: master sync timeout
ptp4l[3742.887]: port 1: master sync timeout
ptp4l[3743.012]: port 1: master sync timeout
ptp4l[3743.138]: port 1: master sync timeout
ptp4l[3743.263]: port 1: master sync timeout
ptp4l[3743.388]: port 1: master sync timeout
ptp4l[3743.513]: port 1: master sync timeout
ptp4l[3743.602]: port 1: delay timeout
ptp4l[3743.603]: delay   filtered      17175   raw      17175
ptp4l[3743.638]: port 1: master sync timeout
ptp4l[3743.688]: port 1: master tx announce timeout
ptp4l[3743.763]: port 1: master sync timeout
ptp4l[3743.888]: port 1: master sync timeout
ptp4l[3744.013]: port 1: master sync timeout
ptp4l[3744.138]: port 1: master sync timeout
ptp4l[3744.263]: port 1: master sync timeout
ptp4l[3744.388]: port 1: master sync timeout
ptp4l[3744.513]: port 1: master sync timeout
ptp4l[3744.602]: port 1: delay timeout
ptp4l[3744.603]: delay   filtered      17175   raw      17196
...
```

- Hardware Bug, incomplete Driver
- Too Large Peer Delay

  → Loss of Sync Domain Membership

Touch-A-Touch-A-Touch-A-Touchy... (11)

```
...
ptp4l[88348.672]: driver changed our HWTSTAMP options
ptp4l[88348.672]: tx_type 1 not 1
ptp4l[88348.672]: rx_filter 1 not 12
ptp4l[88348.672]: port 1: INITIALIZING to LISTENING on INITIALIZE
ptp4l[88348.672]: port 0: INITIALIZING to LISTENING on INITIALIZE
ptp4l[88352.175]: port 1: new foreign master 0026f2.fffe.f25aa0-1
ptp4l[88354.672]: port 1: LISTENING to MASTER on
ANNOUNCE_RECEIPT_TIMEOUT_EXPIRES
ptp4l[88354.672]: selected best master clock a0369f.fffe.1cdd3b
ptp4l[88354.672]: assuming the grand master role
ptp4l[88356.177]: selected best master clock 0026f2.fffe.f25aa0
ptp4l[88356.177]: assuming the grand master role
ptp4l[95586.834]: poll tx timestamp timeout
ptp4l[95586.834]: port 1: send sync failed
ptp4l[95586.834]: port 1: MASTER to FAULTY on FAULT_DETECTED  (FT_UNSPECIFIED)

...
```

- Leader missing a TX Timestamp
- Causes increasing Fault Backoff

Do the time warp - common Pitfalls

- Multiple Time sources in the system (NTP, ...)
  - system time jump
  - NTP also tunes the PHC if using HW timestamping
- PTP profiles
- Missing, incomplete or defective timestamping and clock support in hardware or drivers
- Timestamping in MAC vs in PHY
- Hardware often only supports subset of profiles (1step/2step, Layer2/Layer3, P2P/E2E, ...)

Do the time warp - common Pitfalls

- Multiple Timescales – Offsets, Leap Seconds, etc
- False positive Debug output
- Daemon stability
- Measurement Method – PPS frequency: resolution vs precision, rollovers
- Sporadic dropouts
  - delayed TX timestamps (mitigate with --tx_timestamp_timeout)
- Check selected leader clock
- Never rely solely on datasheets! Measure yourself!

Best Practices

- Choose the correct profile. Often this is dictated by the application anyway
- Do not rely on example commands in the internet™
- Check Hardware Clock Availablity, Stability and Configuration
- Analyze Logs in Endpoints and in bridges
- Read the fine Manpage – subtle differences in parameters
- Thoroughly test over $aLongerPeriodOfTime™
- Always check your assumptions!

Recap

- PTP can offer great performance when done right
- PTP has lots of parameters to tweak
- Performance and stability depends strongly on optimizing your setup properly
- Some effects only sporadically show up
- There’s a light in everybody’s measurement setup…
- Always check your assumptions!

By the way…

… we are hiring

jobs.pengutronix.de

Thank you for your Attention!

Any questions?

j.zink@pengutronix.de