Network Timing: the Good, the Bad, the Ugly, and how to Make it Stop!

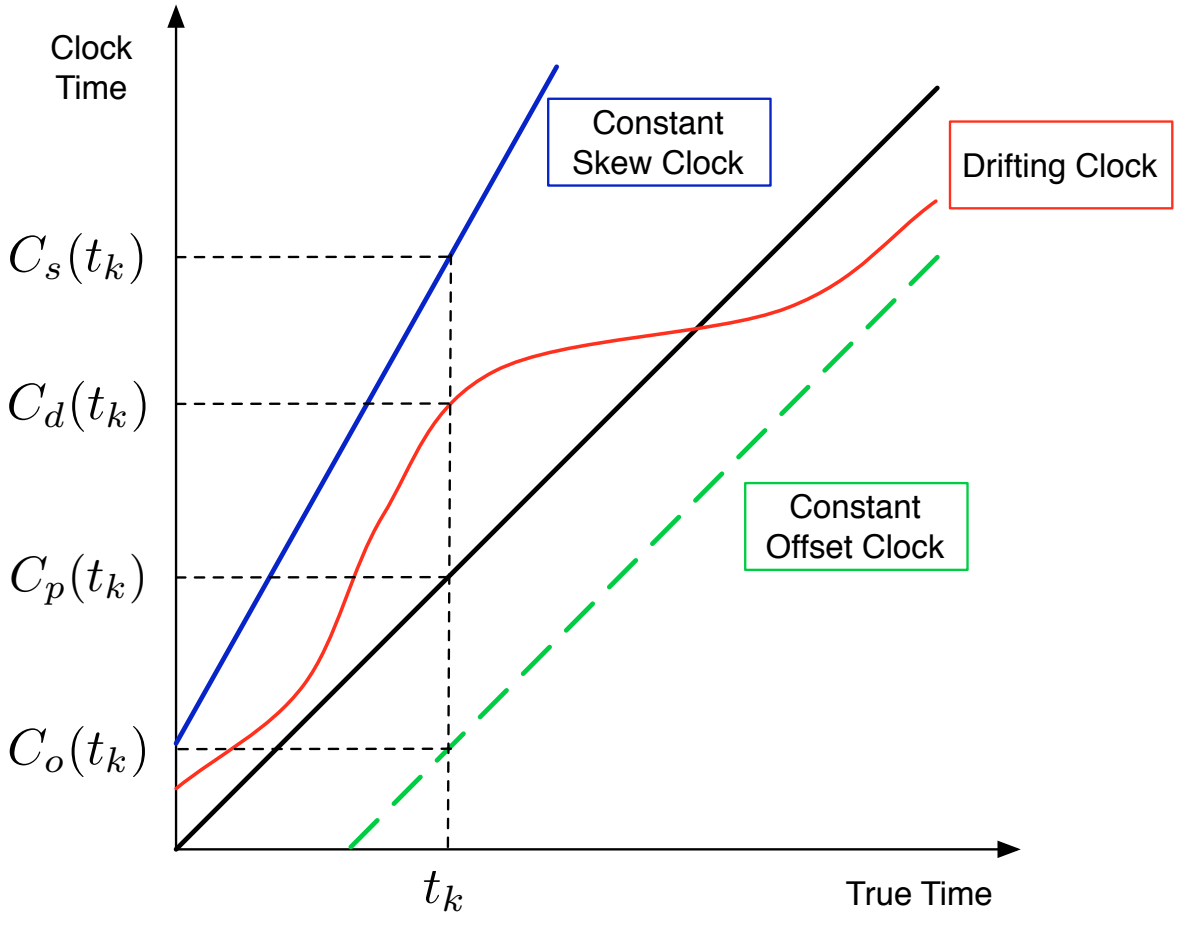
Darryl Veitch

University of Technology Sydney

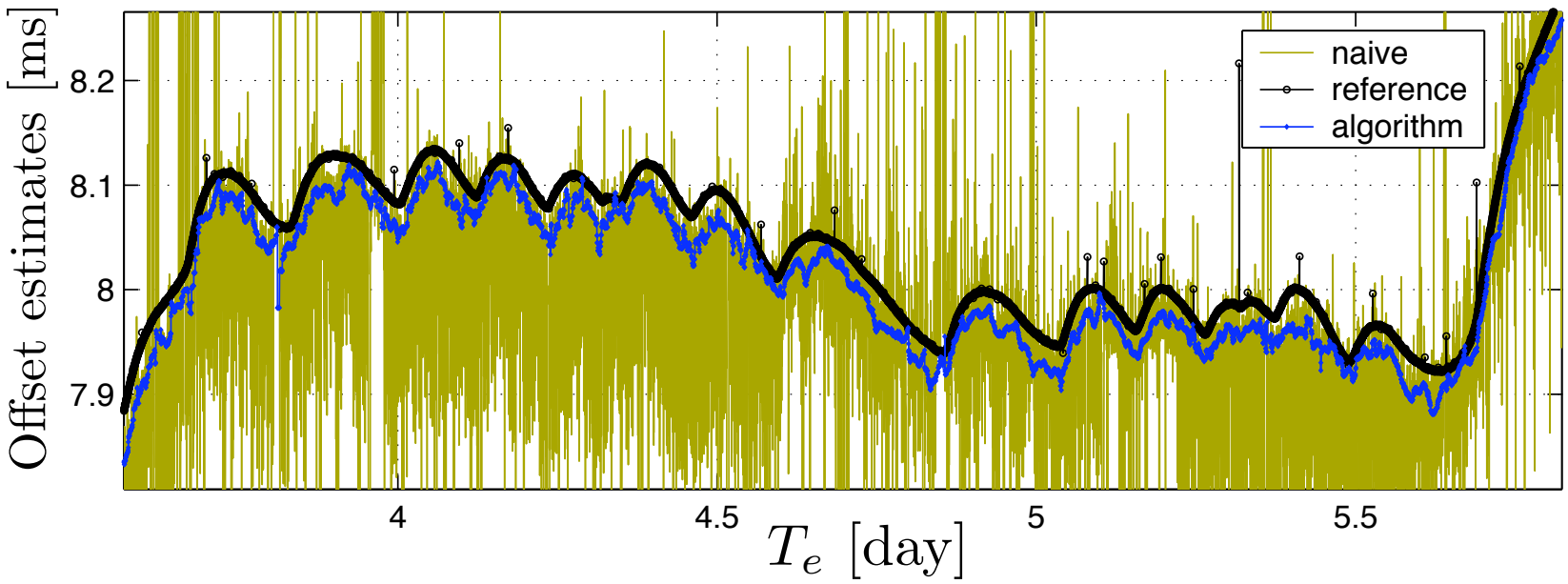
Collaboration with Yi Cao

ACDCN 2018, 28-29 Nov., University of Adelaide

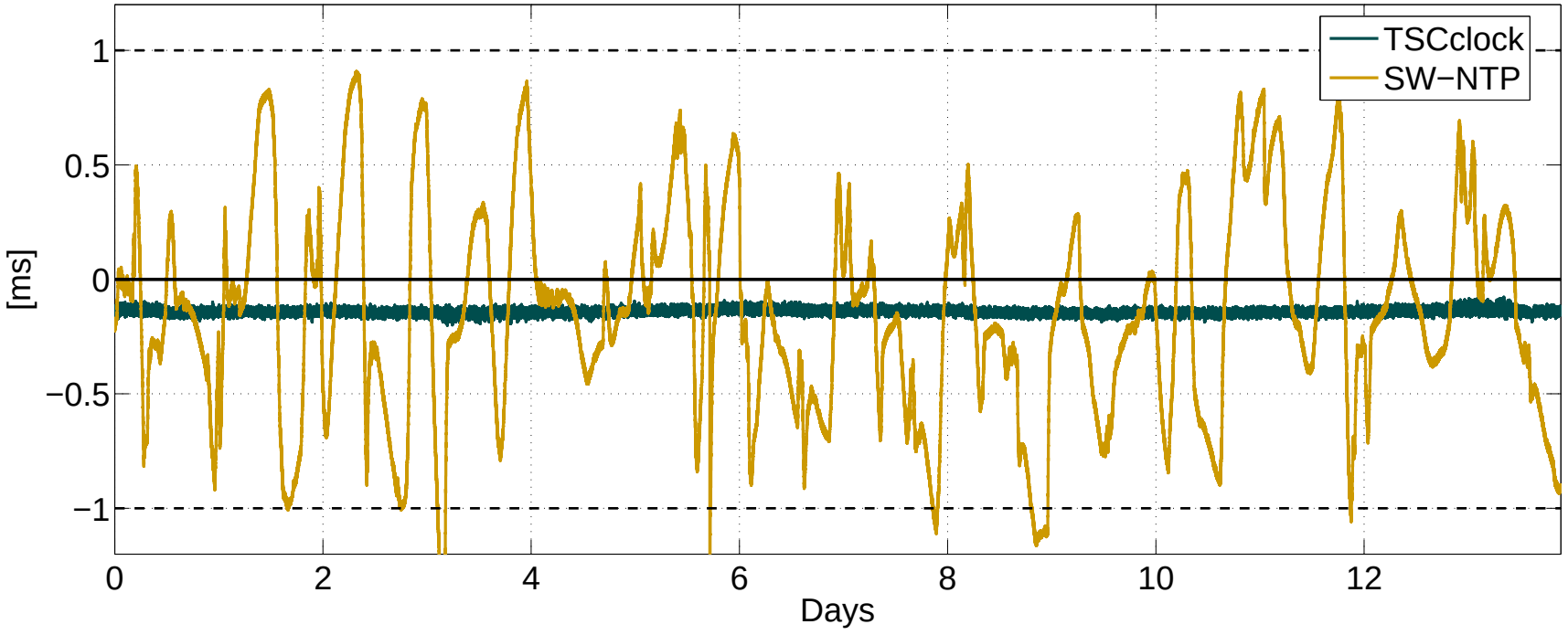
Offset, Skew and Drift



DRIFT ESTIMATION USING RADCLOCK



RADCLOCK IN ONE SLIDE



TIMESCALES

Temps Atomique International (TAI) – PRIMARY STANDARD

- Seconds are constant, each one SI second long
- Continuous time scale
- Origin at HH:MM:SS = 00:00:10, 1st January 1972

TIMESCALES

Temps Atomique International (TAI) – PRIMARY STANDARD

- Seconds are constant, each one SI second long
- Continuous time scale
- Origin at HH:MM:SS = 00:00:10, 1st January 1972

Universal Time (UT1) – DESCENDANT OF GREENWICH MEAN TIME

- Seconds vary to track Earth's variable rotation
- Continuous time scale

TIMESCALES

Temps Atomique International (TAI) – PRIMARY STANDARD

- Seconds are constant, each one SI second long
- Continuous time scale
- Origin at HH:MM:SS = 00:00:10, 1st January 1972

Universal Time (UT1) – DESCENDANT OF GREENWICH MEAN TIME

- Seconds vary to track Earth's variable rotation
- Continuous time scale

Coordinated Universal Time (UTC) – GENERAL STANDARD TIMEKEEPING

- Seconds are constant, except
- Jumps (Leap Seconds) added to keep $|\mathrm{UTC} - \mathrm{UT1}| \leq 0.9$ sec
- Origin at $t_{\mathrm{TAI}} = -10\mathrm{sec}$

NTP Server Stratum 'Hierarchy'

The *Network Time Protocol* (NTP) system distributes UTC.

NTP Server Stratum 'Hierarchy'

The *Network Time Protocol* (NTP) system distributes UTC.

Stratum-0 Reference

- Reference time source
- Typically GPS, or atomic clock

NTP Server Stratum 'Hierarchy'

The *Network Time Protocol* (NTP) system distributes UTC.

Stratum-0 Reference

- Reference time source
- Typically GPS, or atomic clock

Stratum-1 Servers

- Roots of the network timekeeping forest
- Direct access to a Stratum-0
- Could be network appliance or PC running system clock + ntpd

NTP Server Stratum 'Hierarchy'

The *Network Time Protocol* (NTP) system distributes UTC.

Stratum-0 Reference

- Reference time source
- Typically GPS, or atomic clock

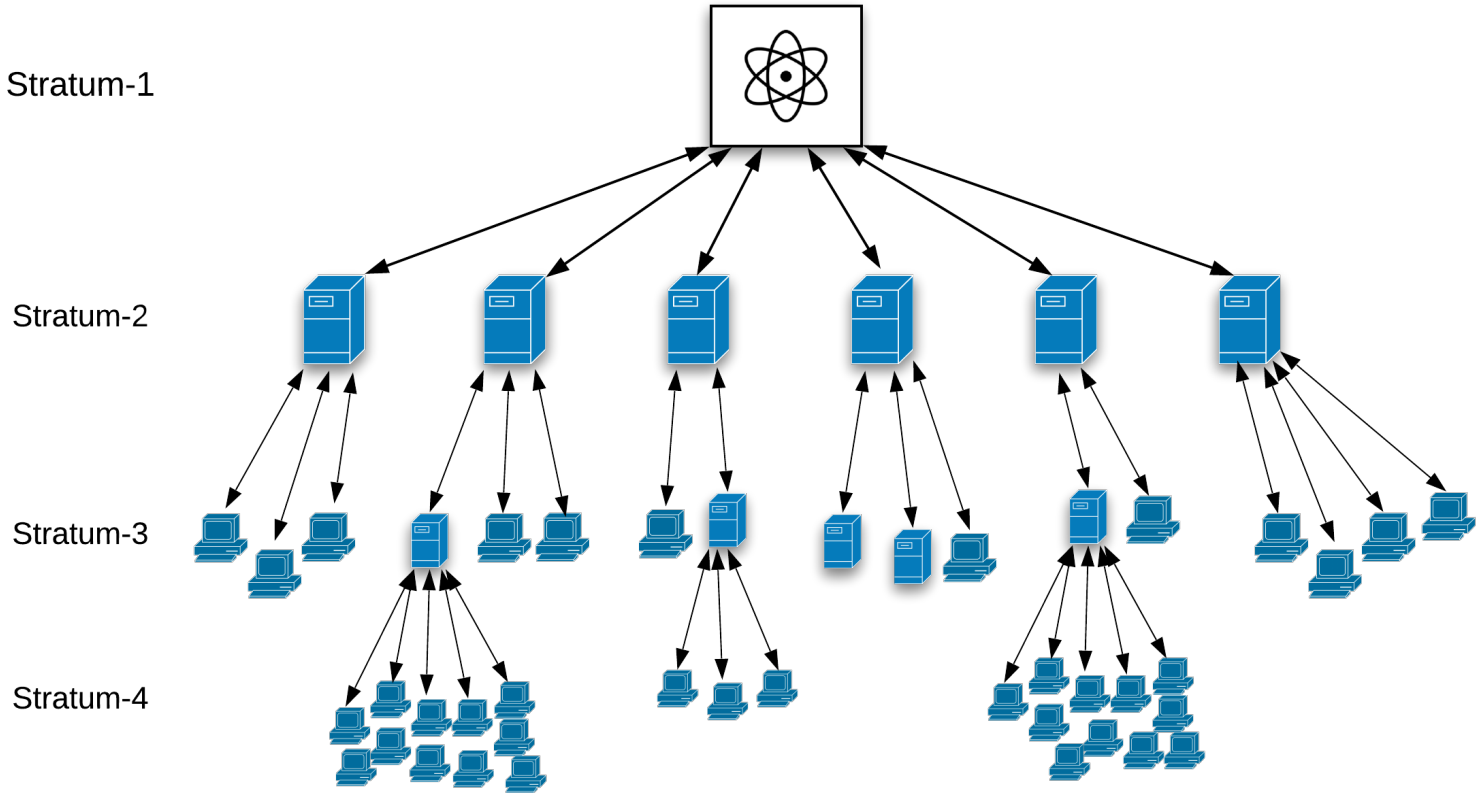
Stratum-1 Servers

- Roots of the network timekeeping forest
- Direct access to a Stratum-0
- Could be network appliance or PC running system clock + ntpd

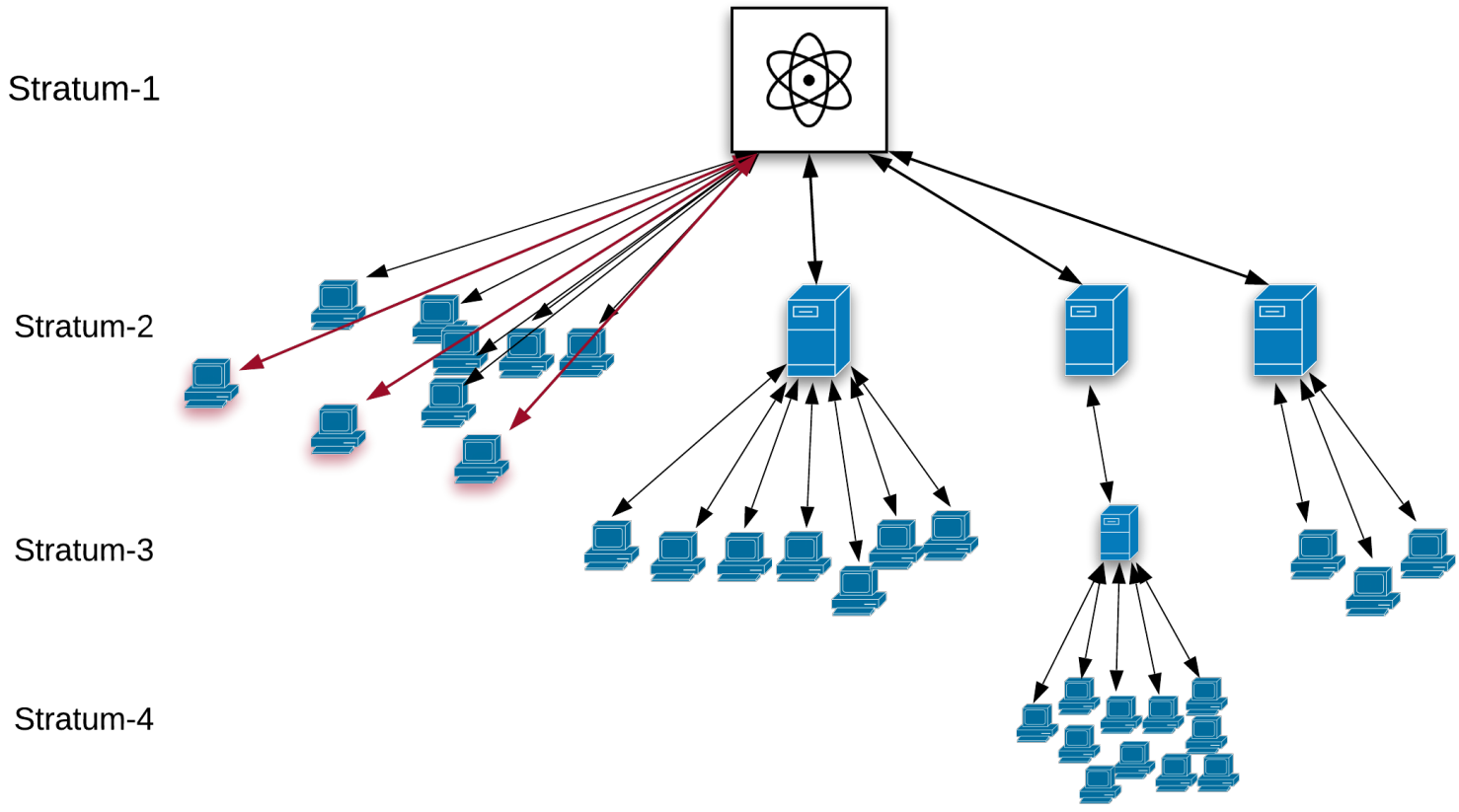
Stratum-n Servers, $n > 1$

- Stratum-n synchronizes over NTP to a Stratum-$(n-1)$ server

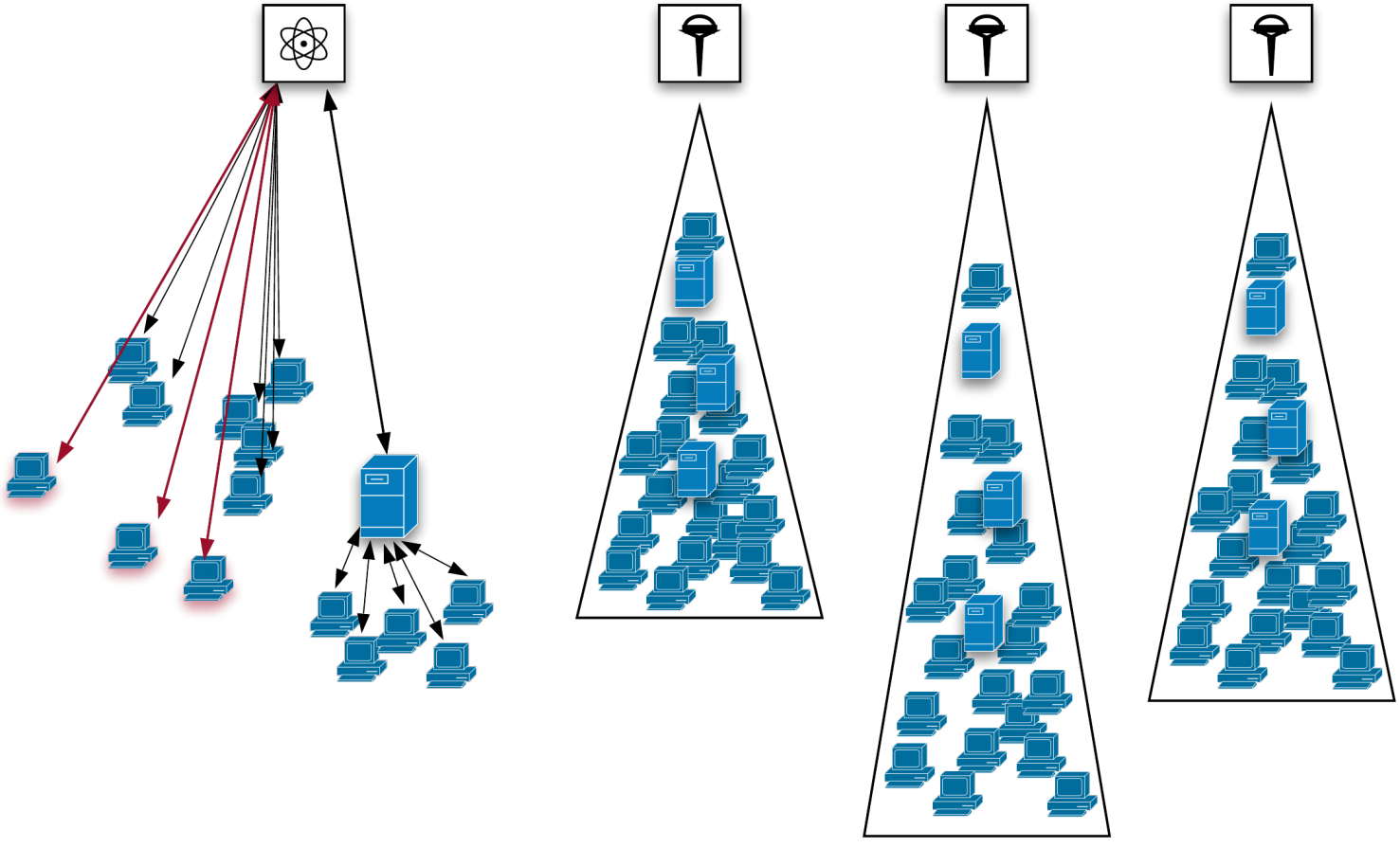
NTP Hierarchy – in principle



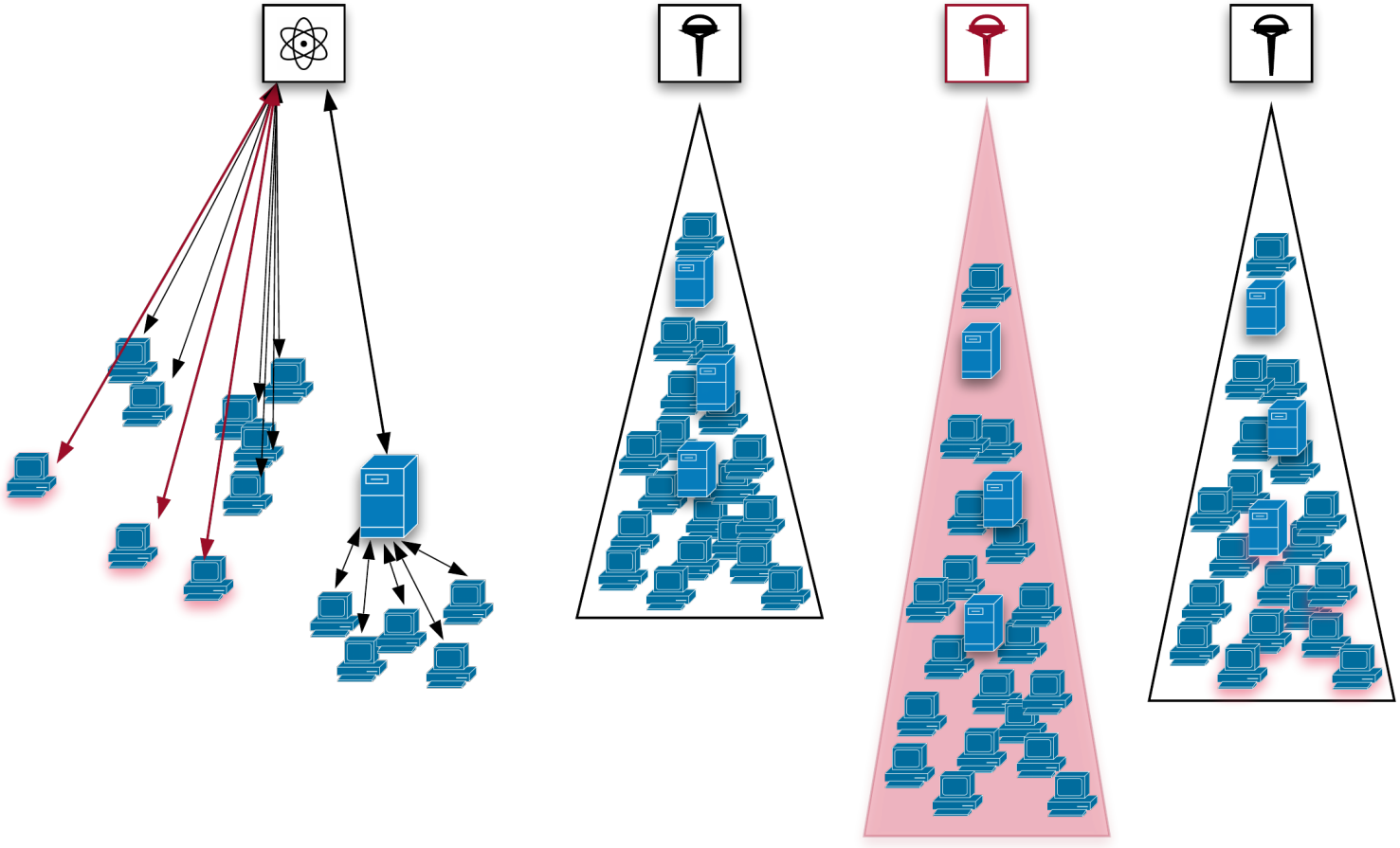
NTP Hierarchy – in reality



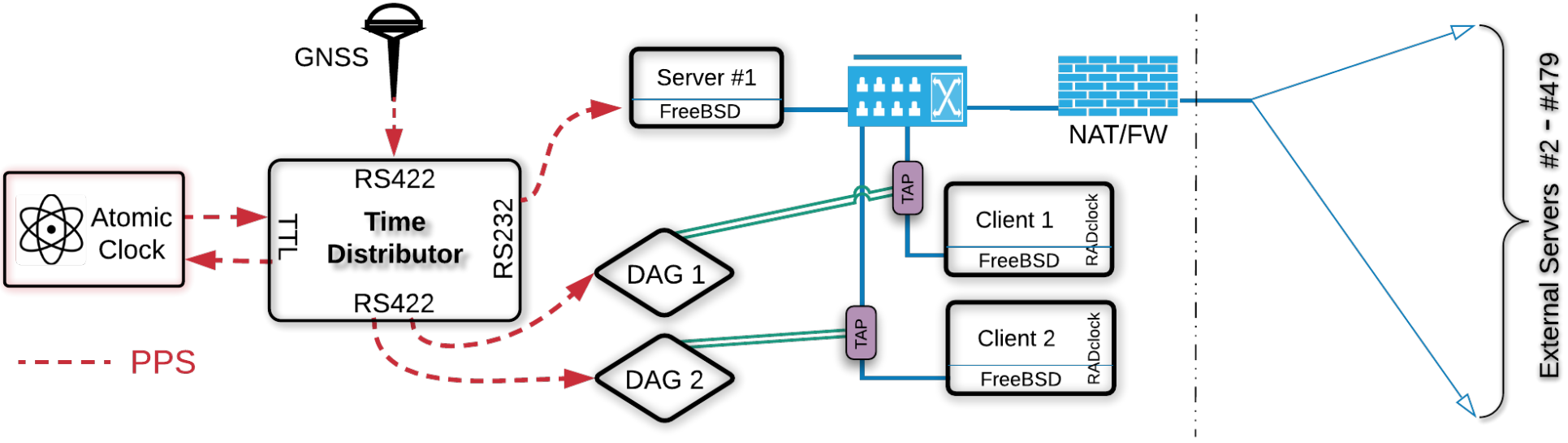
NTP Hierarchy – well actually



NTP Hierarchy – root at the roots?



TESTBED



- Packet timestamping
  - DAG7.5G4 capture card, using passive copper tap
  - roof-mounted Trimble Acutime GG GPS receiver
  - Stanford Research Systems FS725 atomic clock
  - precision $<$100ns
- DAG configuration ignores leap second, acts as 'pre-leap' UTC

Experiment

Experiment

- Over 64 days: 29 Nov. 2016 $\longrightarrow$ 2 Feb. 2017
- Per-server polling server, most with period $\tau = 1$sec
- Load balanced over two client machines

Experiment

Experiment

- Over 64 days: 29 Nov. 2016 $\longrightarrow$ 2 Feb. 2017
- Per-server polling server, most with period $\tau = 1\text{sec}$
- Load balanced over two client machines
- Data problems
  - electricity outage
  - corrupted raw capture file
  - bug causing failure of packet matching

Experiment

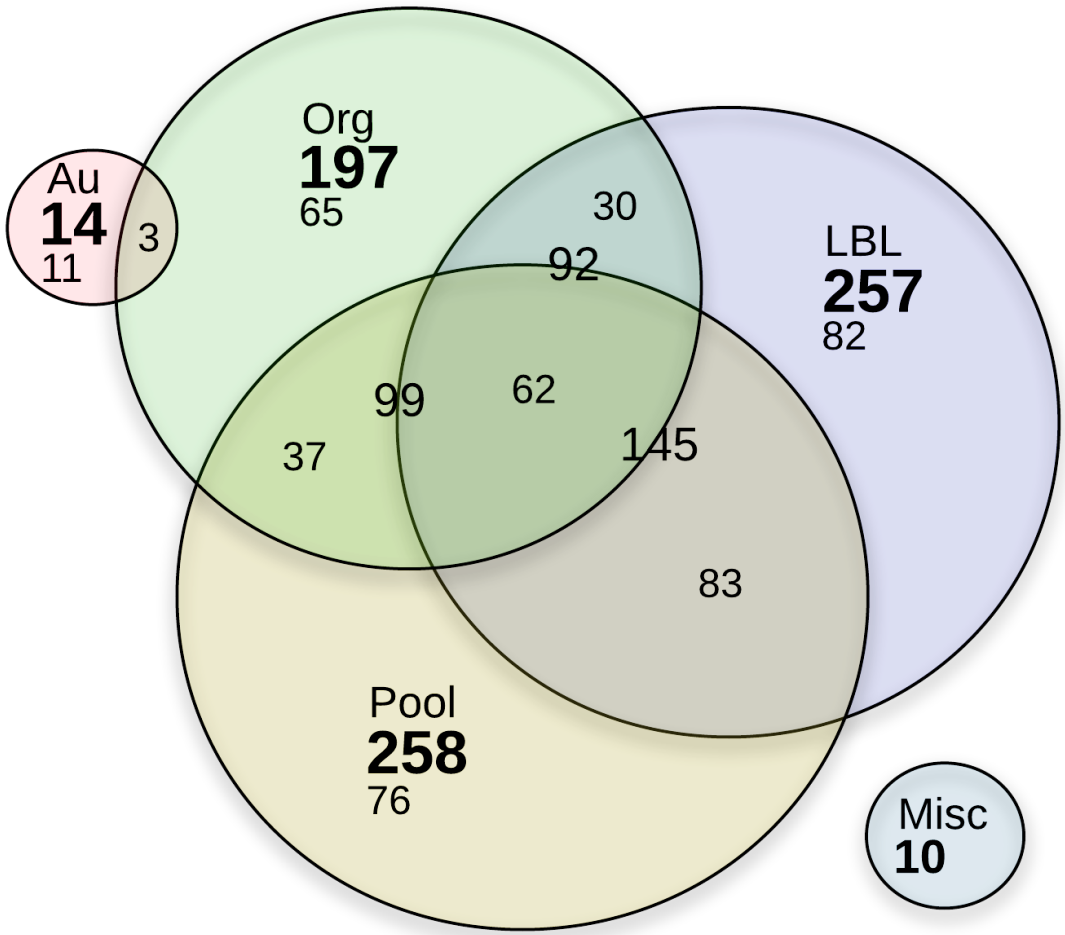
Experiment

- Over 64 days: 29 Nov. 2016 $\longrightarrow$ 2 Feb. 2017
- Per-server polling server, most with period $\tau = 1$sec
- Load balanced over two client machines

Data set for each server

- UTC 4-tuples $\{T_{a,i}, T_{b,i}, T_{e,i}, T_{f,i}\}$ from client<–>server exchange
  – from DAG: $\{T_{a,i}, \quad T_{f,i}\}$
  – from server: $\{ \quad T_{b,i}, T_{e,i} \quad \}$
- LI bits and Stratum level from response pkt headers
- Available at: *https://data.research.uts.edu.au/public/DVTSD/*

SERVER LIST: 459 'STRATUM-1' SERVERS



How Many Stratum-1 Servers?

Marked Sample: $n = |\mathbf{Org} \cup \mathbf{Pool}| = 356$ servers

Random Sample: $K = |\mathbf{LBL}| = 257$ servers

Recaptures: $k = |\mathbf{LBL} \cap (\mathbf{Org} \cup \mathbf{Pool})| = 175$ servers

Population: $N = f(n, K, k)$?

How Many Stratum-1 Servers?

Marked Sample: $n = |\mathbf{Org} \cup \mathbf{Pool}| = 356$ servers

Random Sample: $K = |\mathbf{LBL}| = 257$ servers

Recaptures: $k = |\mathbf{LBL} \cap (\mathbf{Org} \cup \mathbf{Pool})| = 175$ servers

Population: $N = f(n, K, k)$?

Chapman Estimator:

$$\widehat{N} = \lfloor (K+1)(n+1)/(k+1) \rfloor - 1 = 522$$

95% coverage interval:

$$N \in [497, 562]$$

Timestamp Model

$$\text{Forward delay} : D_i^{\uparrow} \equiv T_{b,i} - T_{a,i} = \underline{d}^{\uparrow} + q_i^{\uparrow} + \mathbf{e_i}$$

$$\text{Server delay} : D_i^{\rightarrow} \equiv T_{e,i} - T_{b,i} = \underline{d}^{\rightarrow} + q_i^{\rightarrow}$$

$$\text{Backward delay} : D_i^{\downarrow} \equiv T_{f,i} - T_{e,i} = \underline{d}^{\downarrow} + q_i^{\downarrow} - \mathbf{e_i}$$

$$\text{Round Trip Time (RTT)} : R_i \equiv T_{f,i} - T_{a,i} = \underline{r} + q_i$$

$$\text{Path Asymmetry} : A_i \equiv D_i^{\uparrow} - D_i^{\downarrow} = \underline{a} + a_i + 2\mathbf{e_i}$$

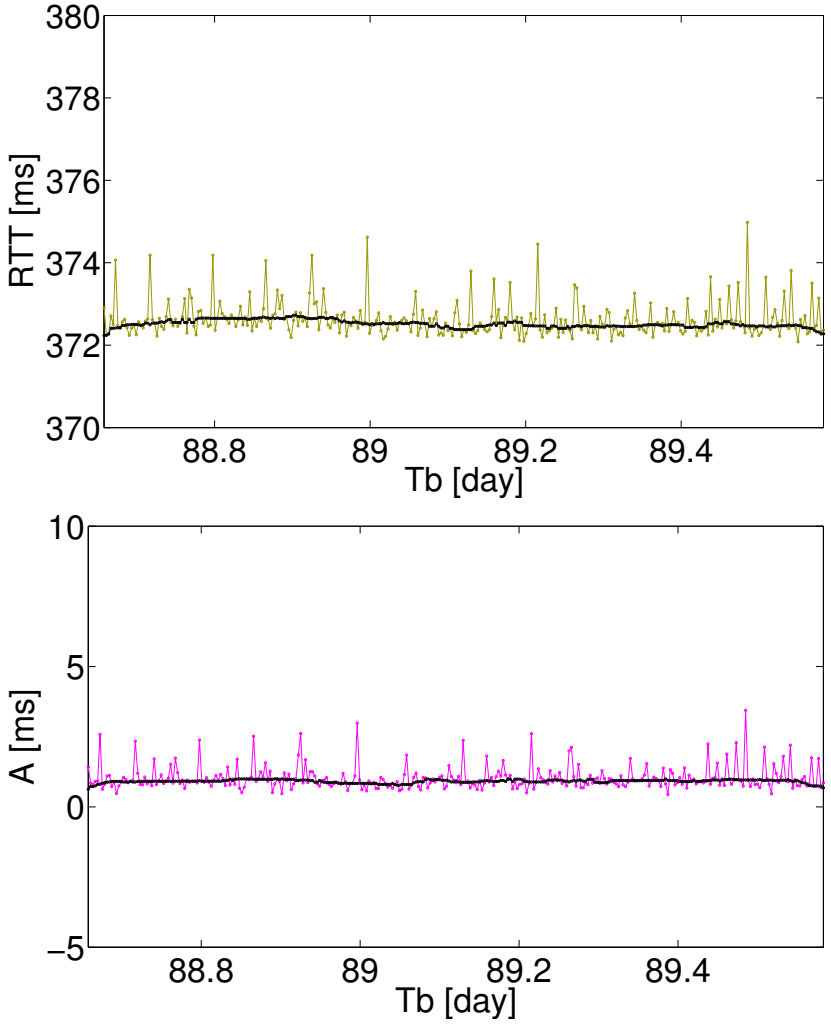
Positive valued OWD, SD, RTT :

- Minima: $\underline{d}^{\uparrow}, \underline{d}^{\rightarrow}, \underline{d}^{\downarrow}, \quad \underline{r} = \underline{d}^{\uparrow} + \underline{d}^{\rightarrow} + \underline{d}^{\downarrow} > 0$
- Congestion Variability: $q_i^{\uparrow}, q_i^{\rightarrow}, q_i^{\downarrow}, \quad q_i = q_i^{\uparrow} + q_i^{\rightarrow} + q_i^{\downarrow}$

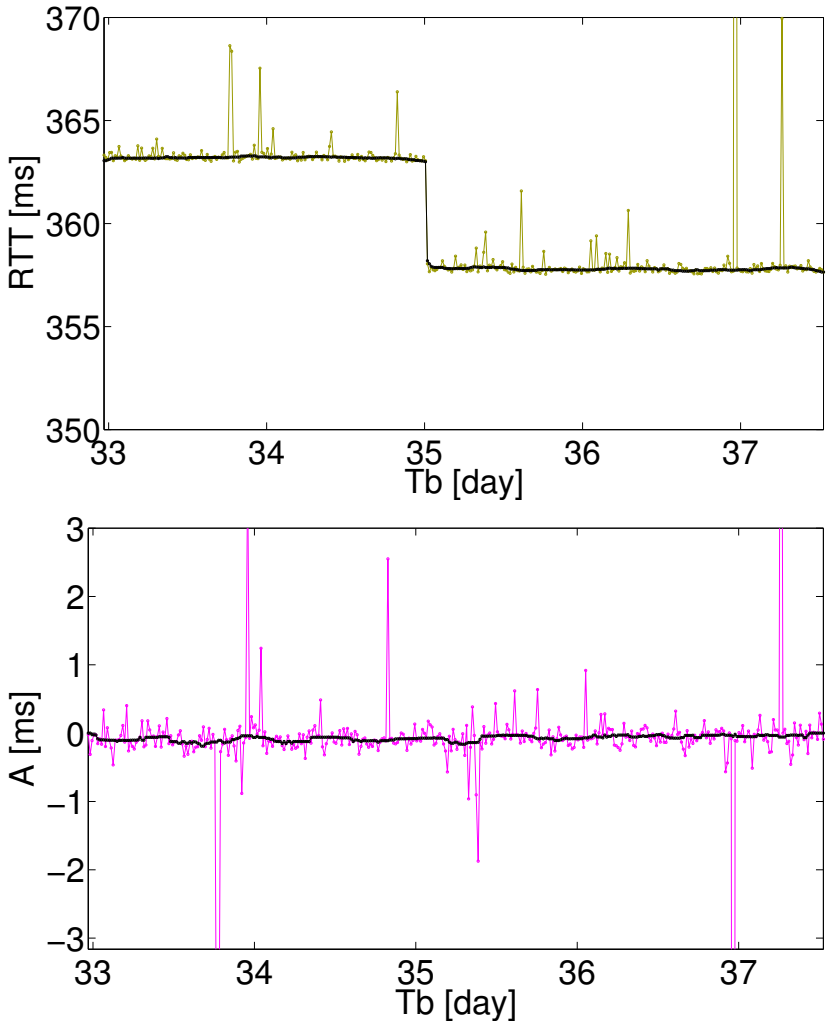
Real valued Asymmetry :

- Real valued underlying: $\underline{a} = \underline{d}^{\uparrow} - \underline{d}^{\uparrow} \in (-\underline{r}, \underline{r})$
- Congestion Variability: $a_i = q_i^{\uparrow} - q_i^{\downarrow}$

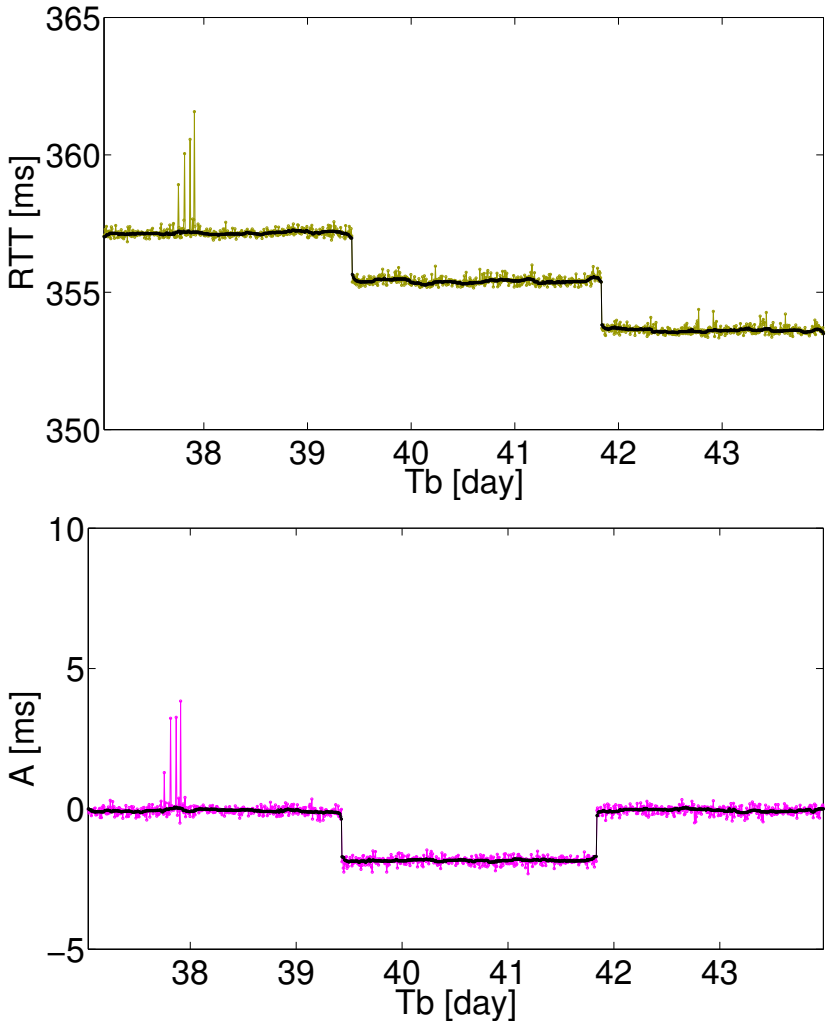
Ideal Case (no server anomaly)



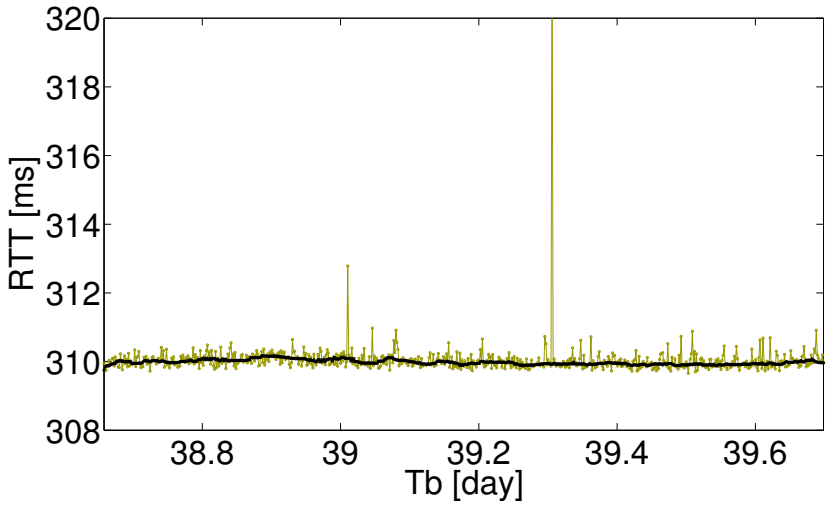
Symmetric Routing Event (no server anomaly)



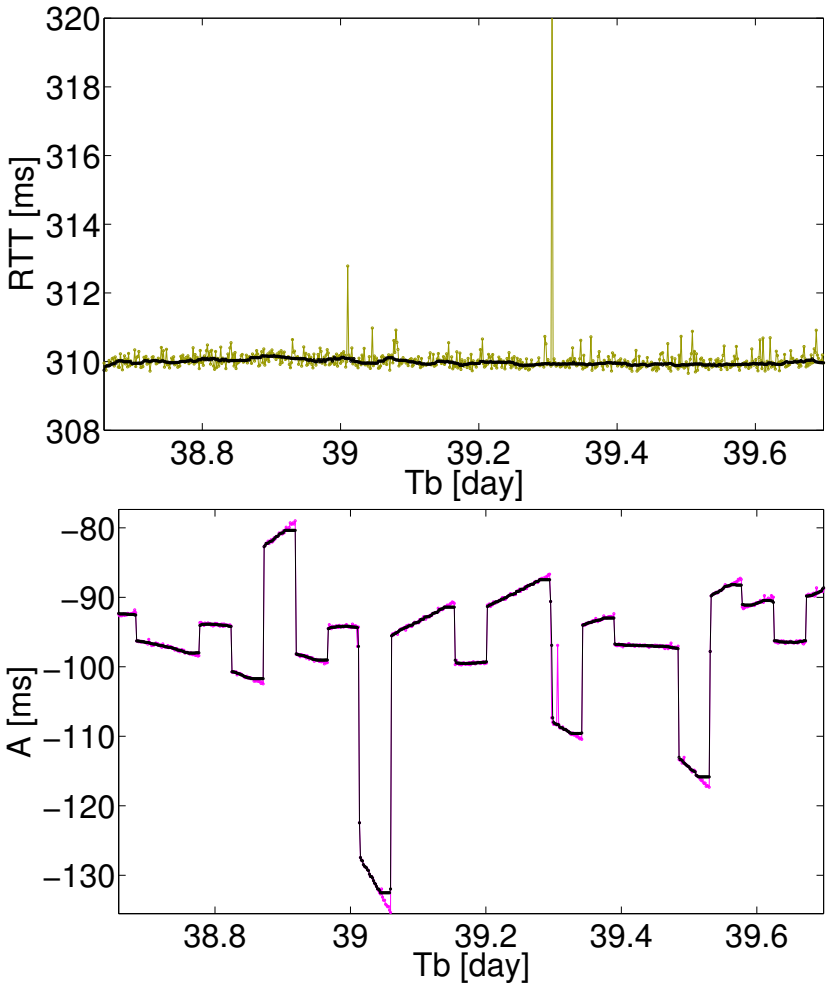
Asymmetric Routing Events (no anomaly)



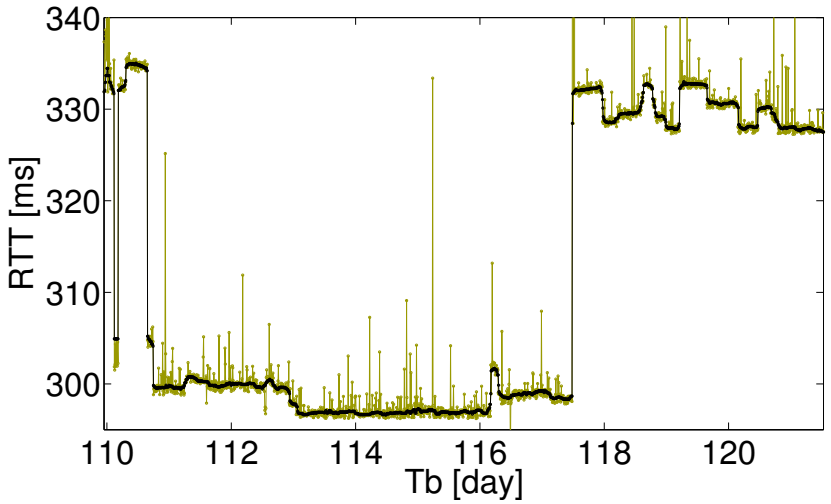
Unmistakeable Anomaly



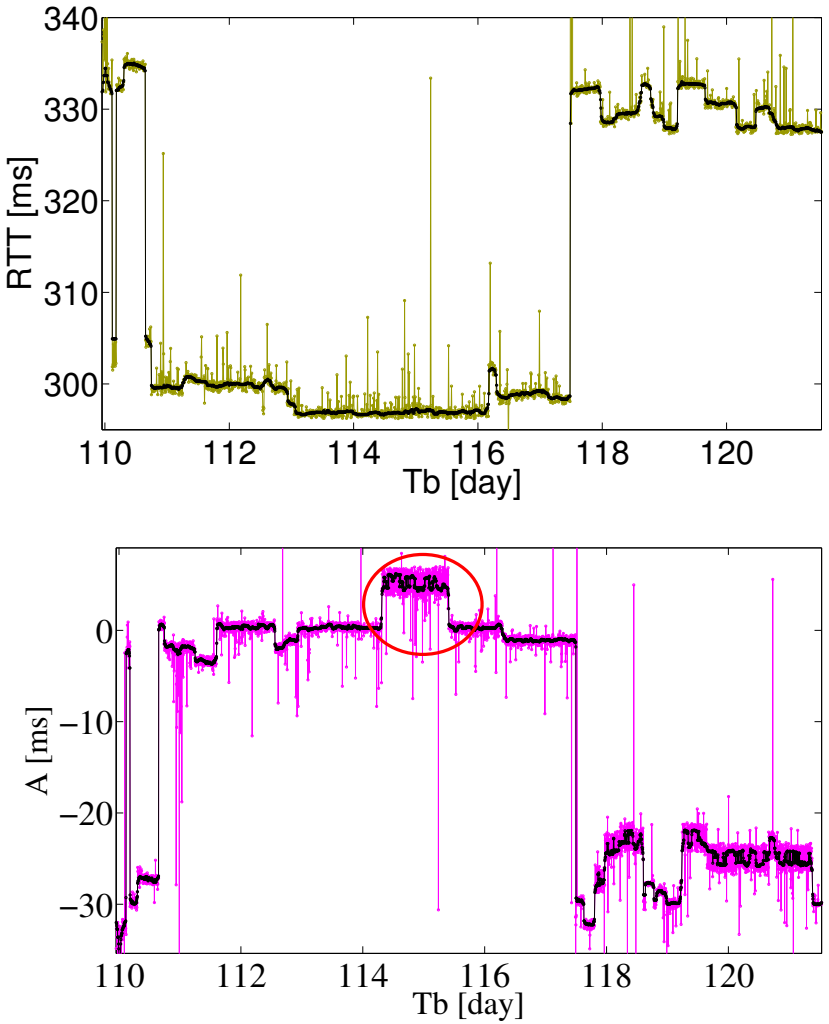
Unmistakeable Anomaly



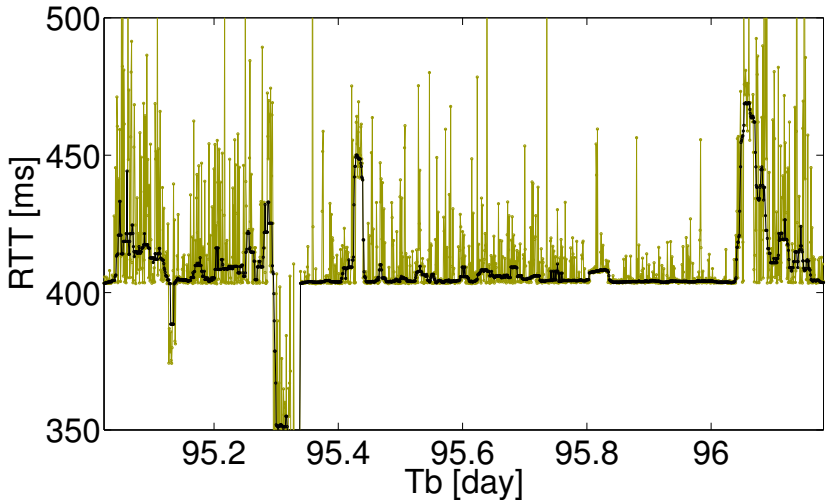
Smaller Anomaly Among Routing Noise



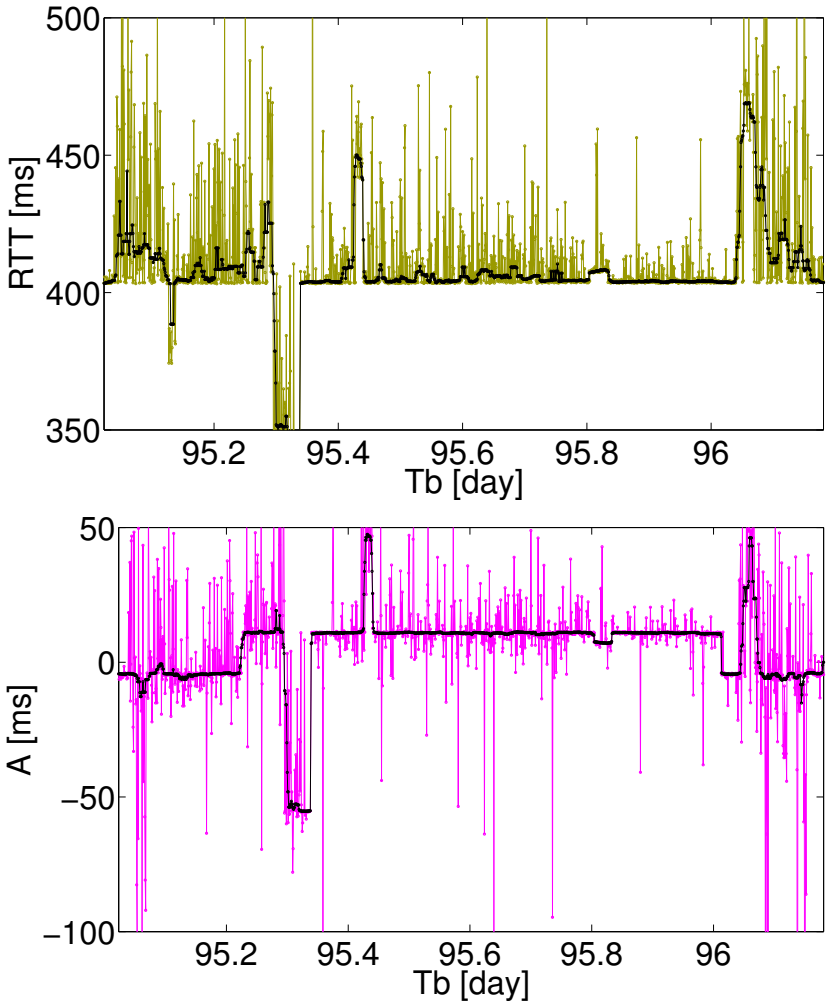
Smaller Anomaly Among Routing Noise



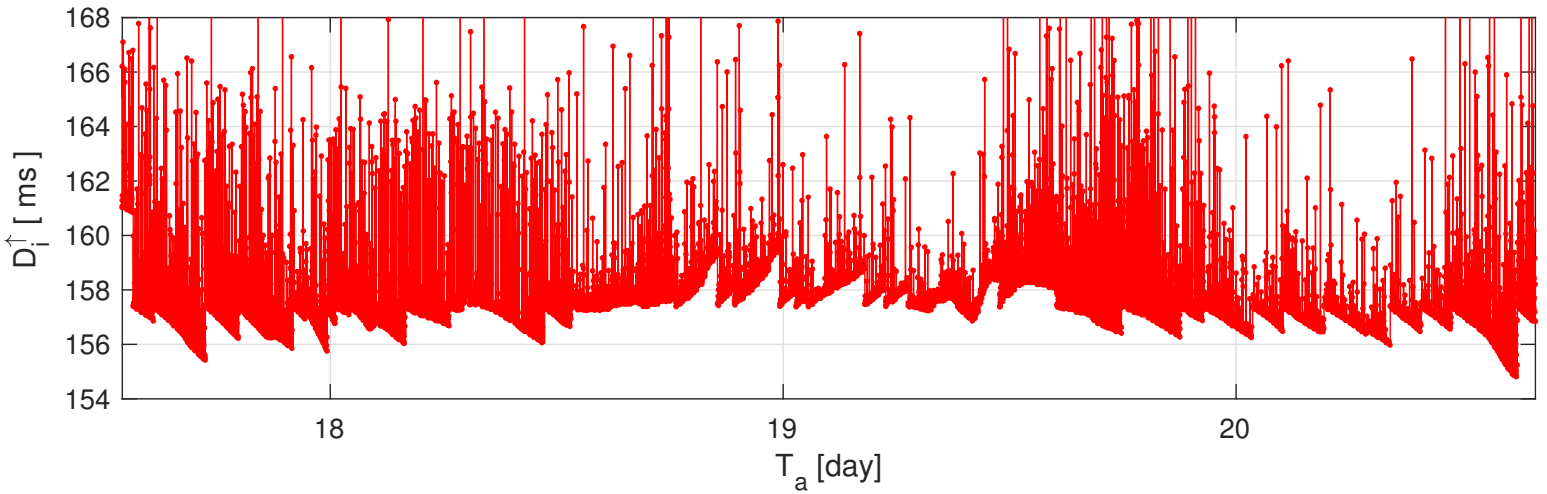
Anomaly with Complex Context



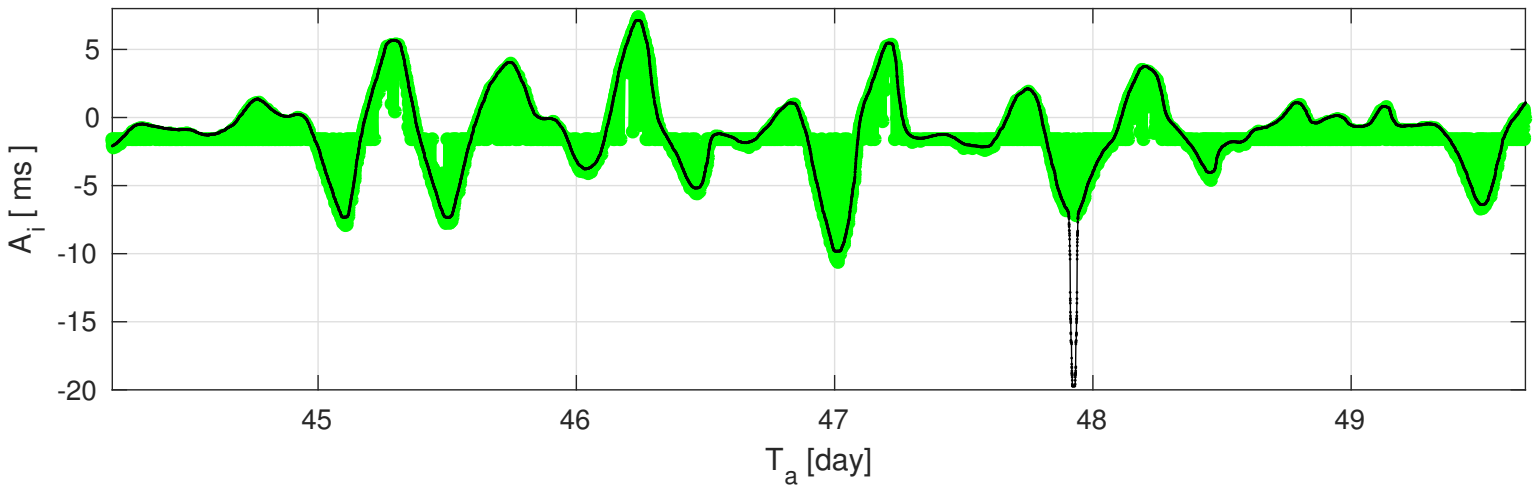
Anomaly with Complex Context



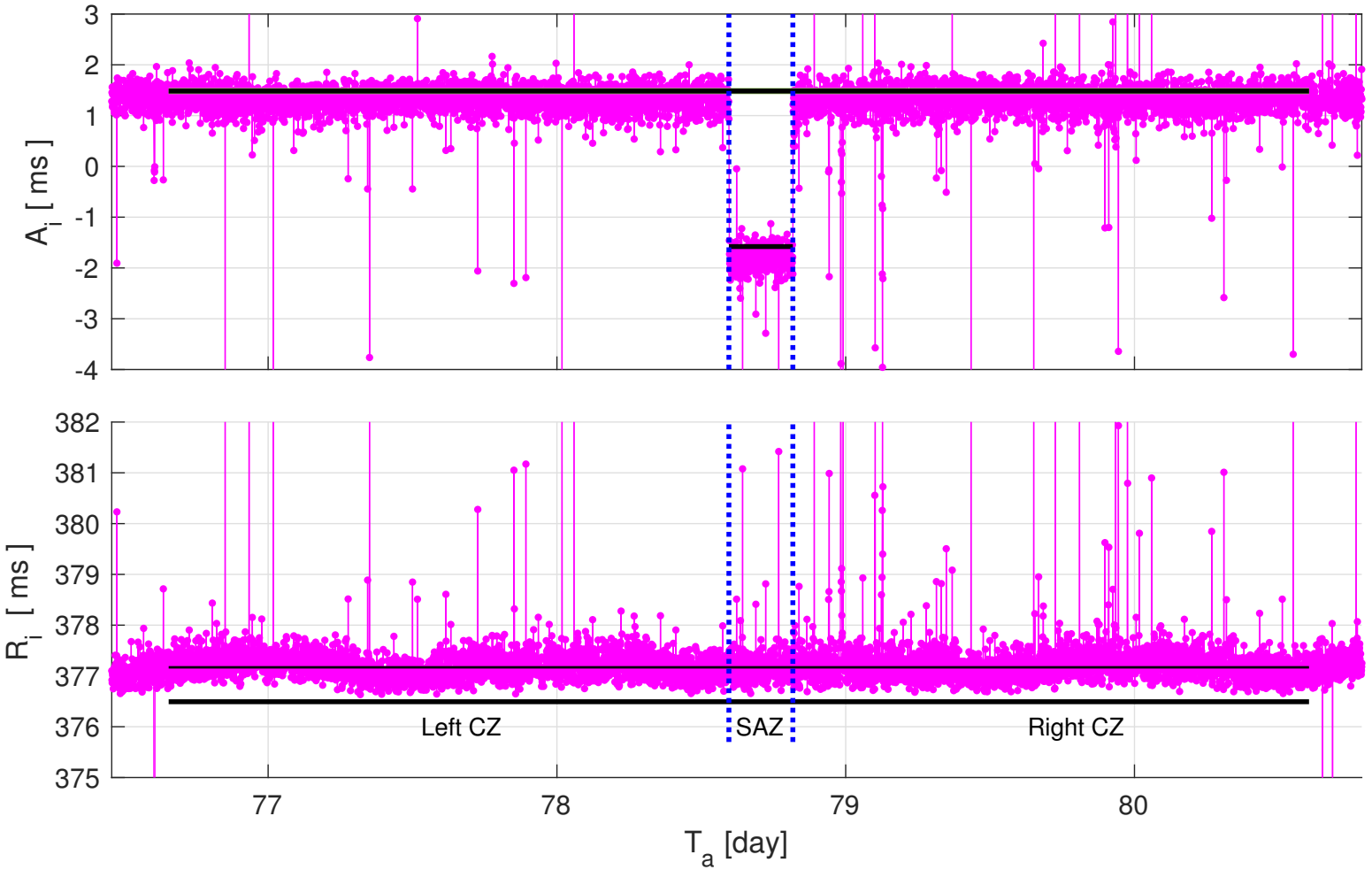
Continuous Skew & Return Errors



Continuous Drift Errors



Error Estimation



Error Estimation

$$\text{Round Trip Time (RTT)} : \; R_i \equiv T_{f,i} - T_{a,i} = \underline{\mathrm{r}} + q_i$$

$$\text{Path Asymmetry} : \; A_i \equiv D_i^{\uparrow} - D_i^{\downarrow} = \underline{\mathrm{a}} + a_i + 2\mathbf{e_i}$$

Error Estimation

$$\text{Round Trip Time (RTT)}: \; R_i \equiv T_{f,i} - T_{a,i} = \underline{\mathrm{r}} + q_i$$

$$\text{Path Asymmetry}: \; A_i \equiv D_i^{\uparrow} - D_i^{\downarrow} = \underline{\mathrm{a}} + a_i + 2\mathbf{e_i}$$

Note $|a_i| < q_i$, estimate $\widehat{q_i} = R_i - \hat{\underline{\mathrm{r}}}$

Adjusted Asymmetry : $\underline{\mathrm{A}}_i(\underline{\mathrm{a}}) = \underline{\mathrm{a}} + \frac{A_i - \underline{\mathrm{a}}}{|A_i - \underline{\mathrm{a}}|} \max(0, |A_i - \underline{\mathrm{a}}| - q_i)$

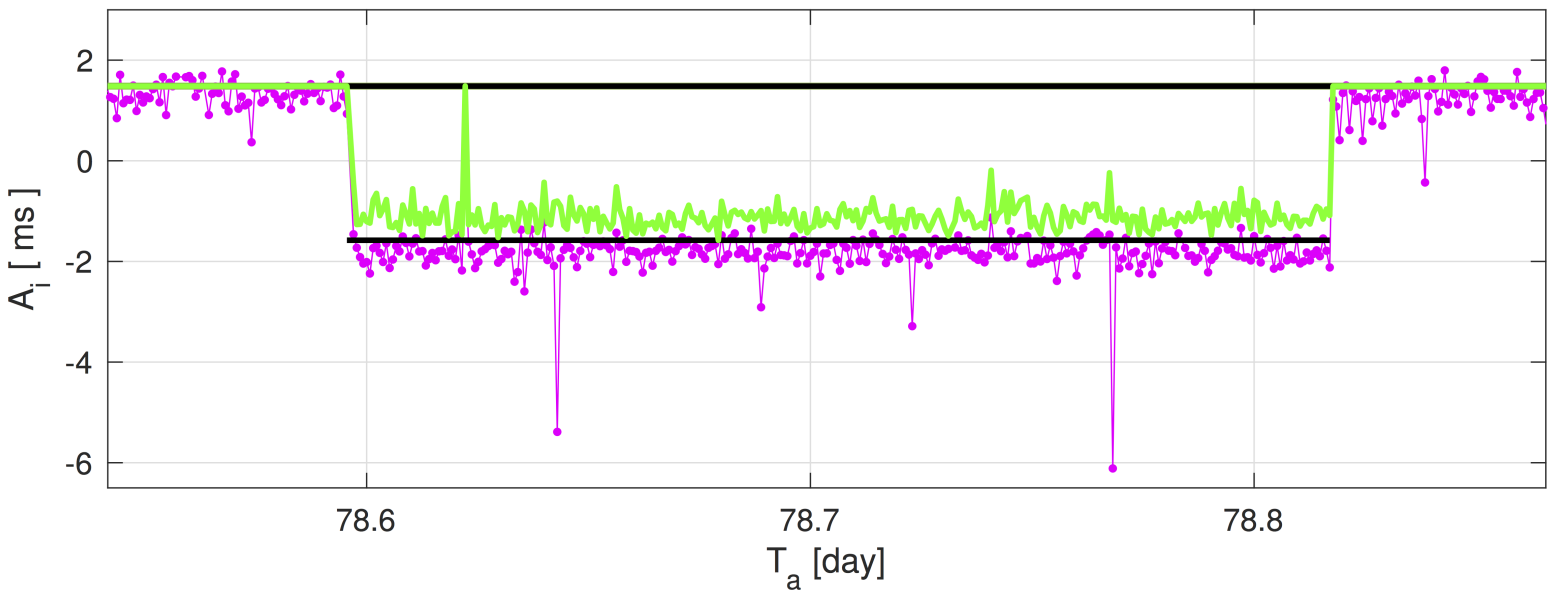
Error Estimation

$$\text{Round Trip Time (RTT)} : \; R_i \equiv T_{f,i} - T_{a,i} = \underline{\mathrm{r}} + q_i$$

$$\text{Path Asymmetry} : \; A_i \equiv D_i^{\uparrow} - D_i^{\downarrow} = \underline{\mathrm{a}} + a_i + 2\mathbf{e_i}$$

Note $|a_i| < q_i$, estimate $\widehat{q_i} = R_i - \hat{\underline{\mathrm{r}}}$

Adjusted Asymmetry : $\underline{\mathrm{A}}_i(\underline{\mathrm{a}}) = \underline{\mathrm{a}} + \frac{A_i - \underline{\mathrm{a}}}{|A_i - \underline{\mathrm{a}}|} \max(0, |A_i - \underline{\mathrm{a}}| - q_i)$



Relative Error (Simulation)

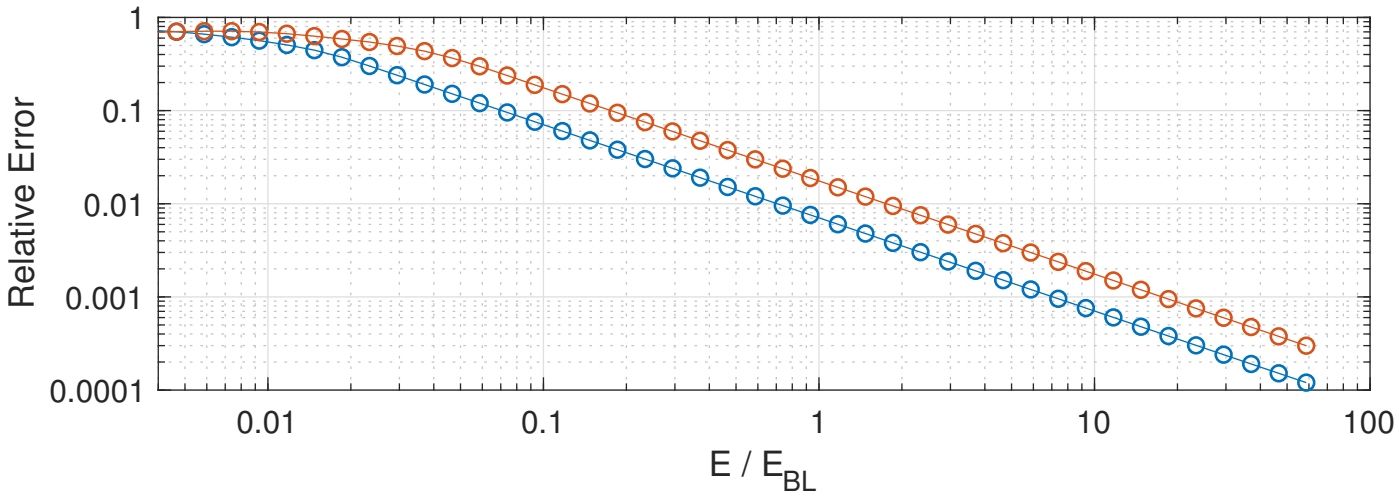


Figure: Relative Error of $\hat{E}$ as a function of the true error size E normalized by the conservative error bound E_{BL}. Top curve: $N = 100$, Bottom: $N = 500$.

Longitudinal Results

Exp	Target Servers (#)	Start	End	duration
Exp1	**List1** (119)	Mar. 4, 2011	Aug. 3, 2011	151 days
Exp2	**List2** (117)	Dec. 5, 2014	April 2, 2015	124 days
Exp3	**List3** (479)	Nov. 29, 2016	Feb. 2, 2017	64 days

Longitudinal Results

Exp	Target Servers (#)	Start	End	duration
Exp1	**List1** (119)	Mar. 4, 2011	Aug. 3, 2011	151 days
Exp2	**List2** (117)	Dec. 5, 2014	April 2, 2015	124 days
Exp3	**List3** (479)	Nov. 29, 2016	Feb. 2, 2017	64 days

Basic Server Status

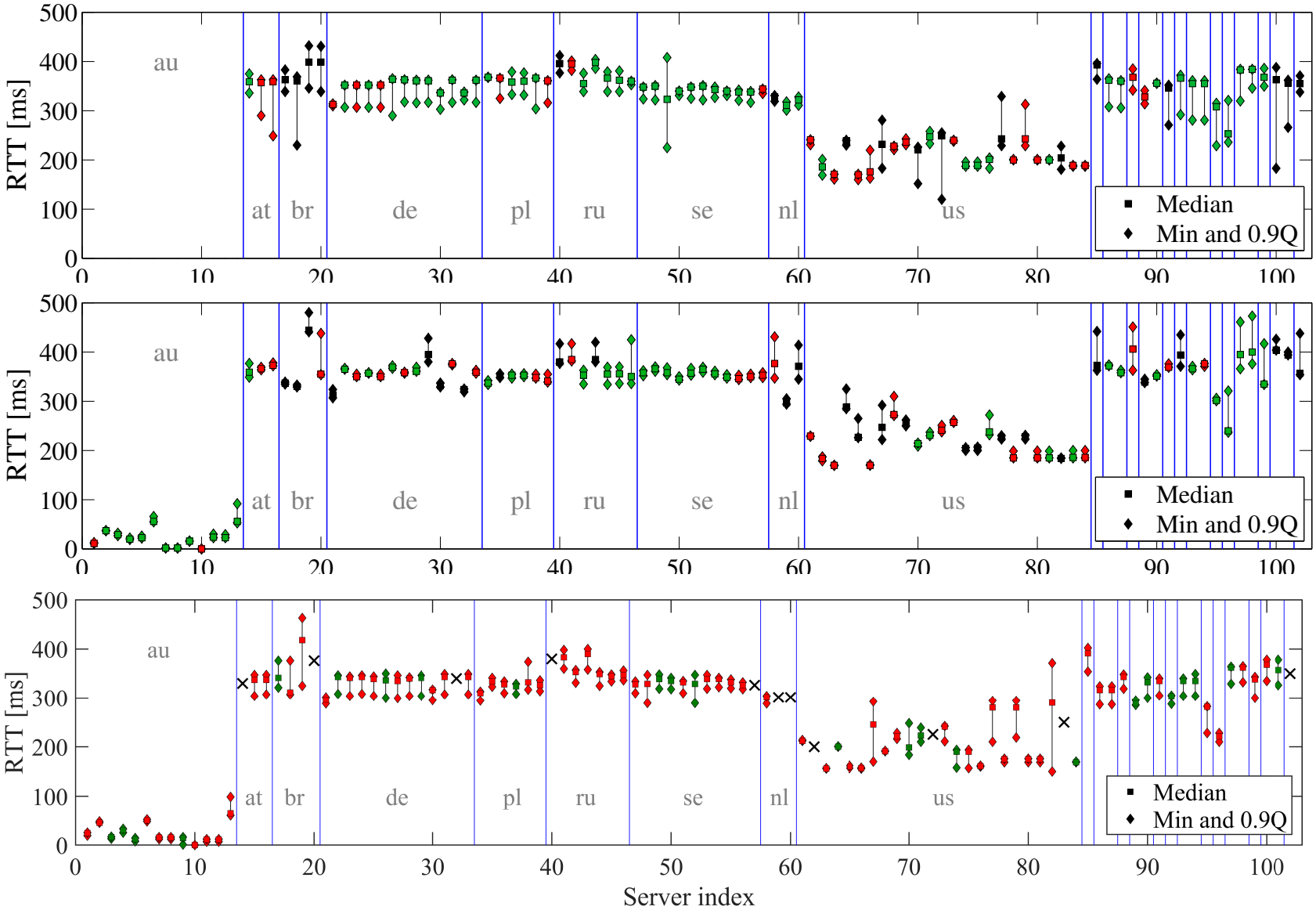
E(rrored): at least one detected error

- **R**(are): less than 1 error per week
- **C**(ommon): more than 1 error per week but not **H**igh
- **H**(igh): in error more than 25% of the time

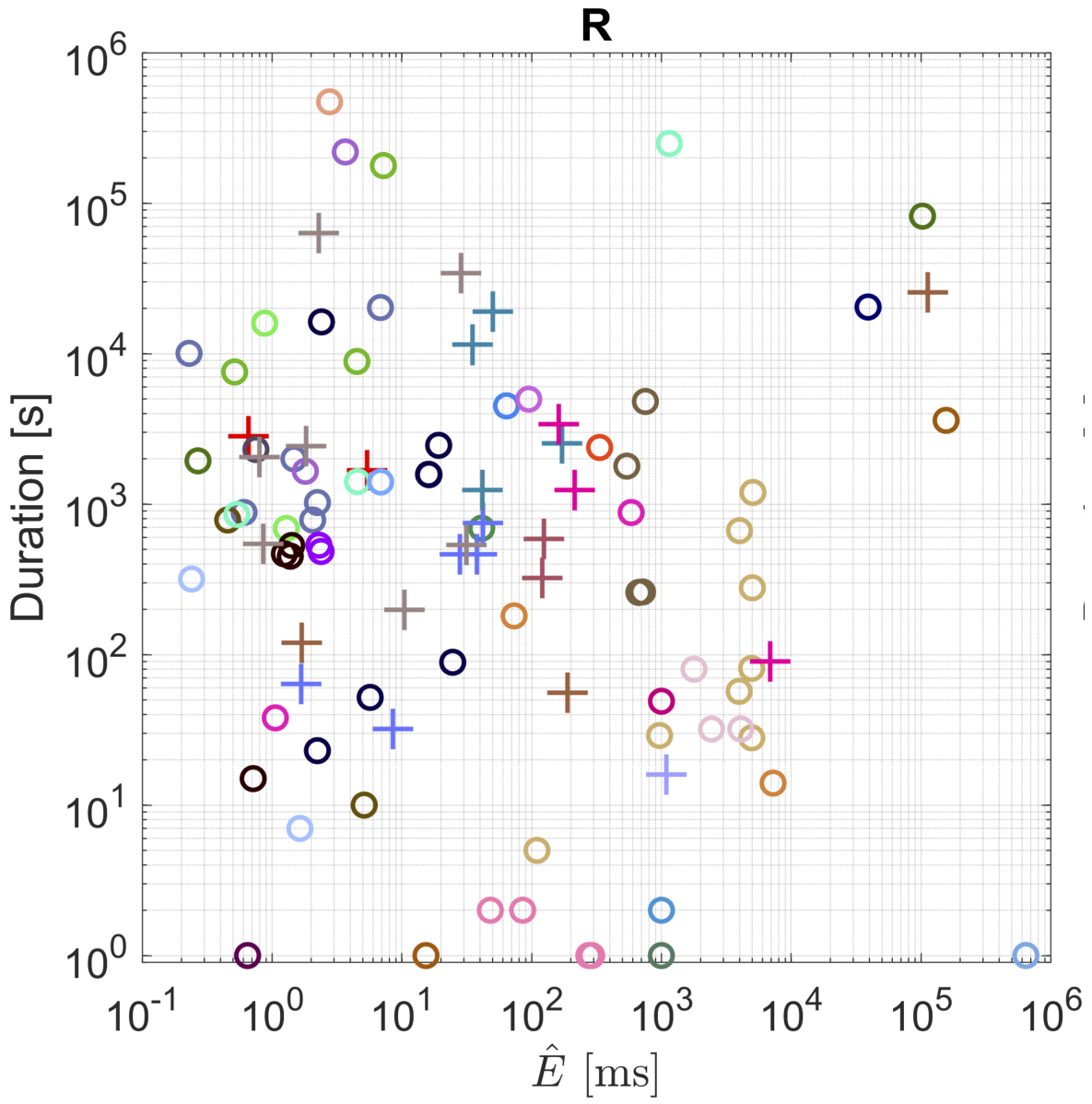
G(ood): no sign of a server error

A(mbiguous): evidence of at least one error, but none certain

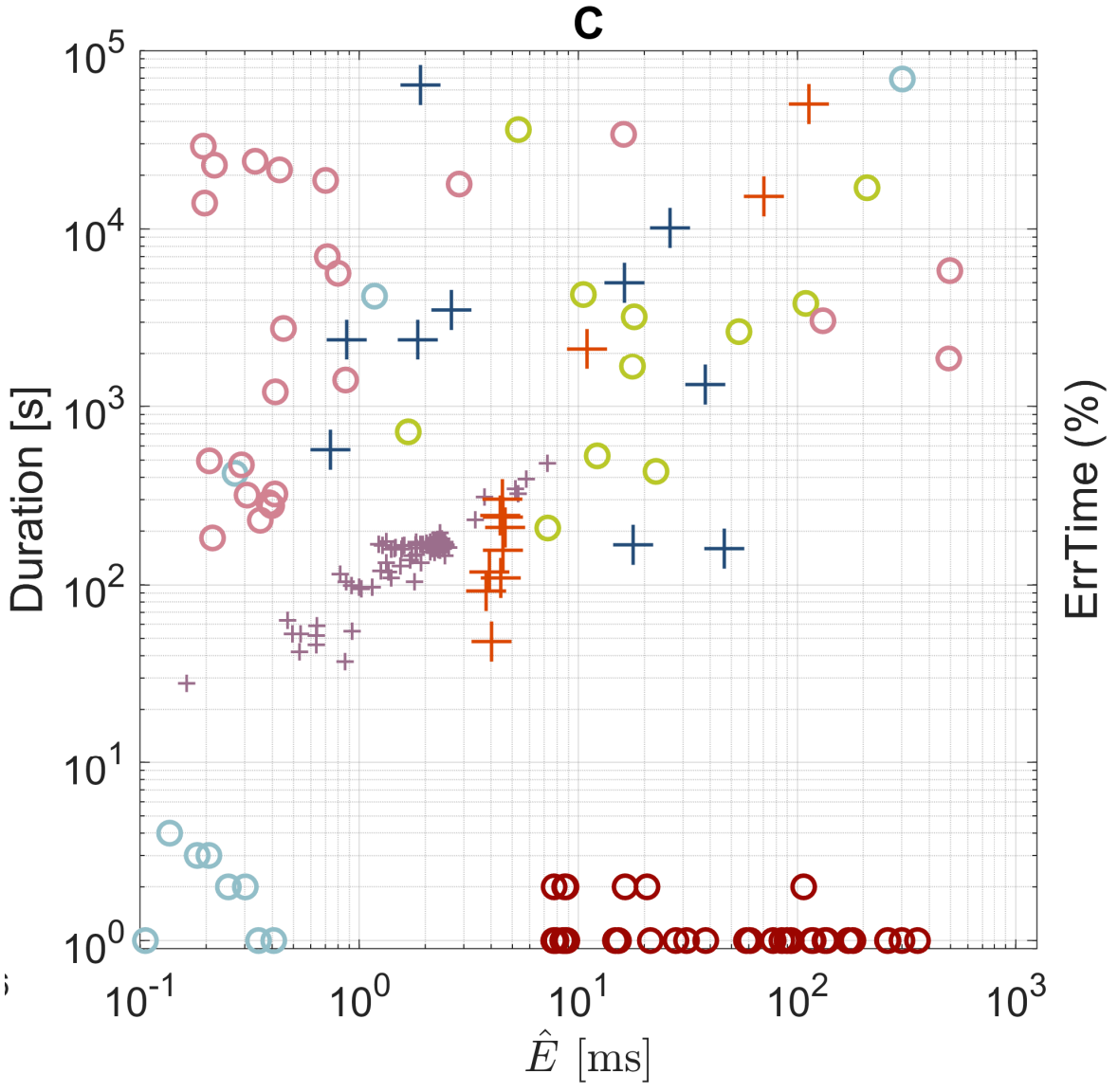
GOOD, BAD, UGLY



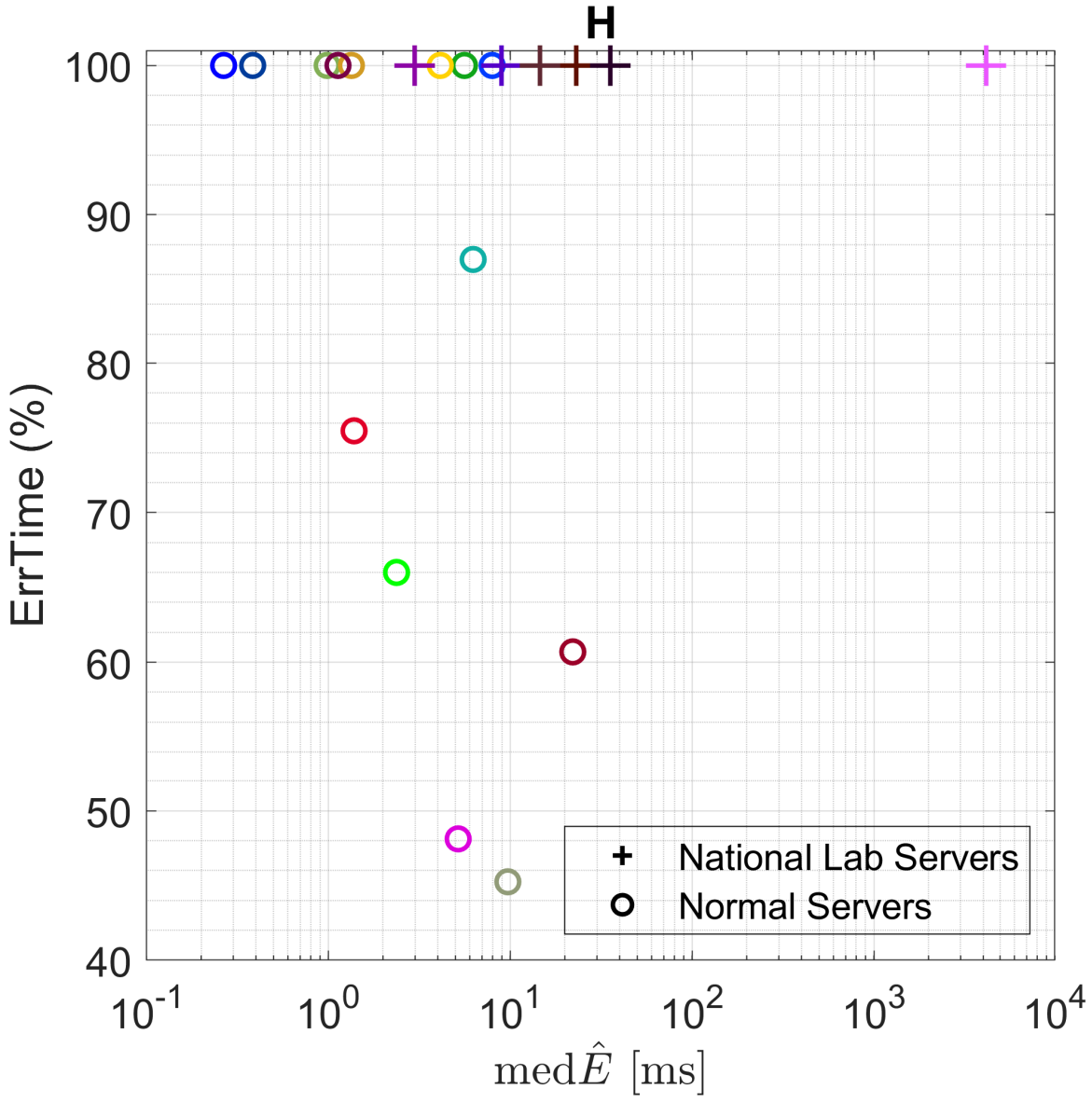
New Data: Servers with Rare Errors



New Data: Servers with Common Errors



New Data: Servers with 'High' Errors



Does the Protocol Know?

$\uparrow$-type

$P_{\uparrow}$: $\mathbb{SN}\ [\uparrow]+\ \mathbb{SN}$

$P_{\uparrow,\oslash}$: $\mathbb{SN}\ [\uparrow]+\ \oslash\ [\uparrow\ \oslash]*\ \uparrow\ \mathbb{SN}$

$\downarrow$-type

$P_{\downarrow}$: $\mathbb{SN}\ [\downarrow]+\ \mathbb{SN}$

L-type

P_{L} : $\mathbb{SN}\ [L]+\ \mathbb{SN}$

$P_{L,\uparrow}$: $\mathbb{SN}\ [L]+\ \uparrow\ [L\ \uparrow]*\ \mathbb{SN}$

$\oslash$-type

$P_{\oslash}$: $\mathbb{SN}\ [\oslash]+\ \mathbb{SN}$

$P_{\oslash,L}$: $\mathbb{SN}\ [\oslash]+\ L\ [\oslash L]*\ \mathbb{SN}$

$P_{\oslash,\uparrow}$: $\mathbb{SN}\ [\oslash]+\ \uparrow\ [\oslash\ \uparrow]*\ \mathbb{SN}$

Does the Protocol Know?

↑-type

$P_{\uparrow}$: $\mathbb{SN}\ [\uparrow]+\ \mathbb{SN}$

$P_{\uparrow,\oslash}$: $\mathbb{SN}\ [\uparrow]+\ \oslash\ [\uparrow\ \oslash]*\ \uparrow\ \mathbb{SN}$

↓-type

$P_{\downarrow}$: $\mathbb{SN}\ [\downarrow]+\ \mathbb{SN}$

L-type

P_{L} : $\mathbb{SN}\ [L]+\ \mathbb{SN}$

$P_{L,\uparrow}$: $\mathbb{SN}\ [L]+\ \uparrow\ [L\ \uparrow]*\ \mathbb{SN}$

⊘-type

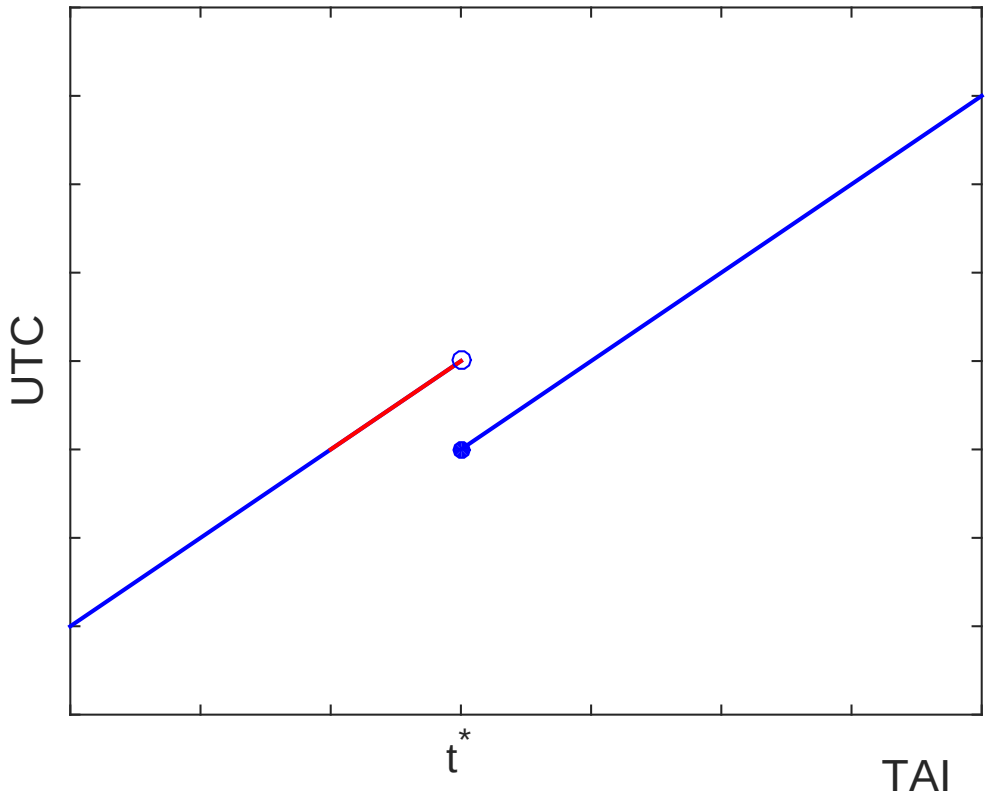
$P_{\oslash}$: $\mathbb{SN}\ [\oslash]+\ \mathbb{SN}$

$P_{\oslash,L}$: $\mathbb{SN}\ [\oslash]+\ L\ [\oslash L]*\ \mathbb{SN}$

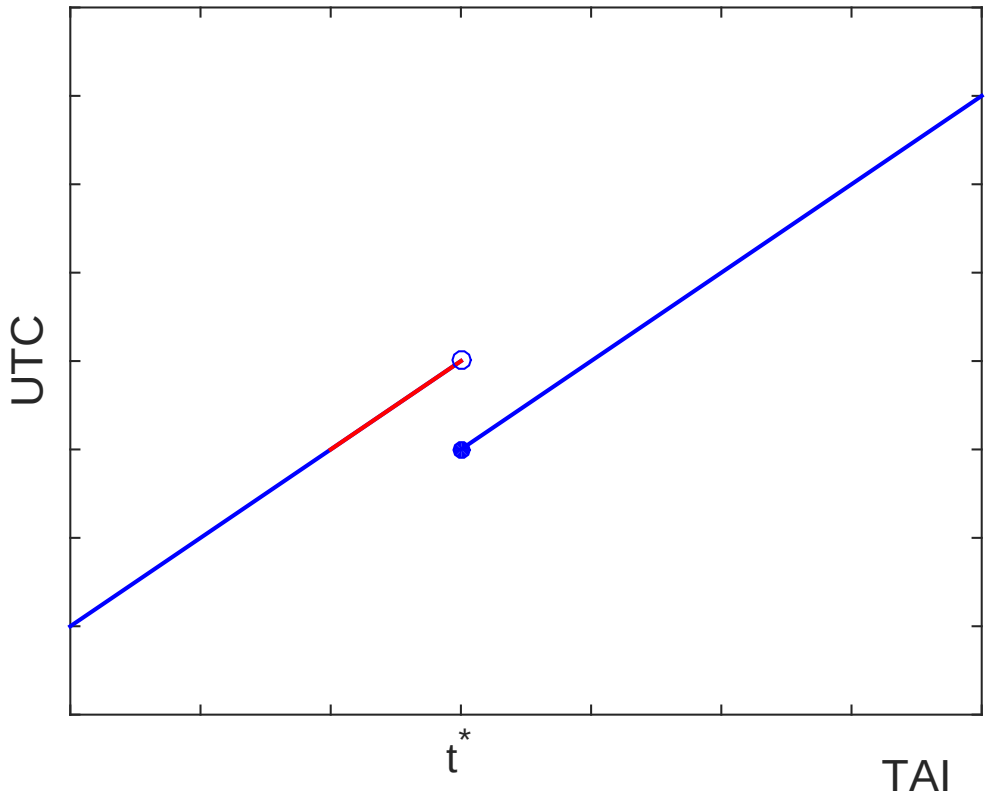
$P_{\oslash,\uparrow}$: $\mathbb{SN}\ [\oslash]+\ \uparrow\ [\oslash\ \uparrow]*\ \mathbb{SN}$

70% of Errors have no warning

UTC Leap Second of end-2016

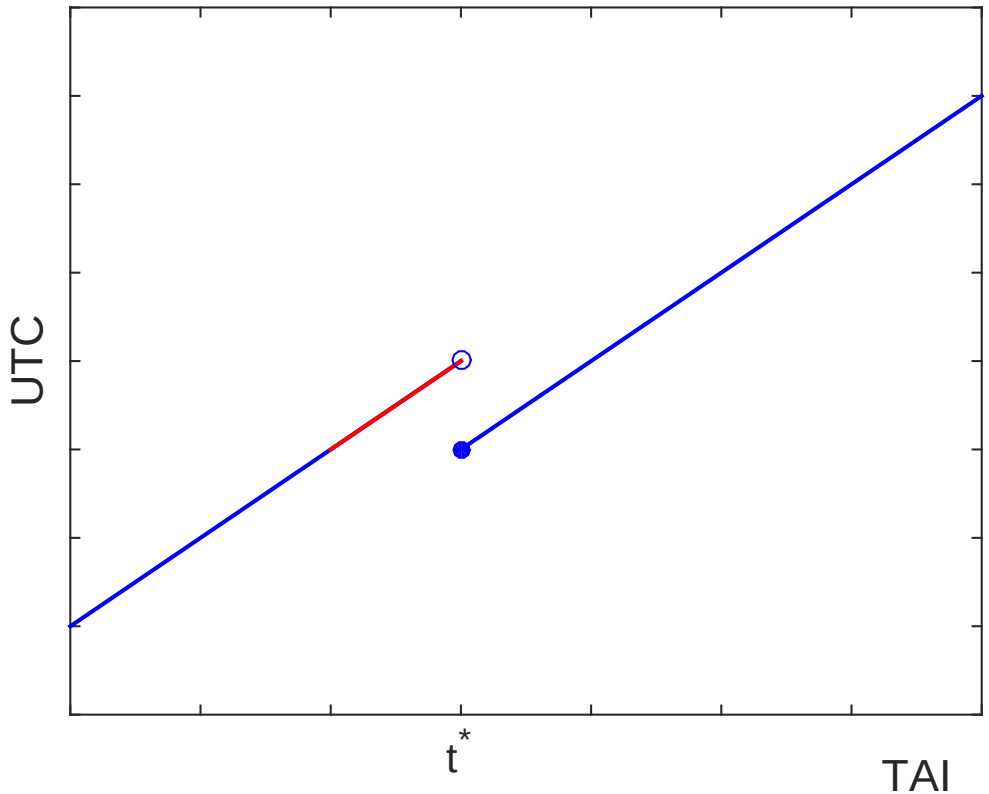


UTC Leap Second of end-2016



- Positive leap second (extra second added = UTC slows down)

UTC Leap Second of end-2016



- Positive leap second (extra second added = UTC slows down)
- Occurred at $t^*_{\text{TAI}} = 1483228837$ sec, when $t^*_{\text{UTC}} = t^*_{\text{TAI}} - 37$ (00:00:00 January 1st UTC)

Server Quality during Leap Second Events

Performance Dimensions

- Clock Compliance: clean jump at t^*?
- Protocol Compliance: LI bits set appropriately?
- Stratum Behaviour: as expected? consistent with LI bits?

Measuring Clock Compliance

Goal: to quantify deviations from ideal server clock behaviour

Measuring Clock Compliance

Goal: to quantify deviations from ideal server clock behaviour

Metrics: Time of Jump (ToJ)
Time to Expected Behaviour (TEB)
Leap Error Duration : $\text{LED} = \max(0, \text{TEB}) - \min(0, \text{ToJ})$

Measuring Clock Compliance

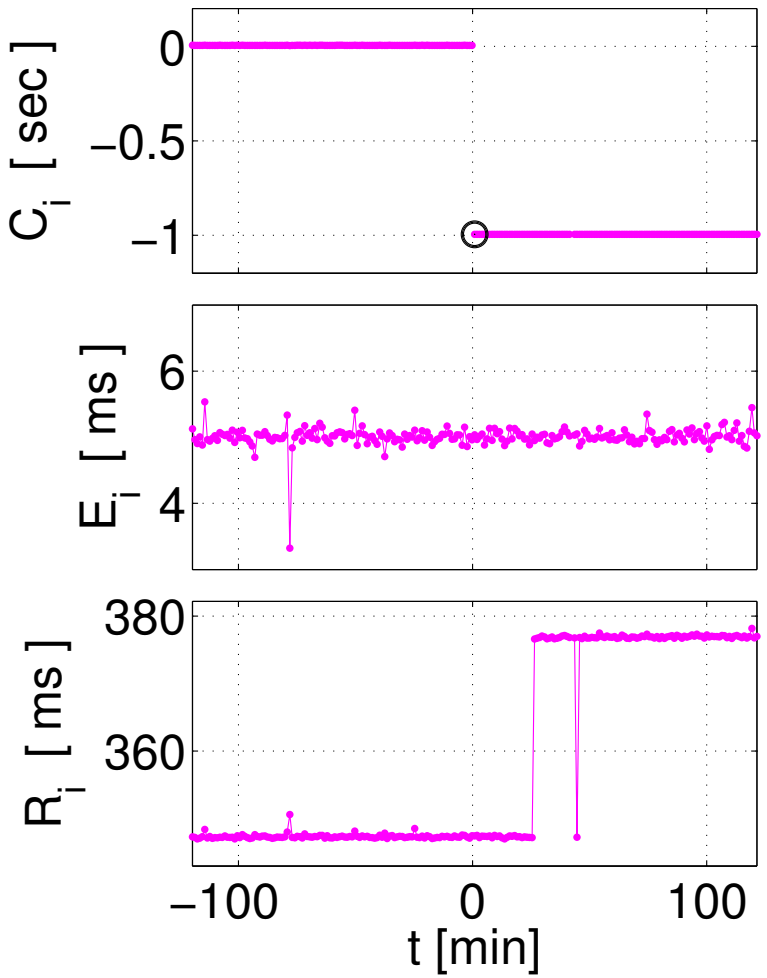
Goal: to quantify deviations from ideal server clock behaviour

Metrics: Time of Jump (ToJ)
Time to Expected Behaviour (TEB)
Leap Error Duration : $\text{LED} = \max(0, \text{TEB}) - \min(0, \text{ToJ})$

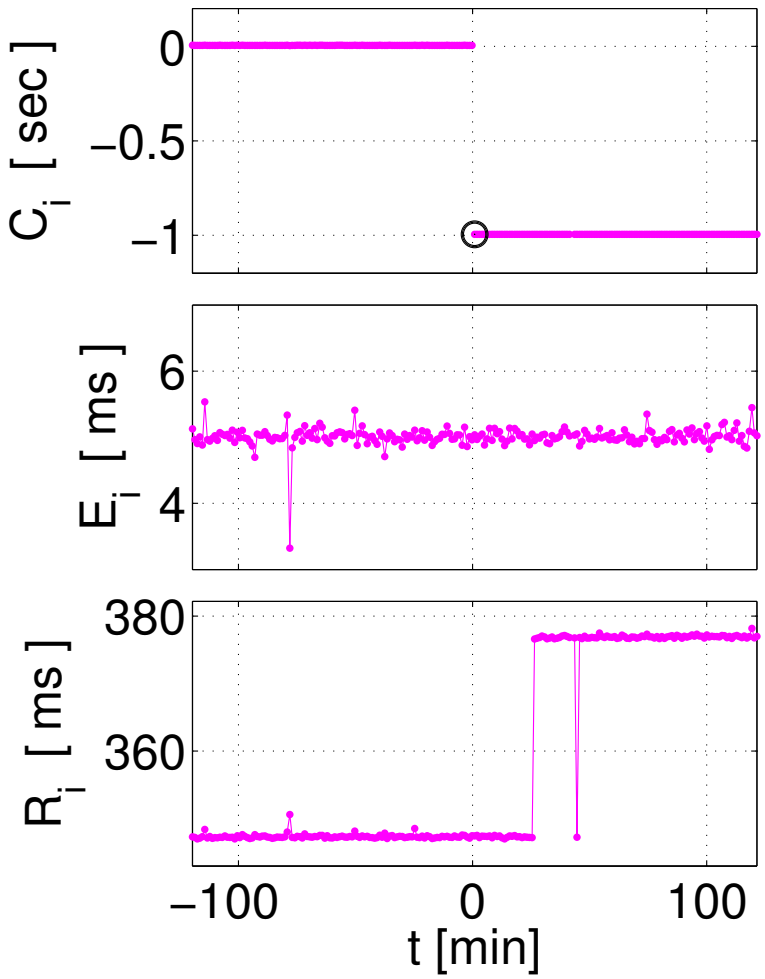
Key time series based on $\{T_{a,i}, T_{b,i}, T_{e,i}, T_{f,i}\}$:

$$\begin{aligned}
\text{Forward delay} &: D_i^{\uparrow} = T_{b,i} - T_{a,i} \\
\text{Backward delay} &: D_i^{\downarrow} = T_{f,i} - T_{e,i} \\
\text{Round Trip Time (RTT)} &: R_i = T_{f,i} - T_{a,i} \\
\text{Server Change} &: C_i = (D_i^{\uparrow} - D_i^{\downarrow})/2 \qquad \text{w.r.t. DAG/TAI} \\
\text{Server Error} &: E_i = C_i + \mathbf{1}_{t_i > t^*} \qquad \text{w.r.t. UTC}
\end{aligned}$$

Good Server

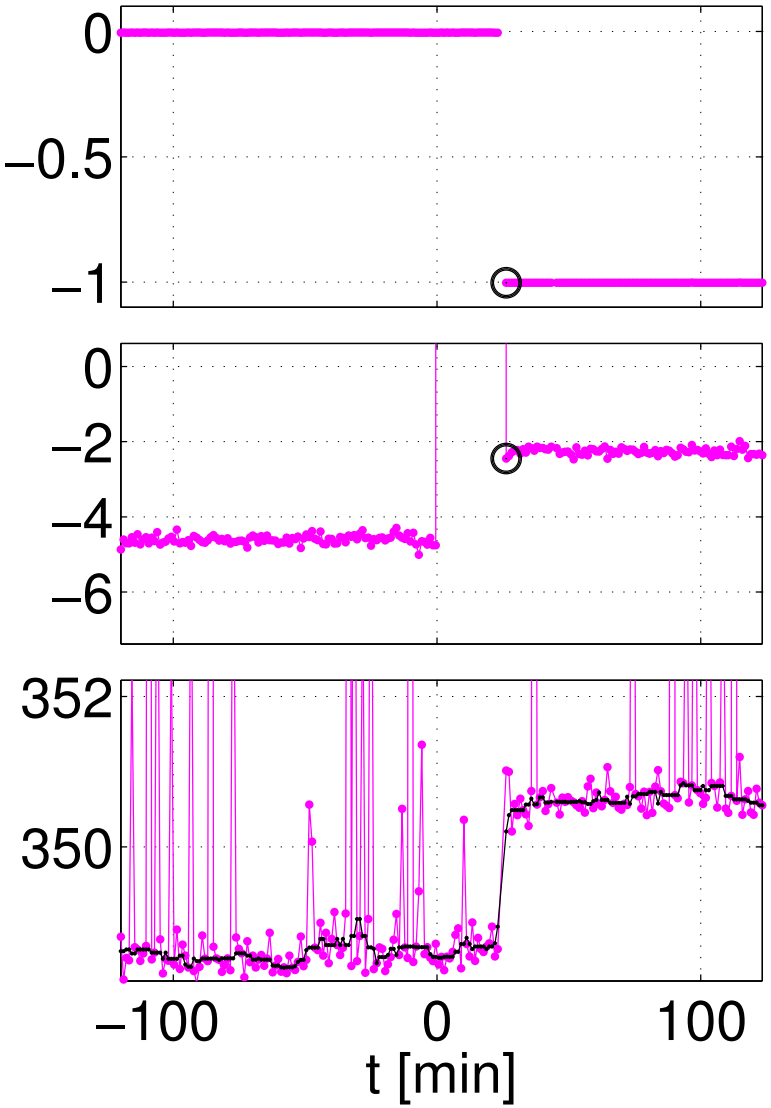


GOOD SERVER



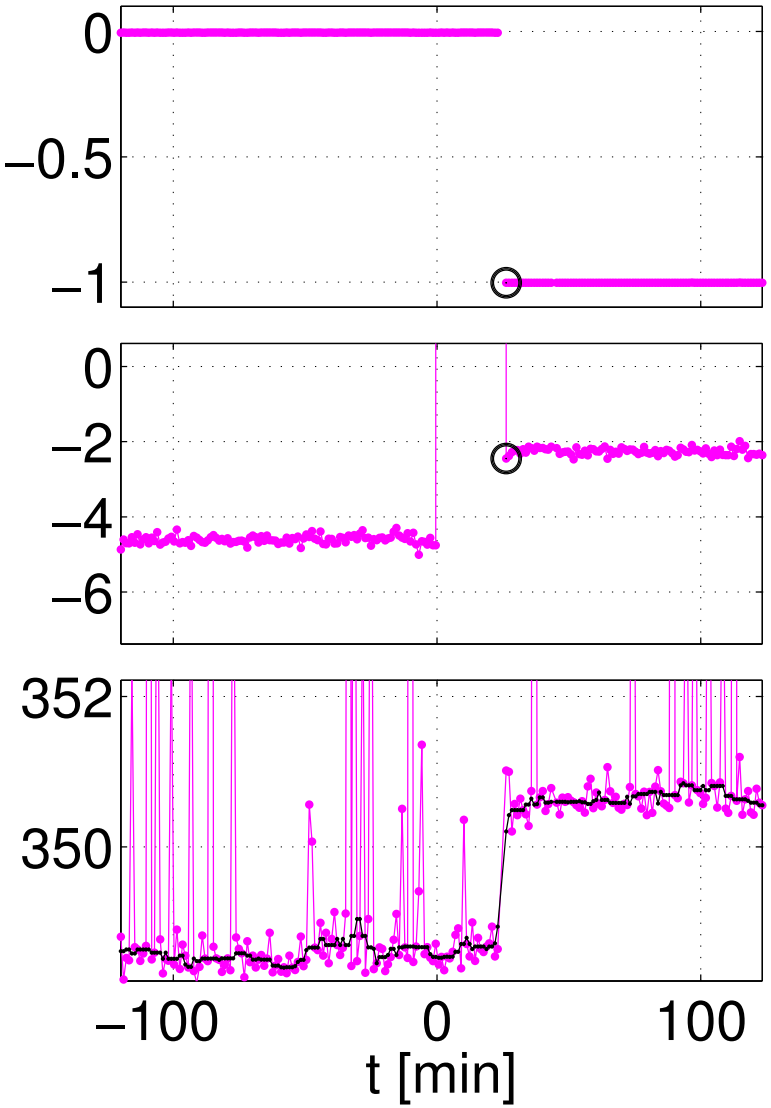
Good Server : *no evidence of error*, clean jump at first sample after leap

Not Good but Clean



$\mathrm{LED} = 26.2\,\mathrm{min}$

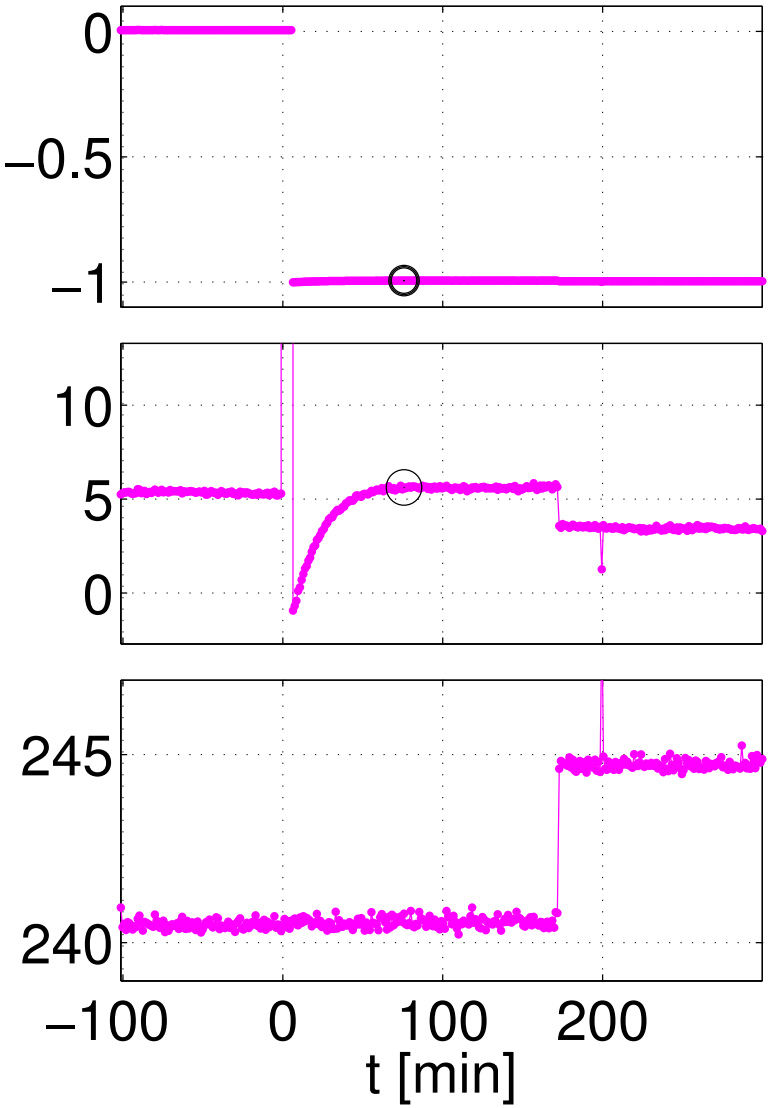
NOT GOOD BUT CLEAN



LED = 26.2 min

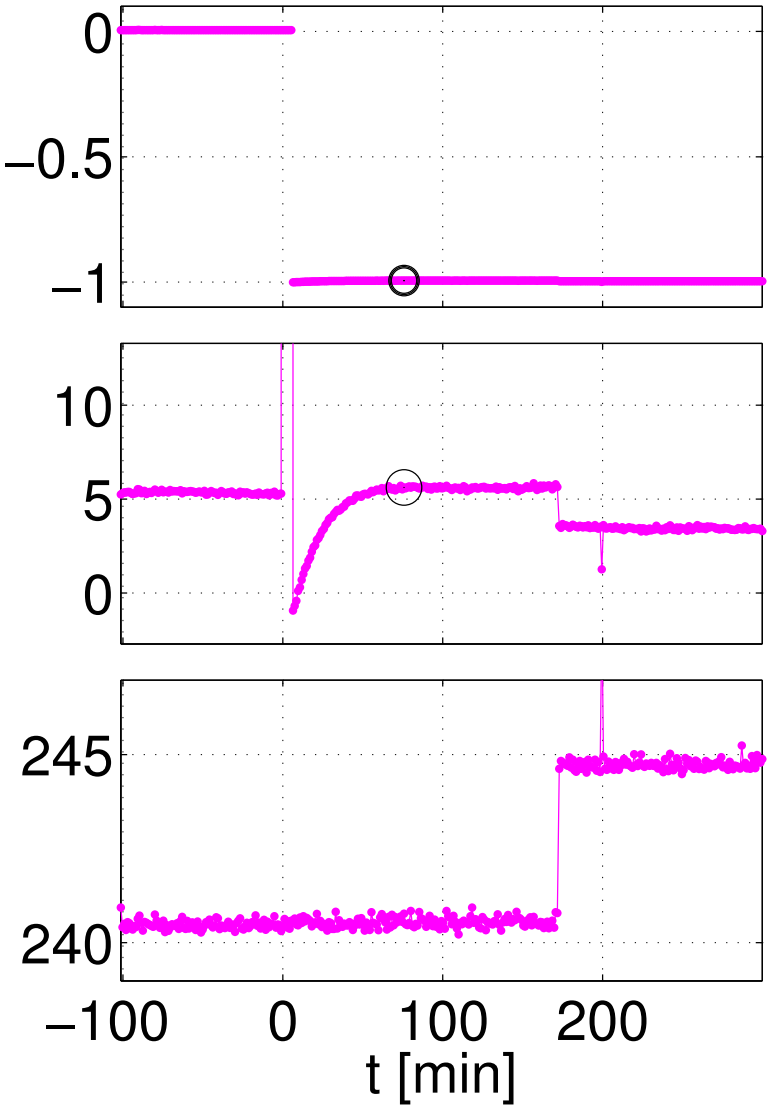
Not Good but Clean : ToJ = TEB (delayed jump)

NOT CLEAN



$$\mathrm{LED} = 75.8\,\mathrm{min}$$

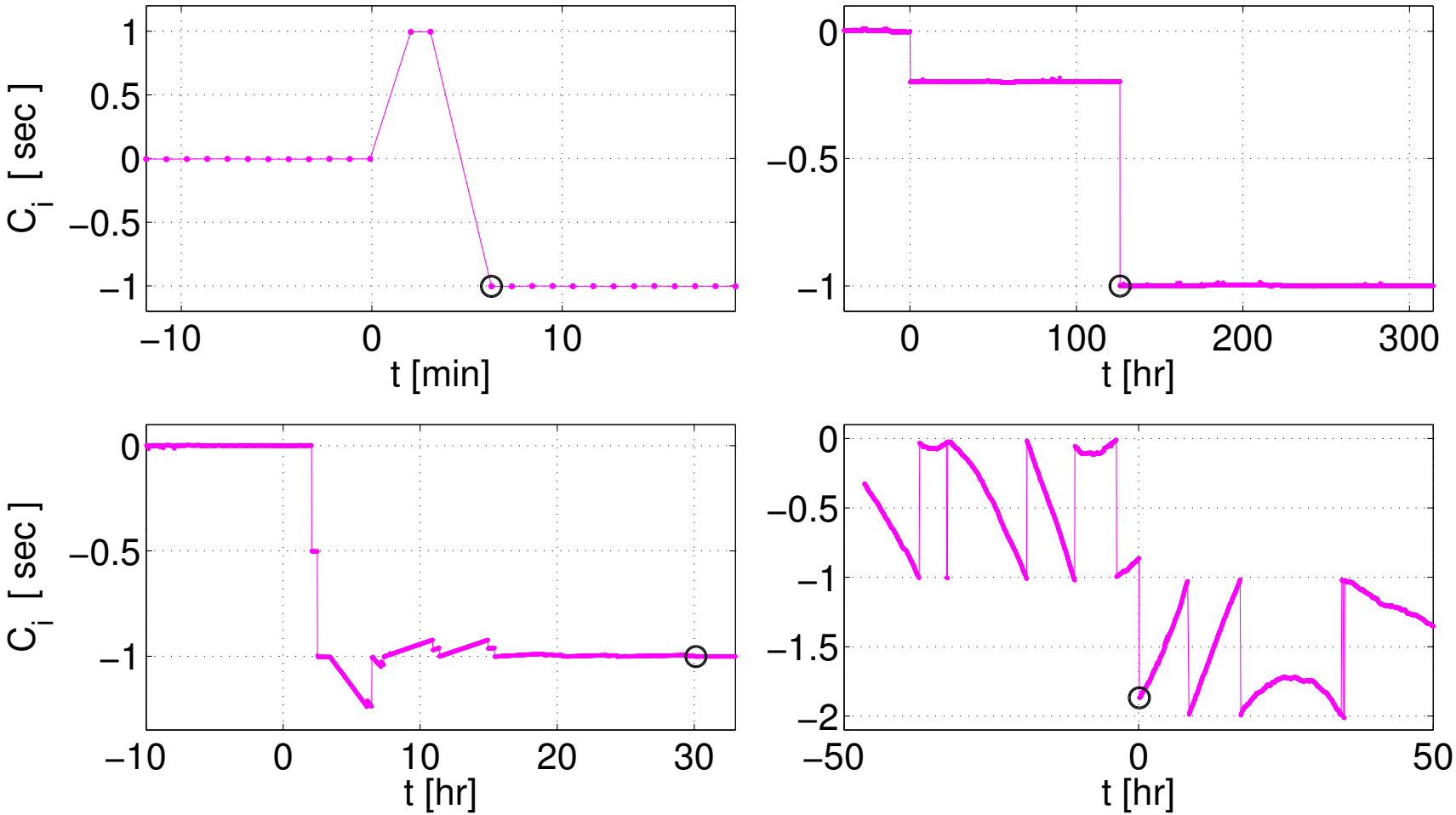
NOT CLEAN



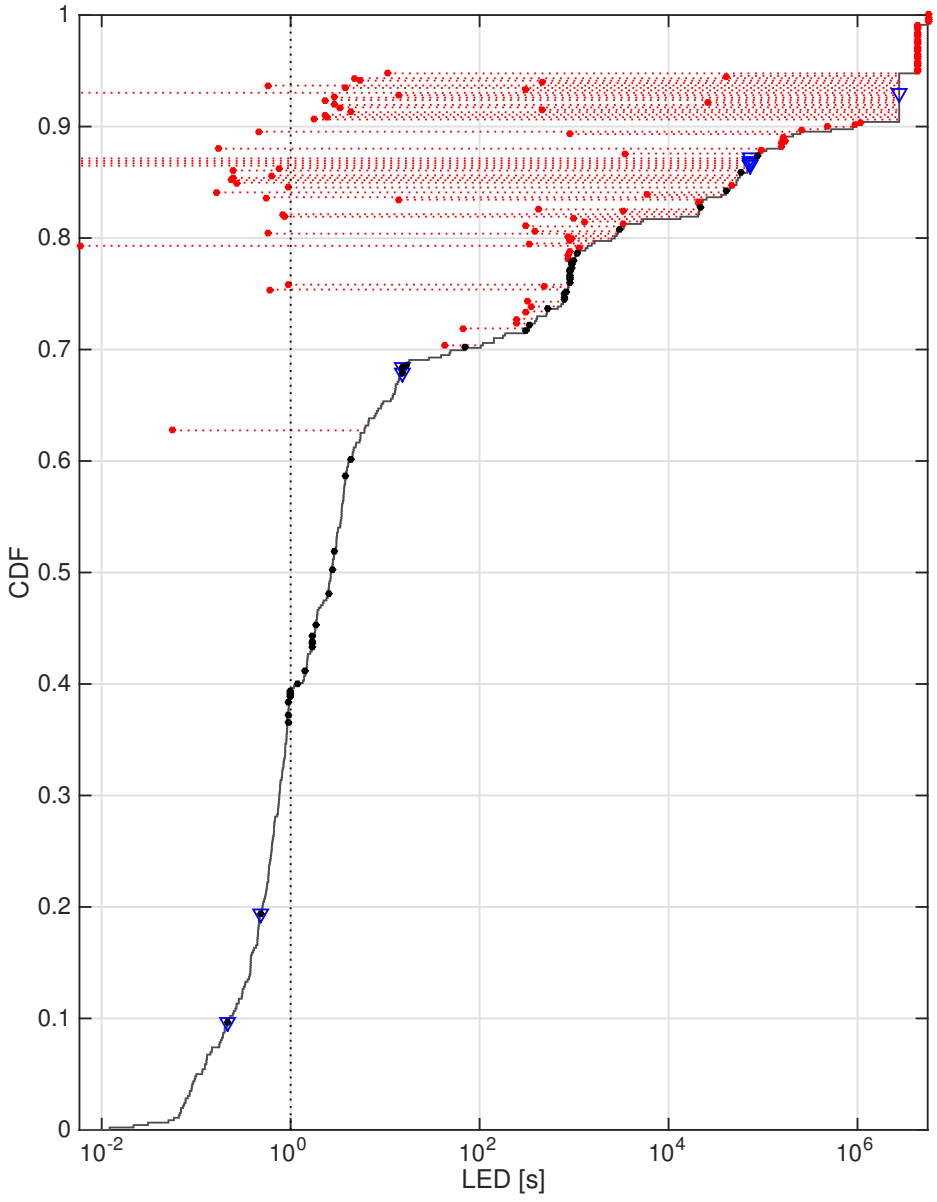
LED = 75.8 min

Not Clean : ToJ < TEB (aftershock after jump)

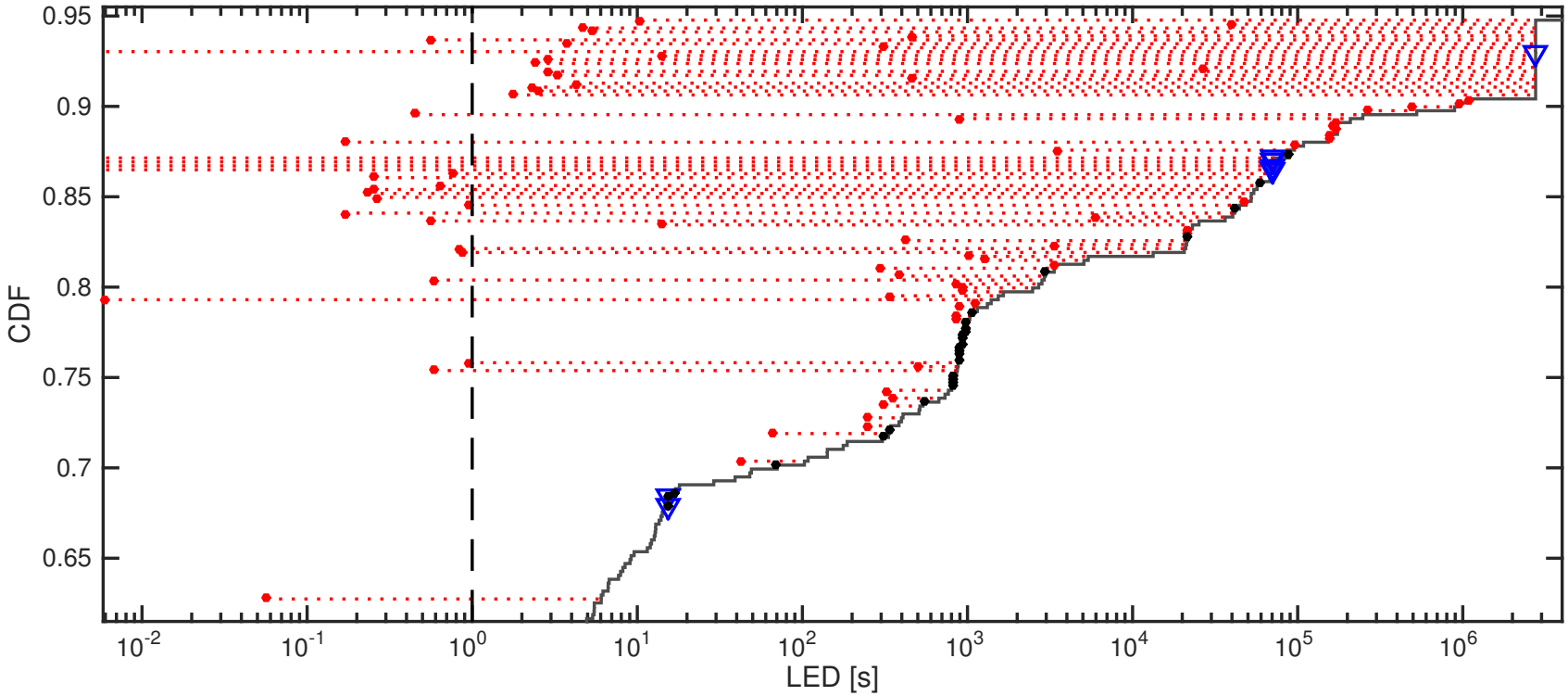
Extreme Servers



Results: Leap Error Duration (LED), and ToJ



LED Spread



Not Good but Clean (ToJ = TEB)

Not Clean (ToJ < TEB)

Preleap (ToJ < 0)

Protocol Compliance

LI values :

00 : default, no leap coming

01 : positive leap coming up

10 : negative leap coming up [never seen in our data]

11 : host clock is unsynchronized [not leap specific !]

Protocol Compliance

LI values :

00 : default, no leap coming

01 : positive leap coming up

10 : negative leap coming up [never seen in our data]

11 : host clock is unsynchronized [not leap specific !]

'Ok' behaviour is to issue LI = 01 warning :

I Starting within 24hrs of the Leap

II In each packet once they start

III Starting early enough so warnings get through

Protocol Compliance

LI values :

00 : default, no leap coming

01 : positive leap coming up

10 : negative leap coming up [never seen in our data]

11 : host clock is unsynchronized [not leap specific !]

'Ok' behaviour is to issue LI = 01 warning :

I Starting within 24hrs of the Leap

II In each packet once they start

III Starting early enough so warnings get through

[i] + [ii] = Standards Compliant (RFC 5904)

Protocol Compliance

LI values :

00 : default, no leap coming

01 : positive leap coming up

10 : negative leap coming up [never seen in our data]

11 : host clock is unsynchronized [not leap specific !]

'Ok' behaviour is to issue LI = 01 warning :

i Starting within 24hrs of the Leap

ii In each packet once they start

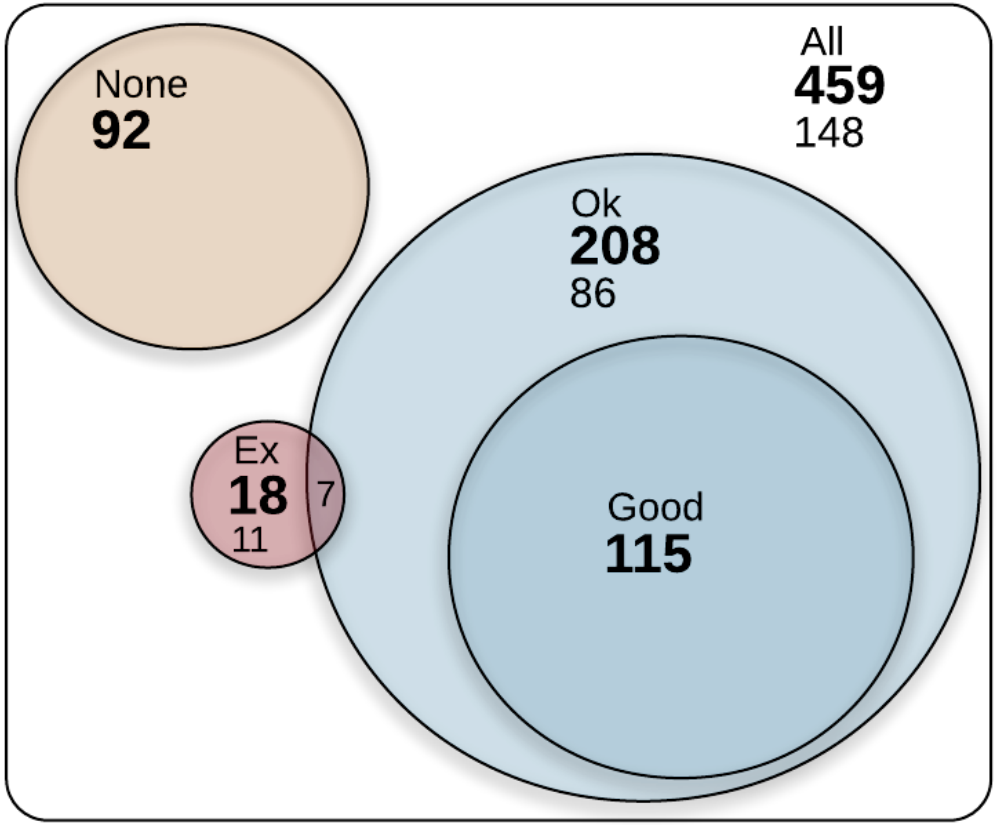
iii Starting early enough so warnings get through

[i] + [ii] = Standards Compliant (RFC 5904)

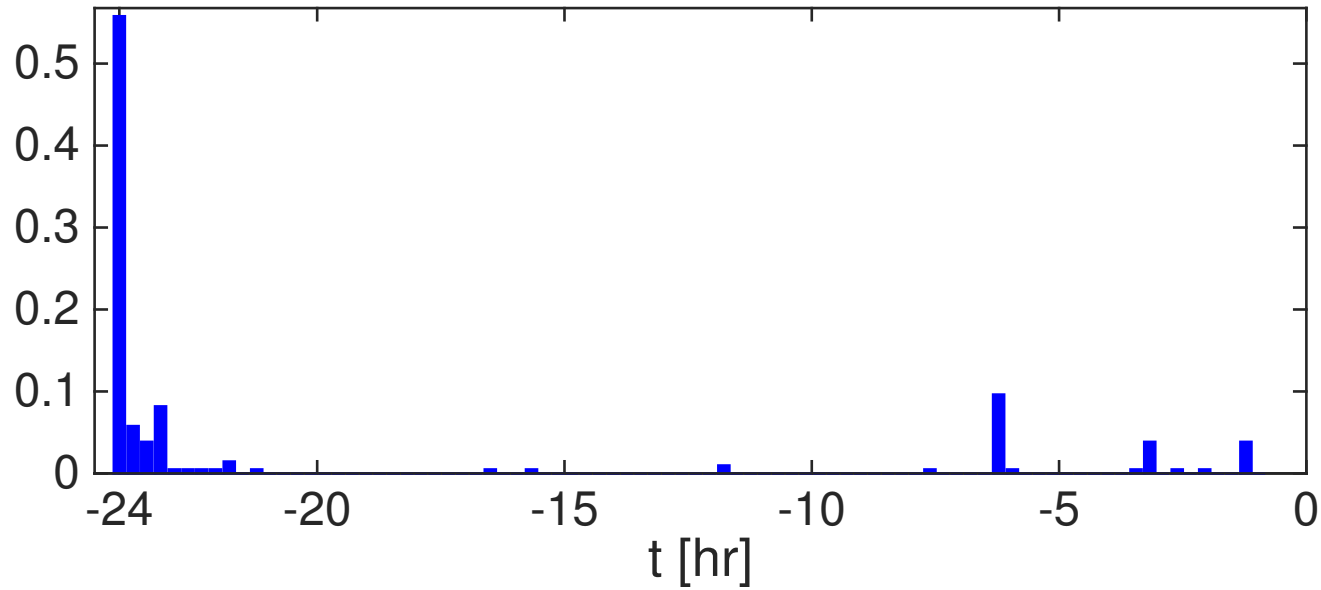
'Good' behaviour is 'Ok' +

- Stopping after the Leap!!

LI Warning Summary



Warning Starting Times (Ok Servers)



Overall Compliance

Ideal Clock-Good & LI01-Good & LI11-None & St-Constant(S1)

Adequate Clock-Good & LI01-Ok \ St-Constant(not S1)

Overall Compliance

Ideal Clock-Good & LI01-Good & LI11-None & St-Constant(S1)

Adequate Clock-Good & LI01-Ok \ St-Constant(not S1)

Server Set	#	Ideal	Adequate
ListLeap	459	**36 (7.8)%**	**171 (37.3)%**
Pool	258	19 (7.4)%	97 (37.6)%
Org	197	11 (5.6)%	69 (35.0)%
NL	66	10 (15.2)%	**32 (48.5)%**
ListLeap\NL	393	26 (6.6)%	**139 (35.4)**%

THE GOOD GUYS

CONT	URL	CY	Strata		Server Error			Avail. (%)	Leap Perf
			List	S1Down time(%)	Class	Size [ms]	Errtime (%)		
AF	stratum1.neology.co.za	ZA	{1}	0	R	2.1	7.0e-5	99.87	Adeq.
AN	–	–	–	–	–	–	–	–	–
OC	ntp1.net.monash.edu.au	AU	{1}	0	R	180	1.4e-4	99.86	Adeq.
EU	ntp1.oma.be	BE	{0,1}	2.9e-4	R	28	0.032	99.04	Adeq.
	ntp.freestone.net	CH	{1}	0	G	–	–	99.80	Ideal
	netopyr.hanacke.net	CZ	{1}	0	G	–	–	99.25	Ideal
	ntp.nic.cz	CZ	{1}	0	G	–	–	99.86	Adeq.
	hora.roa.es	ES	{0,1,2}	2.9e-4	R	120	5.8e-3	99.40	Adeq.
	ntp.i2t.ehu.es	ES	{1}	0	G	–	–	98.94	Ideal
	unknown1	GB	{1}	0	G	–	–	99.71	Ideal
	unknown2	GB	{1}	0	G	–	–	99.71	Ideal
	ntp2.litnet.lt	LT	{1}	0	G	–	–	99.87	Ideal
	chime2.surfnet.nl	NL	{1,2}	3.3e-4	G	–	–	99.80	Adeq.
	metronoom.dmz.cs.uu.nl	NL	{1}	0	G	–	–	99.66	Ideal
	unknown3	NO	{1}	0	G	–	–	98.88	Ideal
	goblin.nask.net.pl	PL	{1}	0	G	–	–	99.79	Ideal
	ntp.certum.pl	PL	{1}	0	R	7.0	0.025	97.55	Adeq.
	ntp.fizyka.umk.pl	PL	{1}	0	G	–	–	99.45	Ideal
	ntp2.tp.pl	PL	{0,1,2}	4.8e-4	G			97.73	Adeq.

How to Make it Stop?

Three Options

- Fix the components and the system
- Ond-off careful infrastructure
- Monitor and Verify what you have

How to Make it Stop?

Three Options

- Fix the components and the system
- Ond-off careful infrastructure
- Monitor and Verify what you have
  – Timing Verification as a Service – see next talk !