Fitting well known distributions to network traffic fails

We’re studying enterprise traffic distributions of flow counts in time intervals of 4 to 512 seconds.

**It’s hard to know if bursty traffic distributions are really heavy-tailed**

Many presumed fits to heavy-tailed distributions don’t truly fit power-laws. [1] This is good news, since, if they fit more well-tempered distributions, prediction and anomaly detection are easier.


Strong dependencies among network flows lead to bursty traffic, which makes modeling and prediction hard.

In general, hierarchical models, e.g. mixture models, are both economical (fewer parameters) and offer better explanations.

More precisely, fits to a mixture model reveal a substantial exponential component...

... whose parameters are stable across time-step size.

Removing *the usual suspects* might explain it

Conditioning on traffic from persist destination addresses shows strong differentiation in spiky-ness. Rare connections (those not persistent) are less spiky. Thus http connections to external websites tend to exhibit better behaved traffic, and...