Applying Stable distribution on congestion latency signatures

AIMS 2018: Workshop on Active Internet Measurements
March 13th, 2018
San Diego Supercomputer Center
University of California San Diego

Esteban Carisimo, J Ignacio Alvarez-Hamelin and Amogh Dhamdhere

UBA
Universidad de Buenos Aires

CONICET

Modeling latency distribution

- **Previous models**
  - Weibull [Papagiannaki,Hernández]
  - Lognormal [Fontugne]
- **Problems**
  - Different acquisition method
  - Did not capture well our data
  - Couldn’t reproduce *extreme values*
- **Proposal**
  - Stable distribution
  - Fairly popular in Econophysics

Papagiannaki, Konstantina, et al. "Measurement and analysis of single-hop delay on an IP backbone network." *IEEE Journal on Selected Areas in Communications* 21.6 (2003): 908-921.
Hernández, José-Alberto, and Iain W. Phillips. "Weibull mixture model to characterise end-to-end Internet delay at coarse time-scales." *IEE*
Fontugne, Romain, Johan Mazel, and Kensuke Fukuda. "An empirical mixture model for large-scale RTT measurements." *Computer Communications (INFOCOM), 2015 IEEE Conference on*. IEEE, 2015. *Proceedings-Communications* 153.2 (2006): 295-304.

The Stable distribution

$$g(k) = \exp\{\delta[ik\gamma - |k|^{\alpha} w(k;\alpha,\beta)]\}$$

$$w(k;\alpha,\beta) = \begin{cases} exp[-i\beta\Phi(\alpha)\mathrm{sign}(k)], & \alpha \neq 1 \\ \pi/2 + i\beta \log|k|\mathrm{sign}(k), & \alpha = 1 \end{cases}$$

$$\Phi(\alpha) = \begin{cases} \alpha\pi/2 & \alpha < 1 \\ (\alpha-2)\pi/2 & \alpha > 1 \end{cases}$$

- g(k) is the characteristic function
- It is defined by four parameters
- a.k.a “paretian stable”, “levy stable”, “alpha-stable"

Formula as in page xvi. Uchaikin, Vladimir V., and Vladimir M. Zolotarev. Chance and stability: stable distributions and their applications. Walter de Gruyter, 1999.

CONICET

The Stable distribution

$$g(k) = \exp\{\delta[ik\gamma - |k|^{\alpha} w(k;\alpha,\beta)]\}$$

$$w(k;\alpha,\beta) = \begin{cases} exp[-i\beta\Phi(\alpha)\text{sign}(k)], & \alpha \neq 1 \\ \pi/2 + i\beta \log|k|\text{sign}(k), & \alpha = 1 \end{cases}$$

$$\Phi(\alpha) = \begin{cases} \alpha\pi/2 & \alpha < 1 \\ (\alpha-2)\pi/2 & \alpha > 1 \end{cases}$$

Drawbacks

- Fairly complex expression
- It cannot be expressed in terms of elementary functions.
- Extremely hard to fit Stable to data [McCulloch1986]

Formula as in page xvi. Uchaikin, Vladimir V., and Vladimir M. Zolotarev. Chance and stability: stable distributions and their applications. Walter de Gruyter, 1999.

[McCulloch1986] McCulloch, J. Huston. "Simple consistent estimators of stable distribution parameters." *Communications in Statistics-Simulation and Computation* 15.4 (1986): 1109-1136.

CONICET

3 important things about the Stable

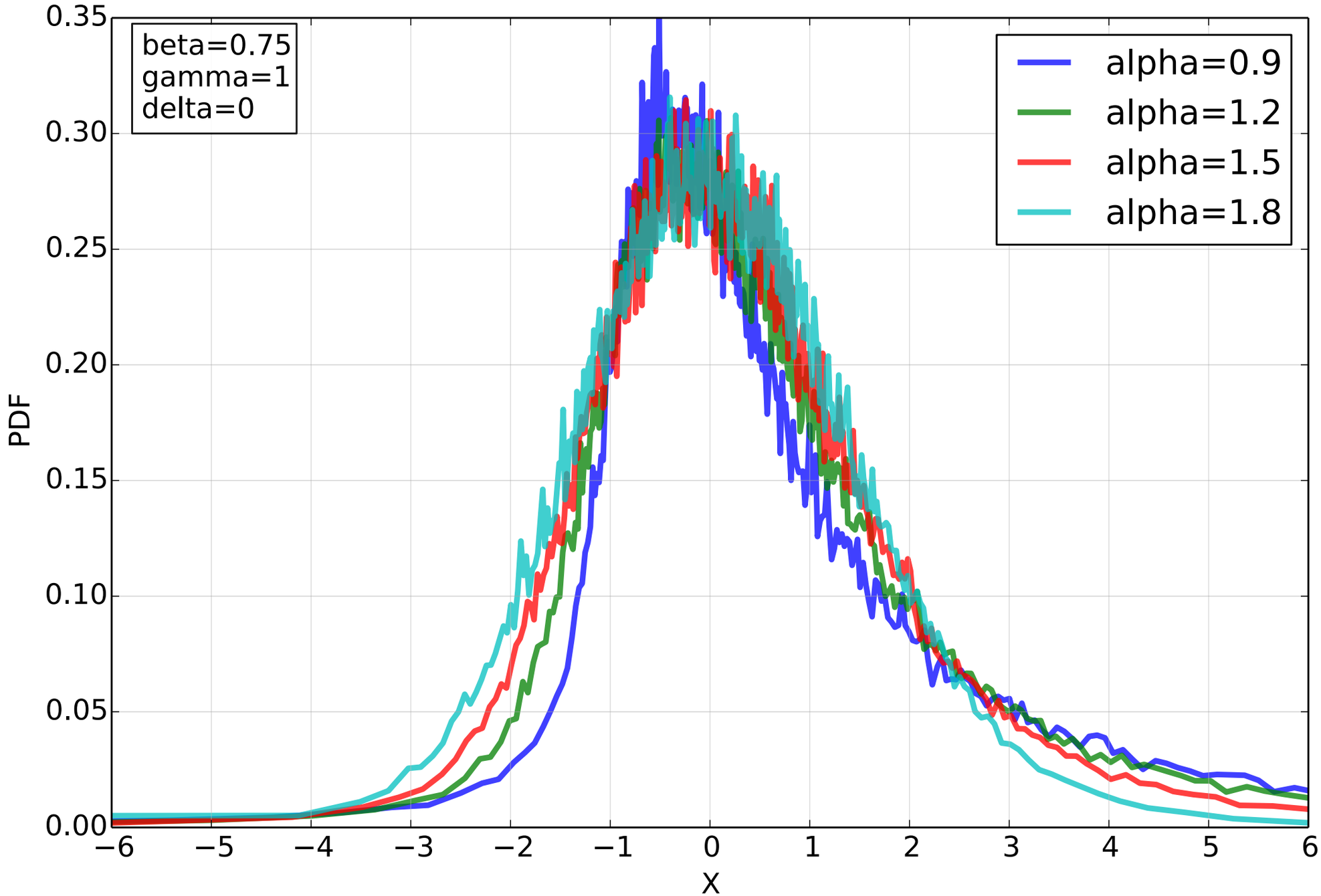
1. It is four-parameter distribution
2. Some parametrization can yield a heavy tail distribution
3. Normal distribution belongs to the Stable distribution family

Parameters of the Stable distribution

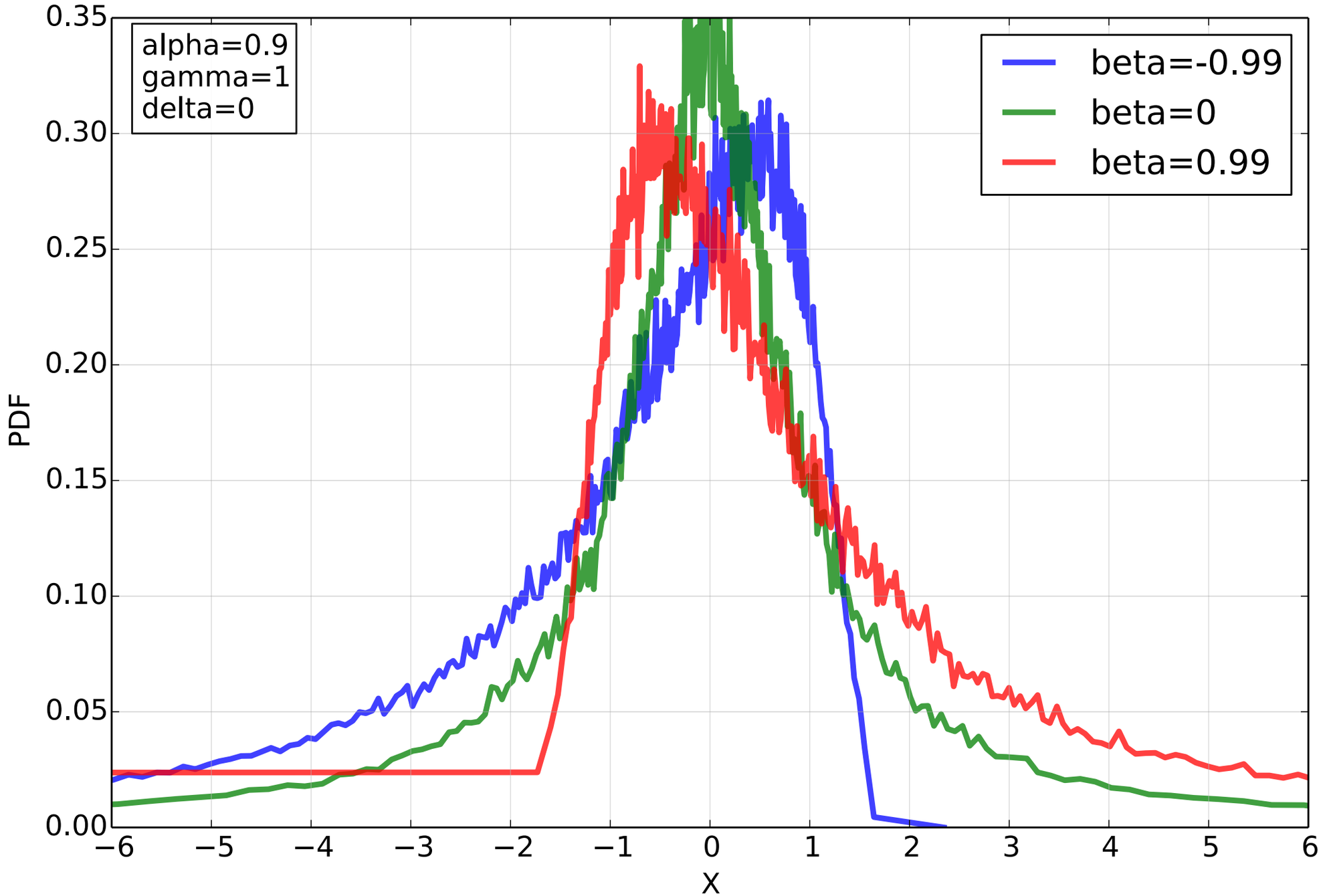
- $\alpha \in (0, 2]$: Characteristic parameter. Defines the decrease of the tail
- $\beta \in (-1, 1)$: Skew parameter
- $\gamma \in \mathbb{R}$: Scale or stretching parameter
- $\delta \in \mathbb{R}$: Location parameter

Uchaikin, Vladimir V., and Vladimir M. Zolotarev. Chance and stability: stable distributions and their applications. Walter de Gruyter, 1999.

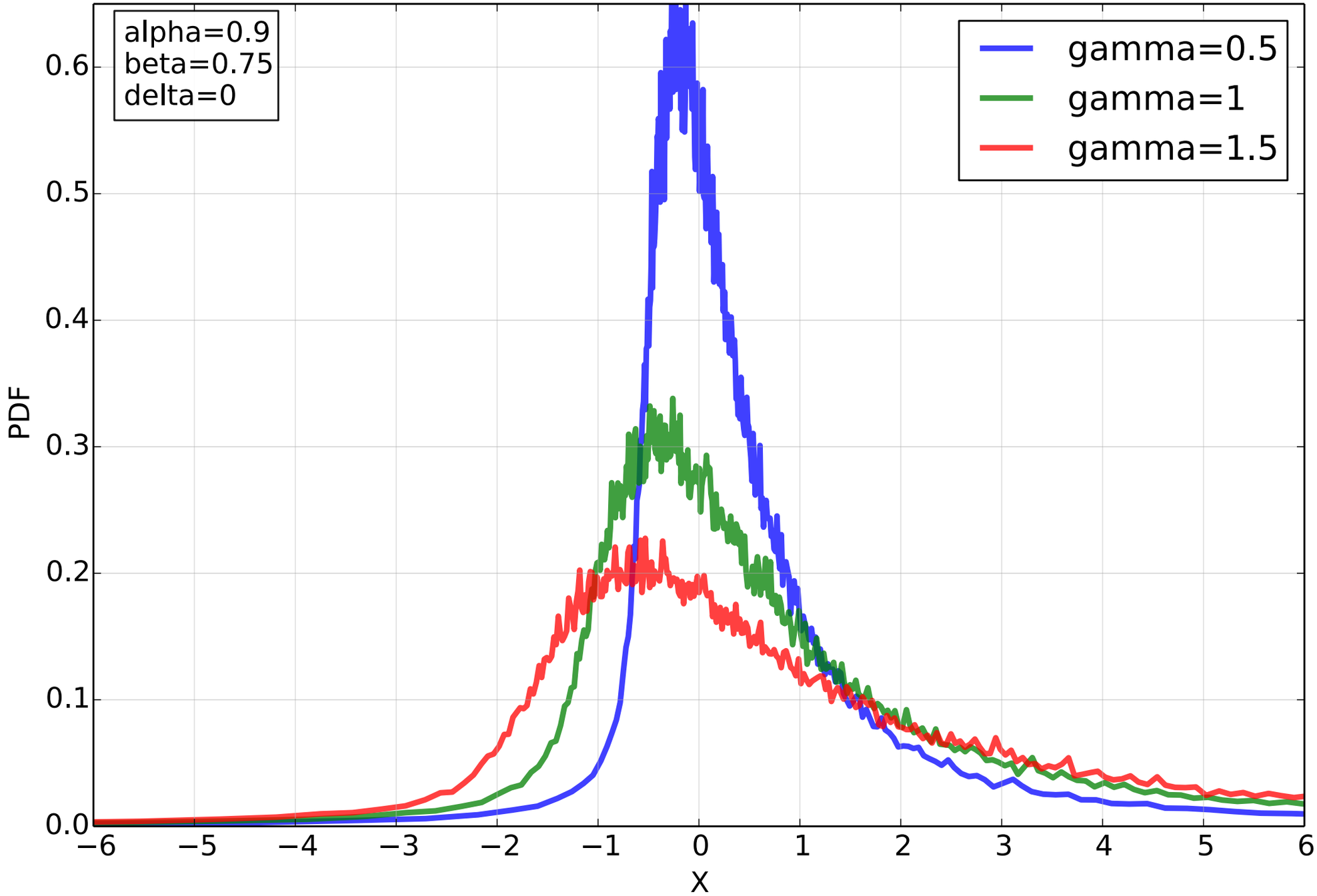
Parameters of Stable distribution



Parameters of Stable distribution

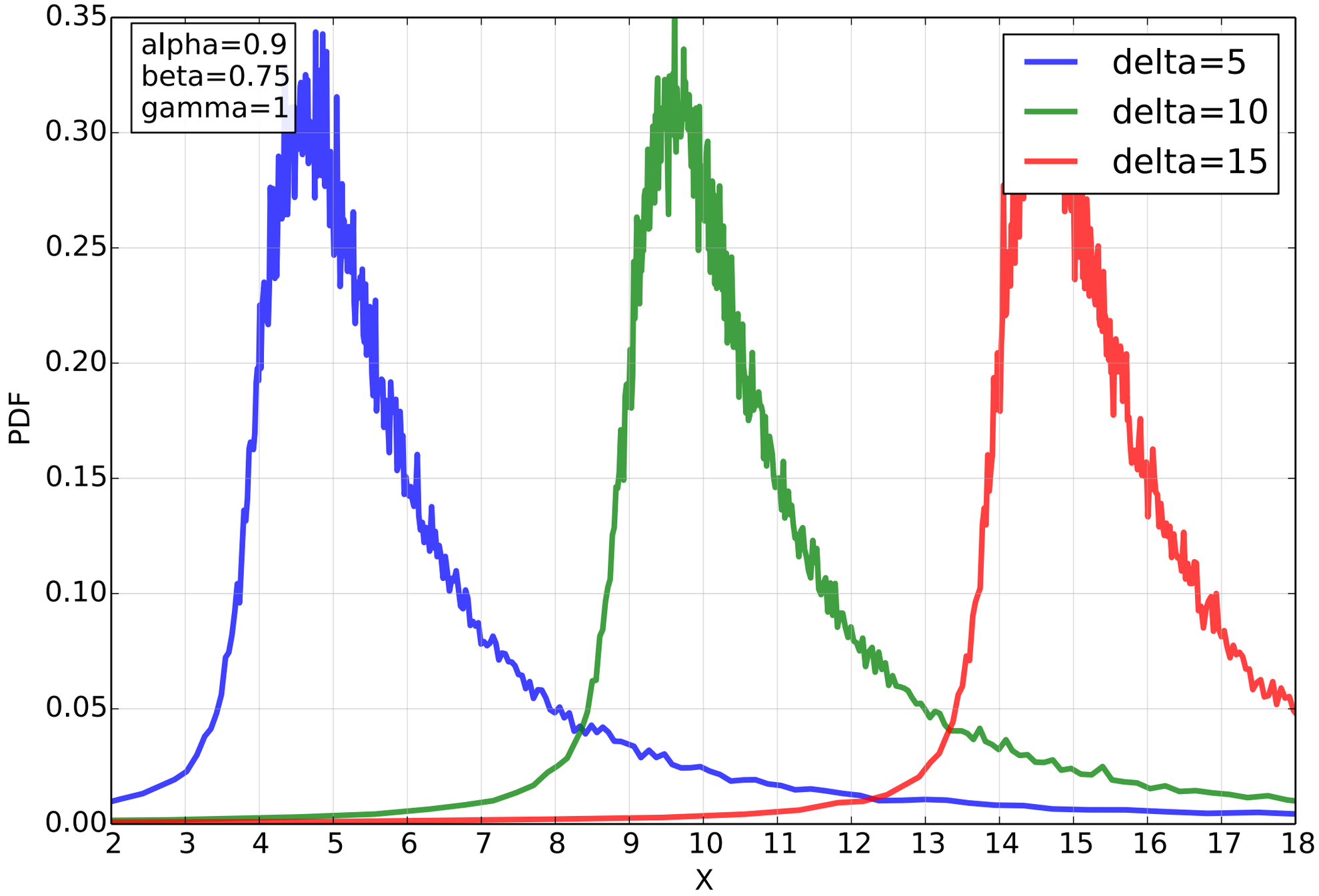


Parameters of Stable distribution



CONICET

Parameters of Stable distribution



Dataset & procedure

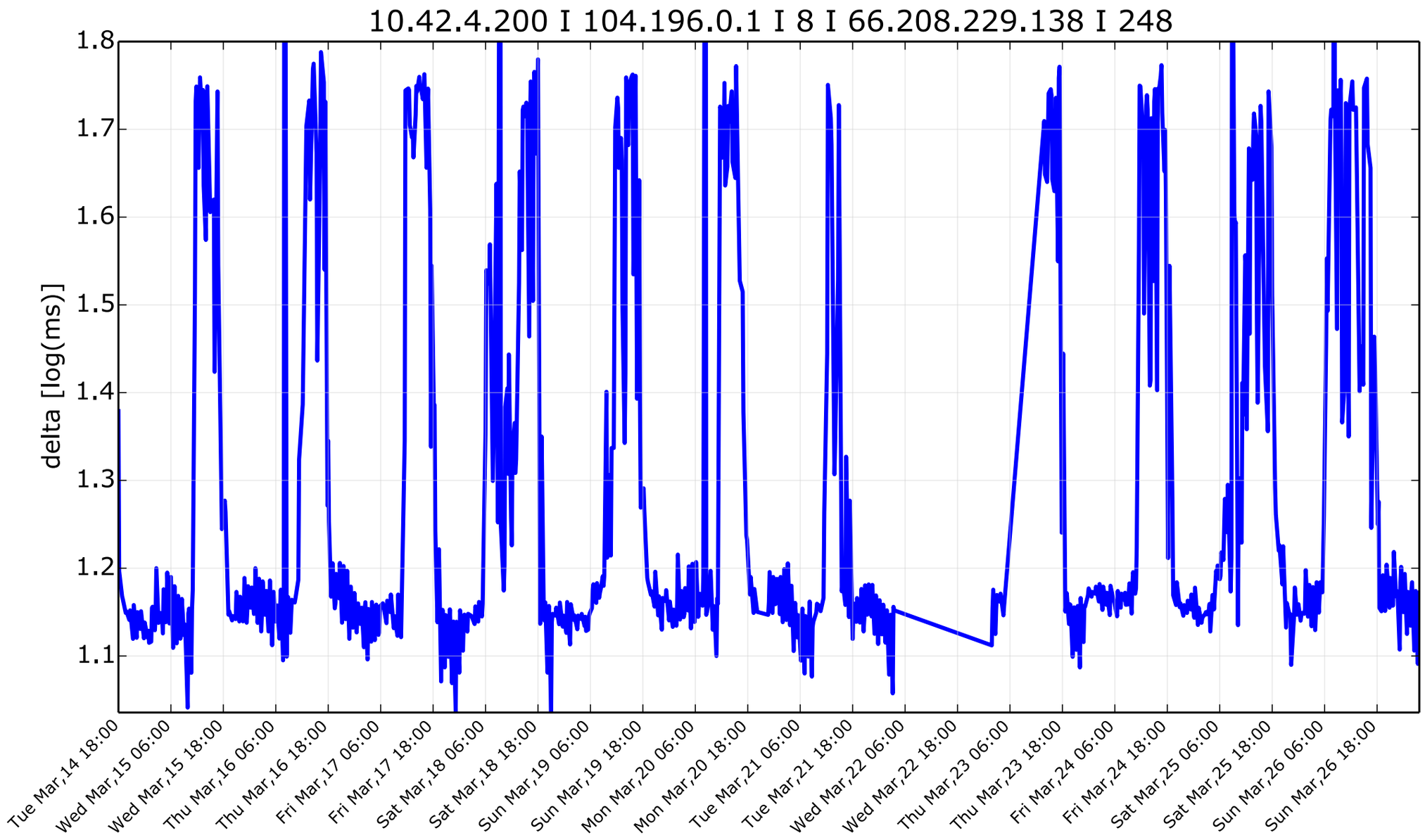
- Data acquisition
  - pings from Ark to neighbor ASes
  - RTT samples from interdomain links
  - High-frequency Probing
- Fitting Stable distribution to RTTs
  - Libstable (GPU) [Julian-Moreno16]
  - Time window: 10 minutes
  - Latency model: $\log(RTT) \sim S(\alpha, \beta, \gamma, \delta)$

[Julian-Moreno16] Julián-Moreno, Guillermo, et al. "Fast parallel $ $\alpha $ $ α-stable distribution function evaluation and parameter estimation using OpenCL in GPGPUs." *Statistics and Computing* 27.5 (2017): 1365-1382.

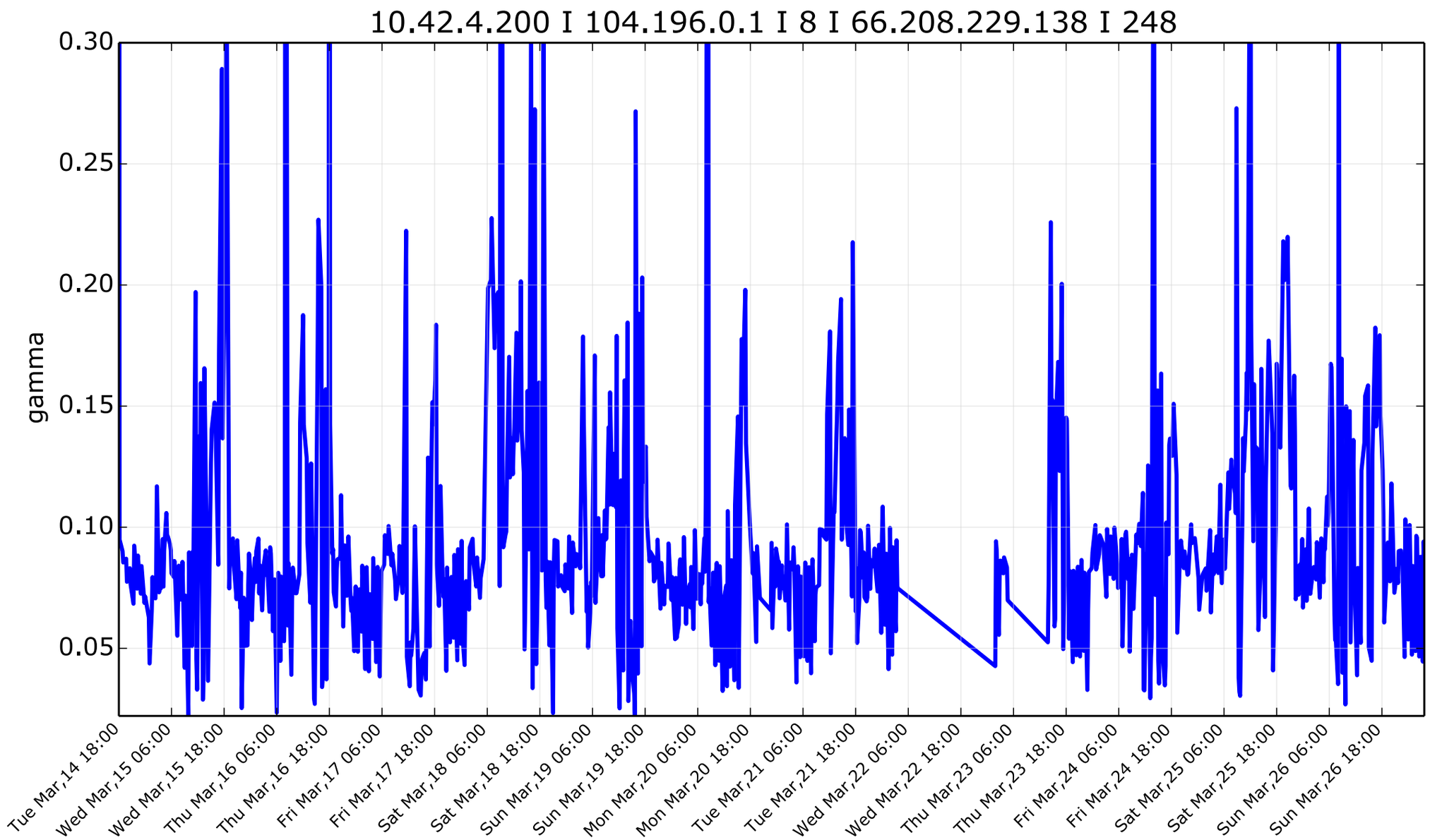
Summary in numbers

- 416M RTT samples
- 5 Ark monitors in 3 major ISPs
- 16 neighbor ASes
- 125 far IPs
- 1667 unique tuples (monitor,farIP,day)

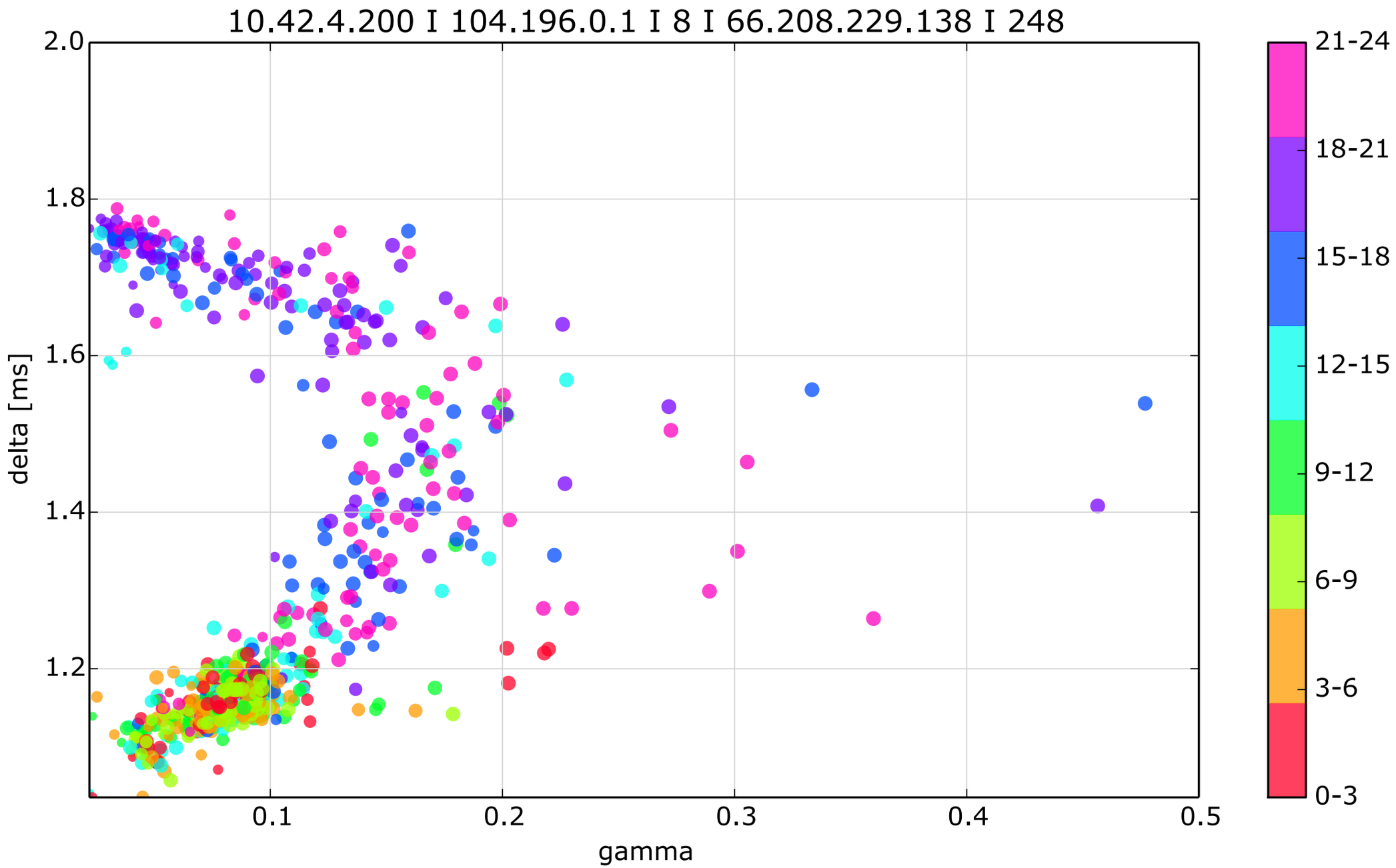
First look at DELTA throughout some days



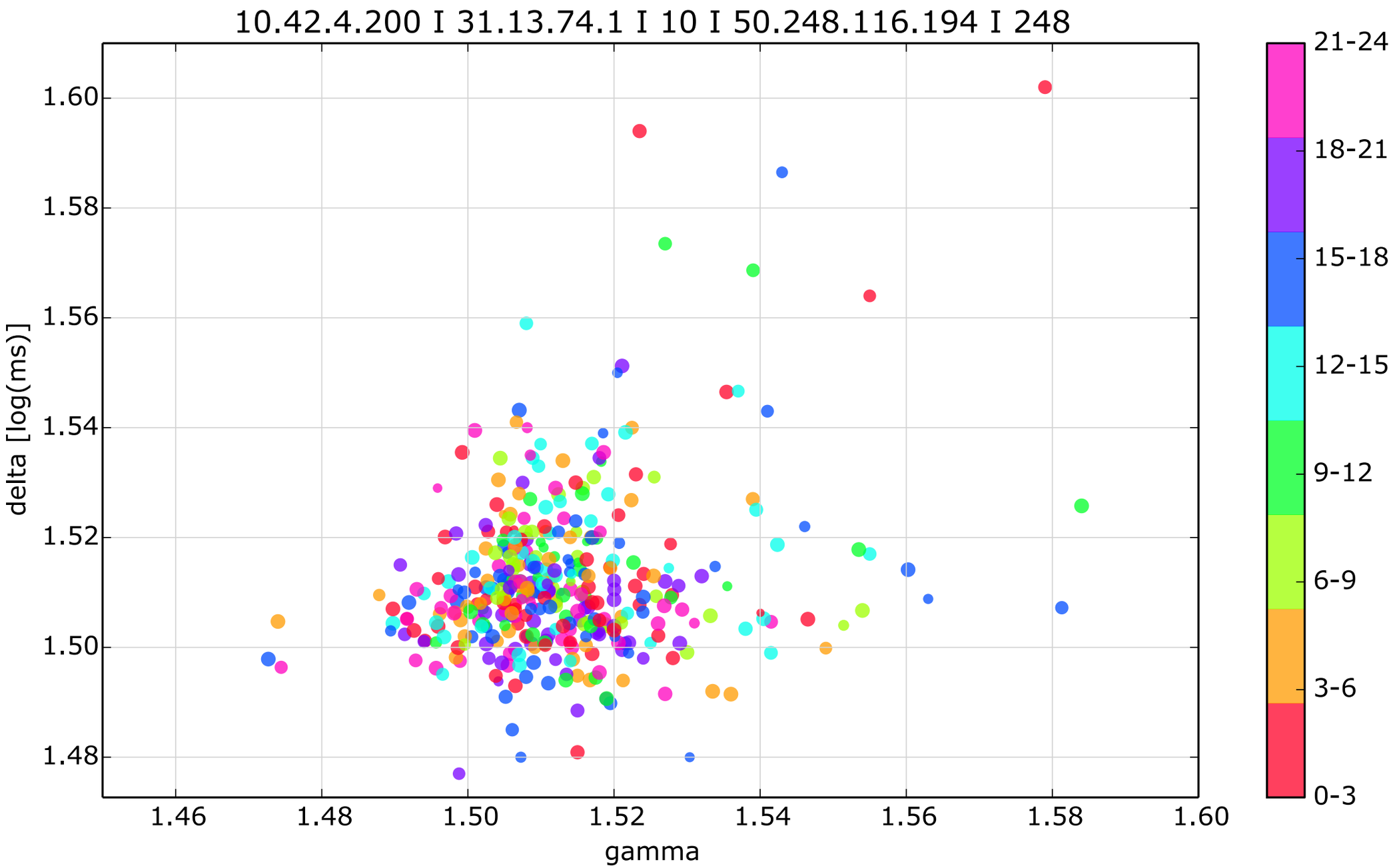
First look at GAMMA throughout some days



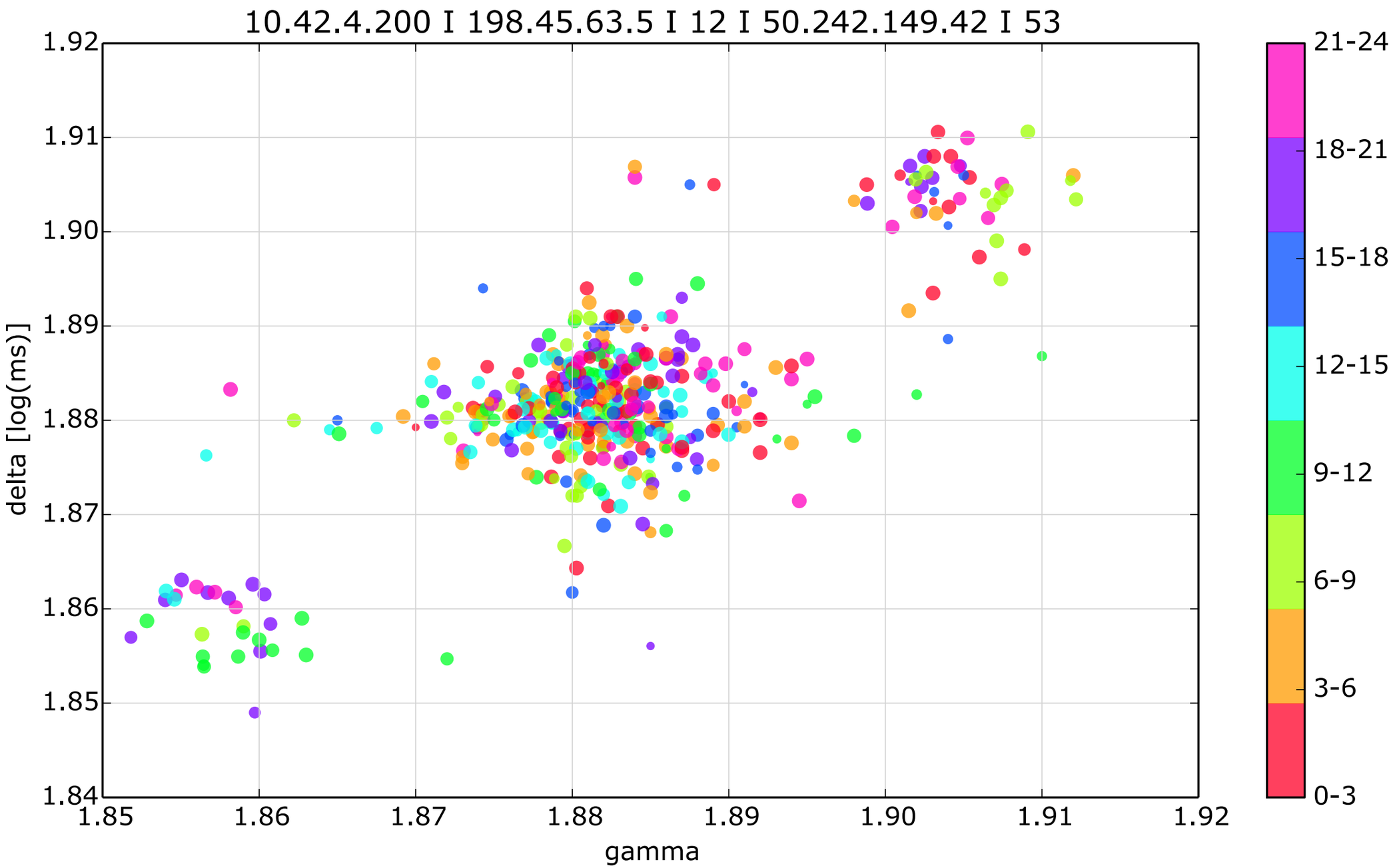
Scatter plot: DELTA and GAMMA altogether COMCAST (atl2-us)—> Google



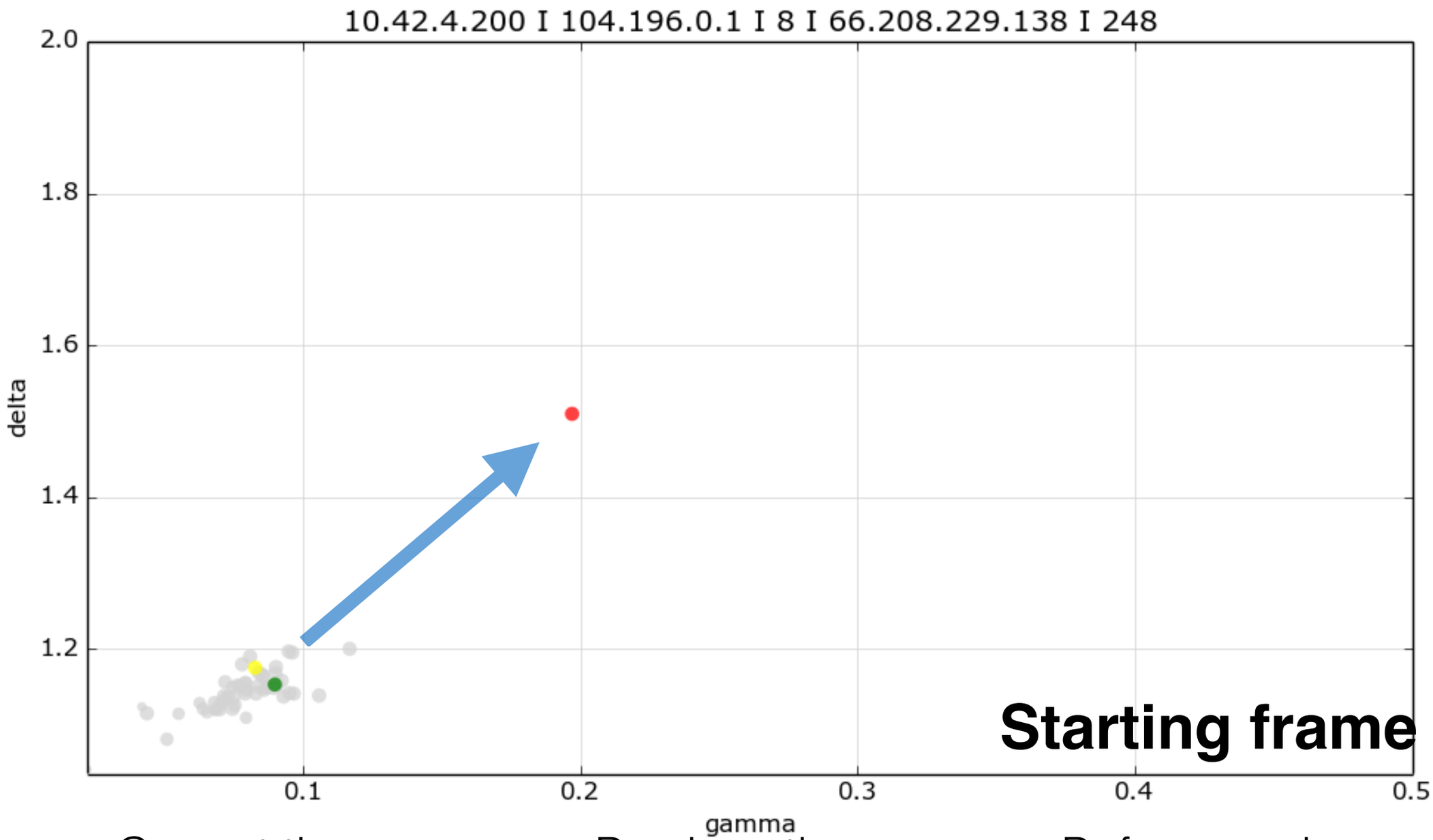
Scatter plot: DELTA and GAMMA altogether COMCAST (atl2-us)—> Facebook



Scatter plot: DELTA and GAMMA altogether COMCAST (atl2-us)—> Netflix

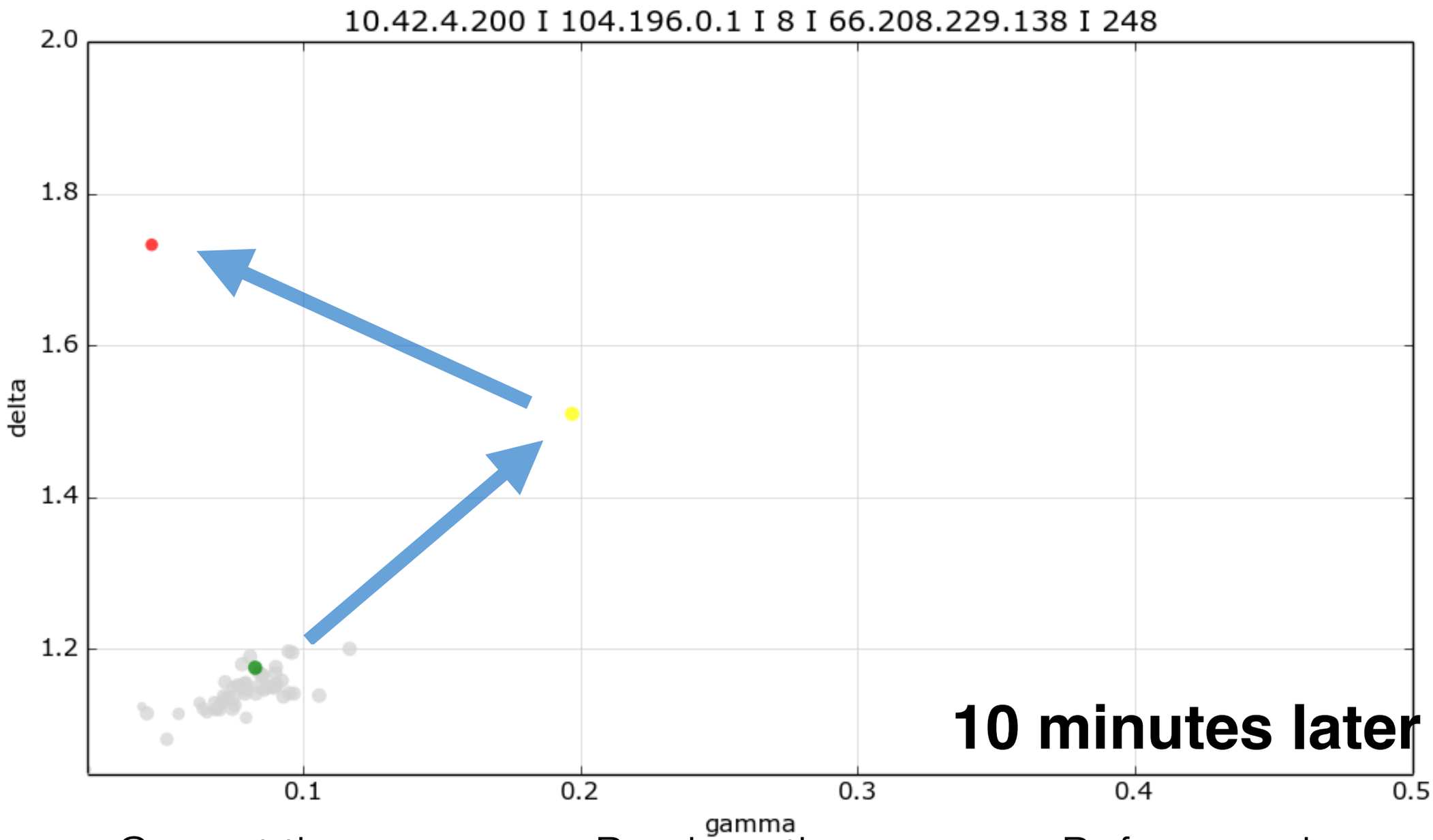


Scatter plot frame by frame



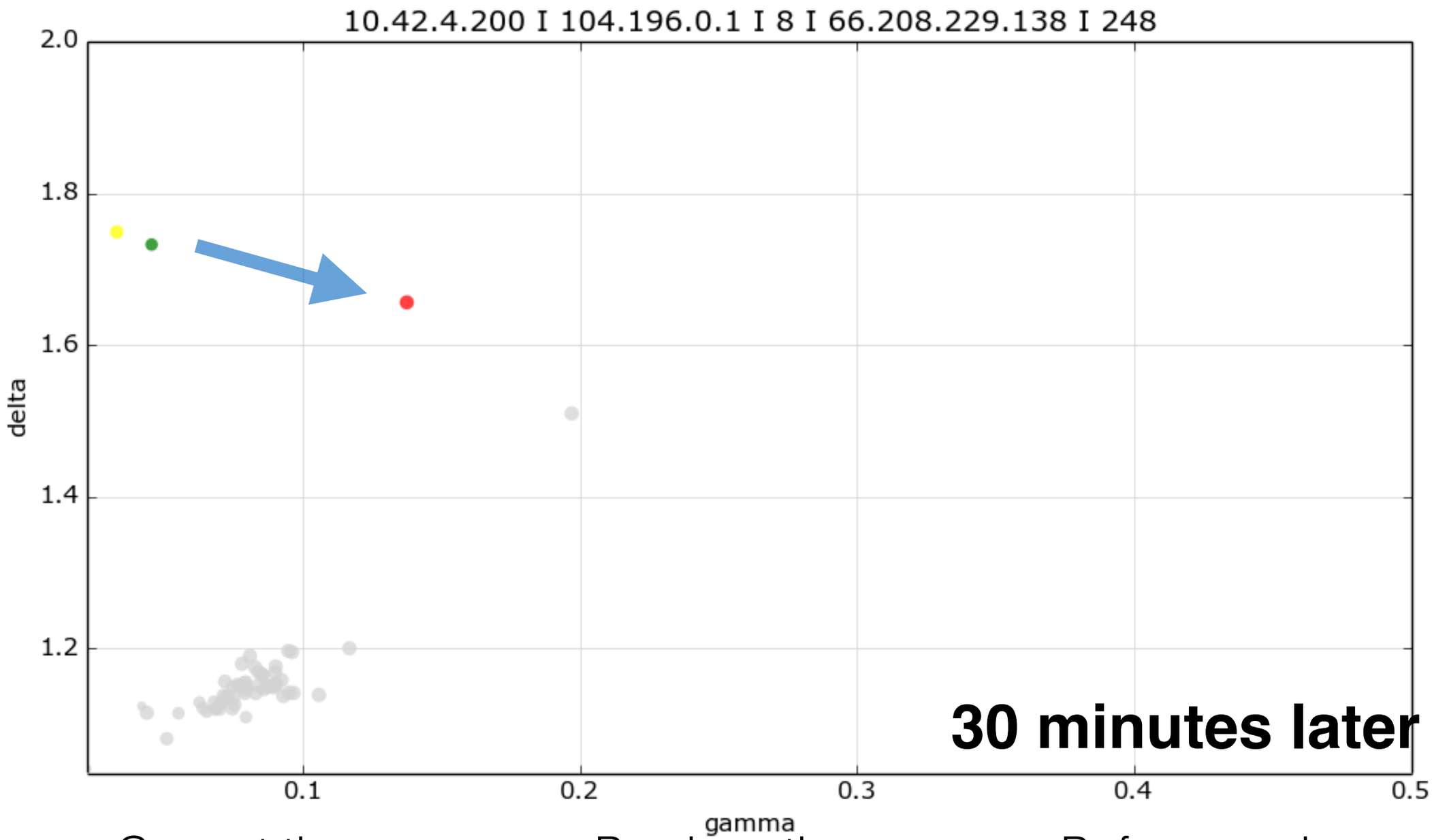
- Current time window
- Previous time window
- Before previous time window

Scatter plot frame by frame



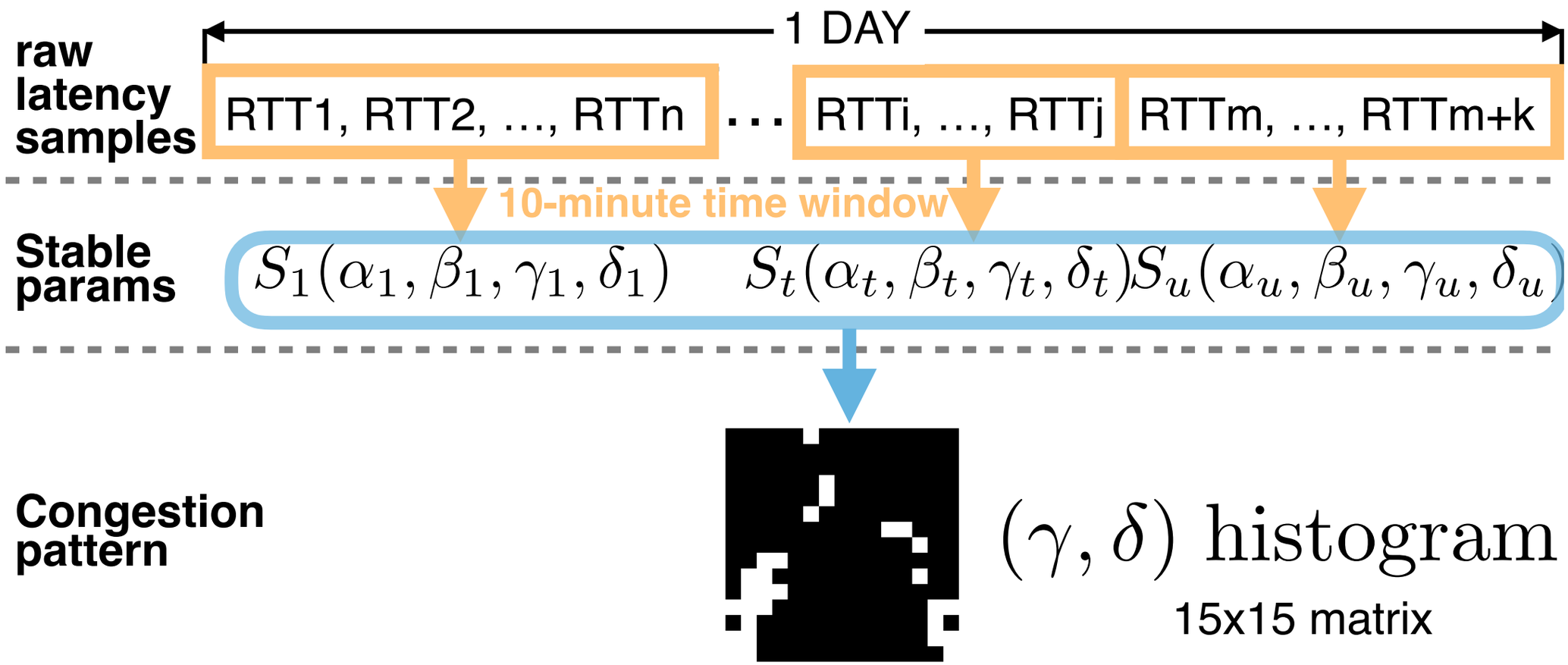
- Current time window
- Previous time window
- Before previous time window

Scatter plot frame by frame

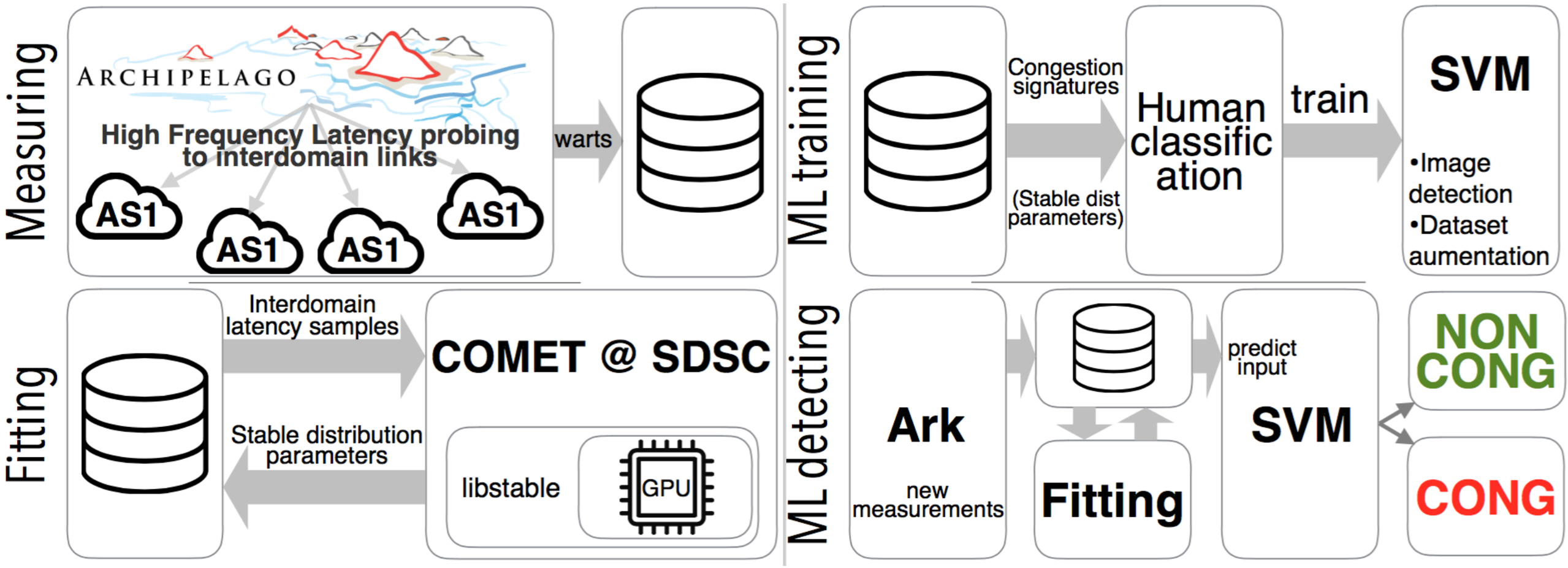


- Current time window
- Previous time window
- Before previous time window

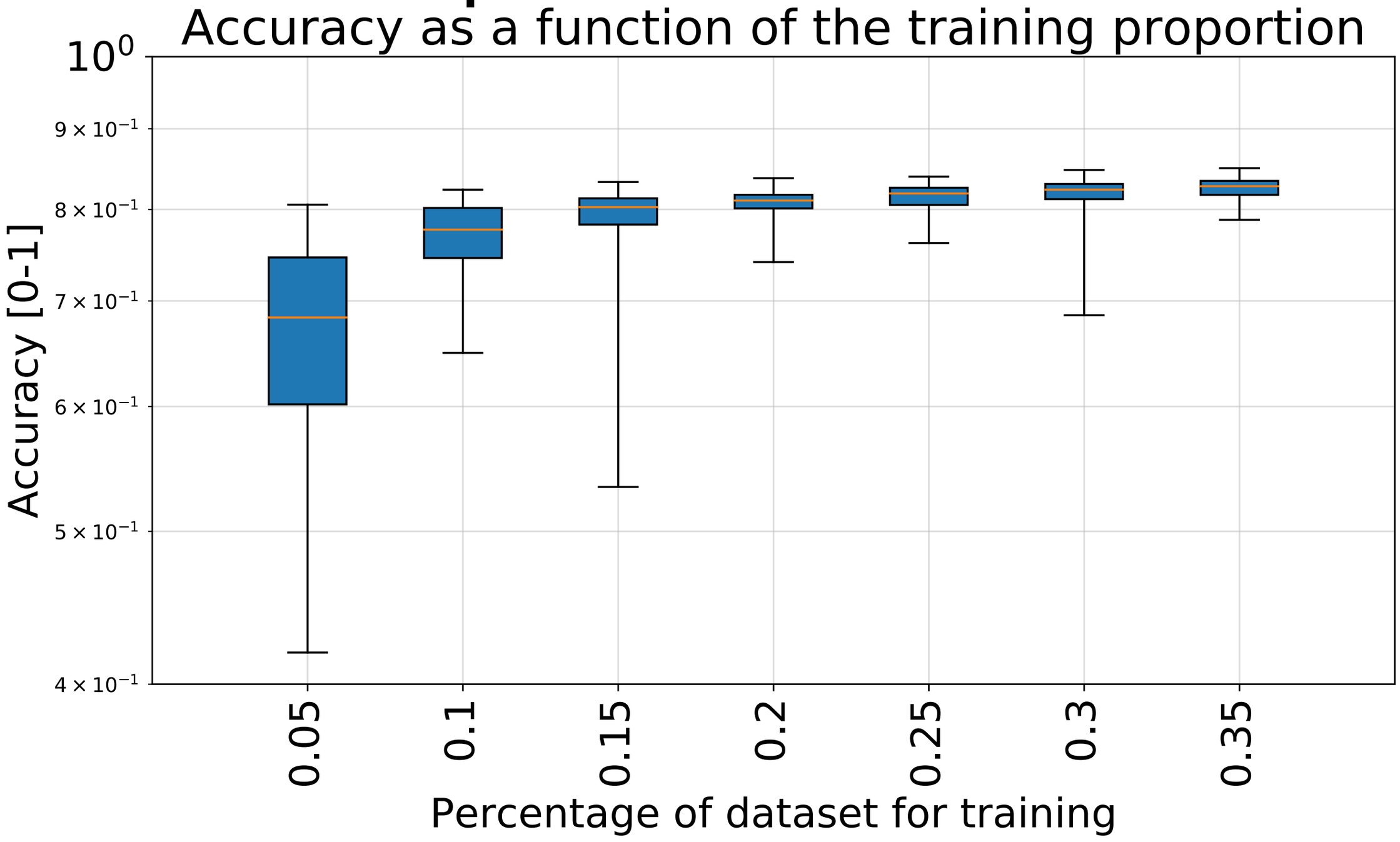
Data reduction and congestion signature



System architecture



ML performance



Conclusions

- We introduce the Stable to model latency
- We found delta-gamma patterns as an innovative way to detect congestion
- We applied Stable+ML to built an automatic system to detect congestion