Non-stationarity and high-order scaling in TCP flow arrivals: a methodological analysis

INGI
Département d'ingénierie informatique

Steve UHLIG

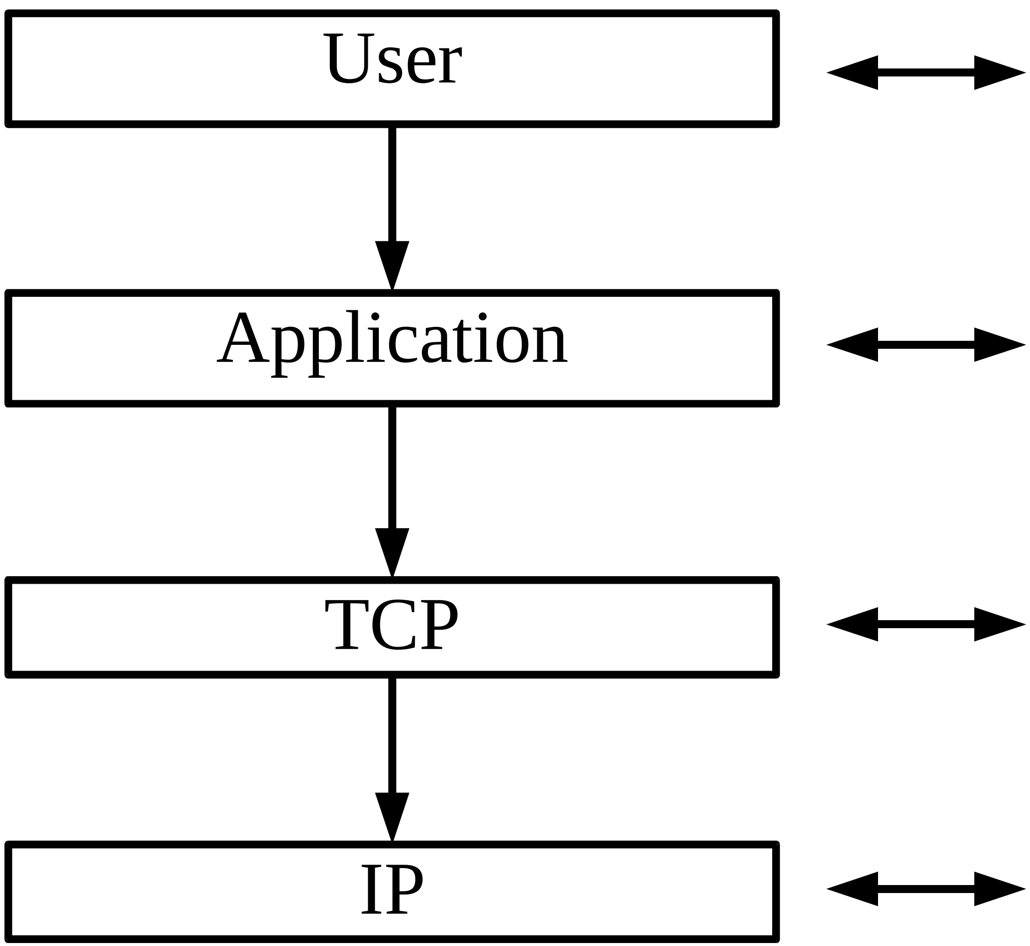
Computing Sciences and Engineering Dept.
Université catholique de Louvain, Belgium

Agenda

- Internet traffic tutorial
- Scaling
- Second-order scaling
- High-order scaling
- Conclusions

Internet traffic tutorial

The Internet stack



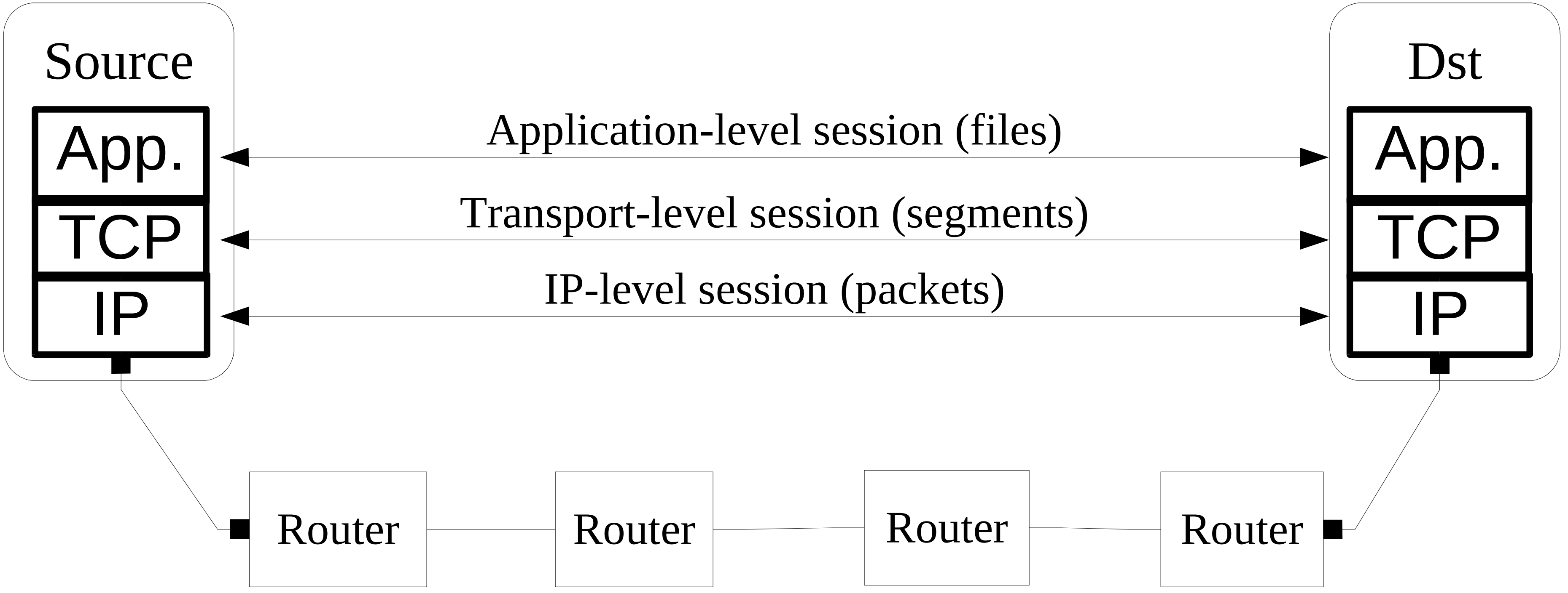
User asks for content (web pages, file transfers, interactive sessions,...)

Application exchanges data between user and remote host

TCP breaks data into segments, provides reliability and controls sending rate

TCP segments encapsulated into IP packets sent on physical link

The Internet stack illustrated

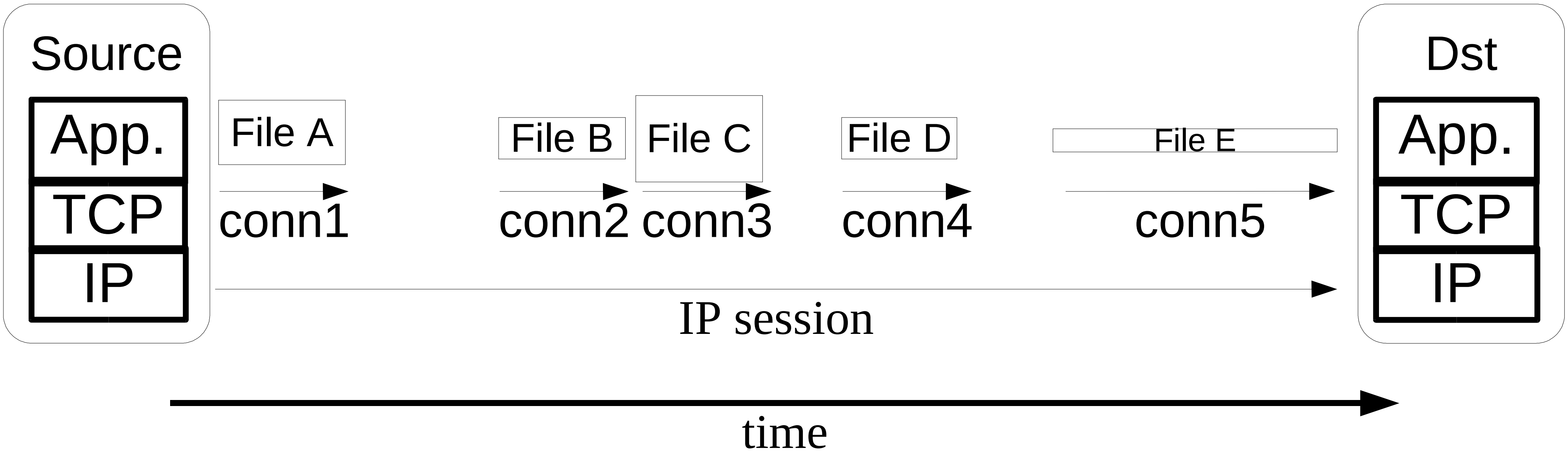


Application-transport interface

1. Application decides to exchange data between source and destination

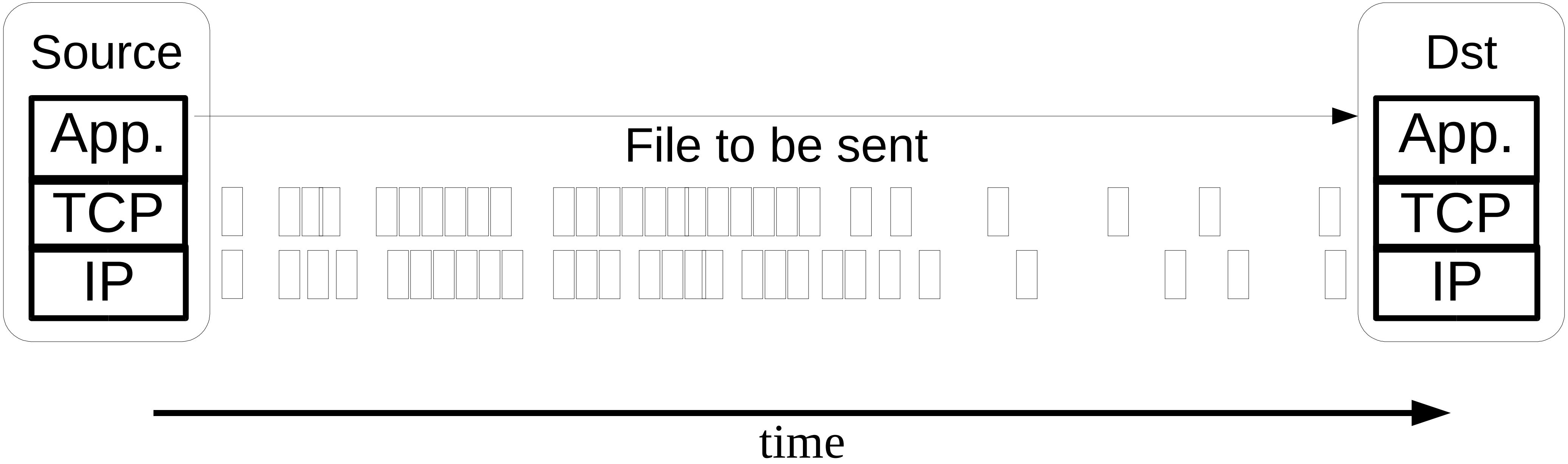
2. Several TCP connections established to exchange each "object"

3. Single IP session as seen from the outside as an ON/OFF process



Transport-IP interface

1. Application sends data to transport layer
2. TCP breaks data into segments
3. TCP state machine decides how and when segments are to be sent to IP layer
4. IP layer sends packets over network



Scaling

Scaling

- **<u>Definition:</u> statistics over some variable scales as a power-law of time (or another index)**
- **If increments of the process X(t) are replaced by wavelet coefficients *d(j,k)*, scaling can be studied through the behavior of** $E|d(j,k)|^q$
- **Power-law relationship can be linear or non-linear:**
  - **self-similarity:** $E|d(j,k)|^q \propto \exp(qH \ln(2^j))$
  - **multiscaling:** $E|d(j,k)|^q \propto \exp(H(q) \ln(2^j))$

Warning: scaling vs. physical models

- **Scaling is a *statistical property* of the signal**
- **Scaling models don't need to have anything to do with the *physics* of the signal**
- **Scaling is *one* particular way to describe a signal**

Second-order scaling

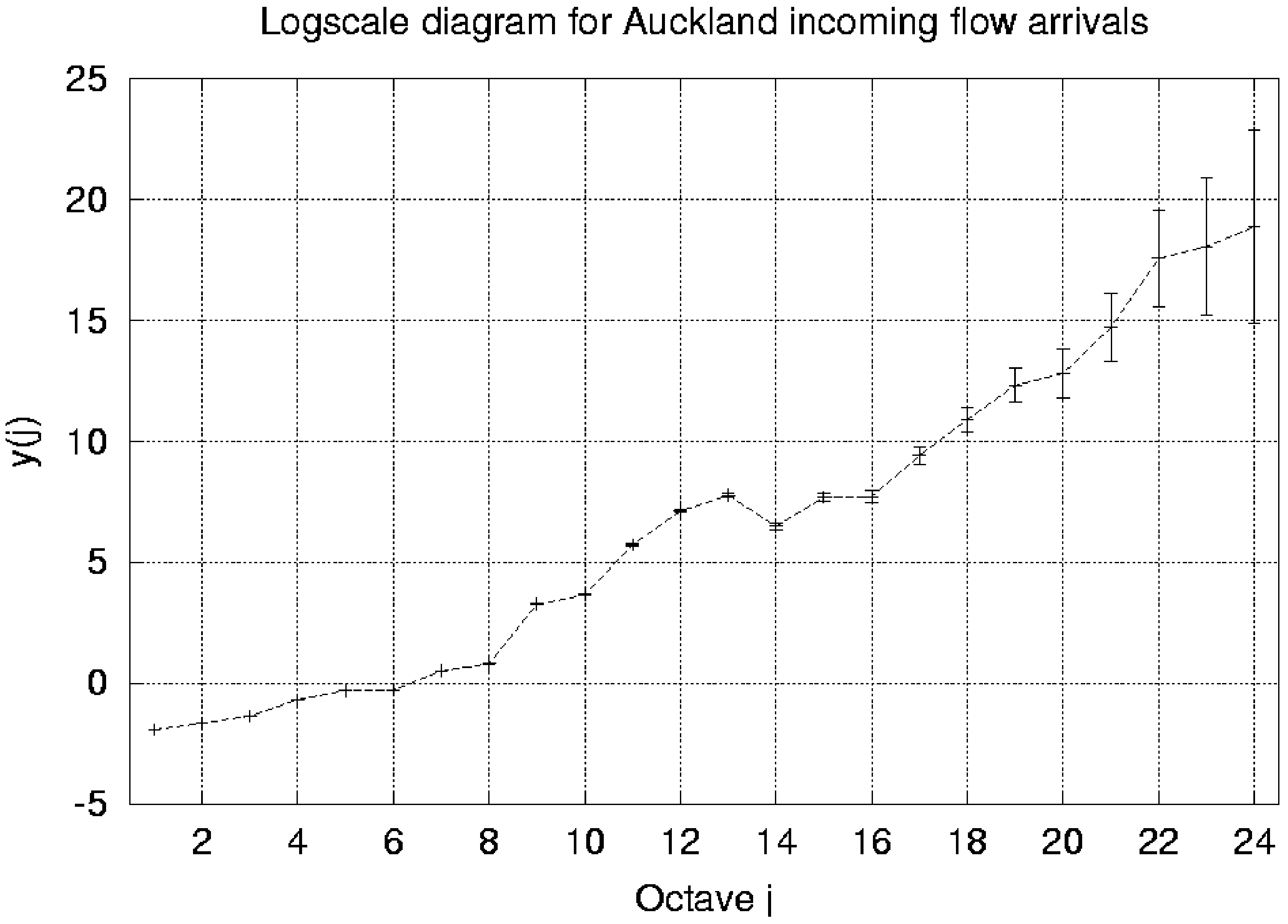
Second-order scaling

- **Identification of second-order scaling can be done based on the *logscale diagram* :**
  - **plot $\log_2(\mu_j)$ against j where**

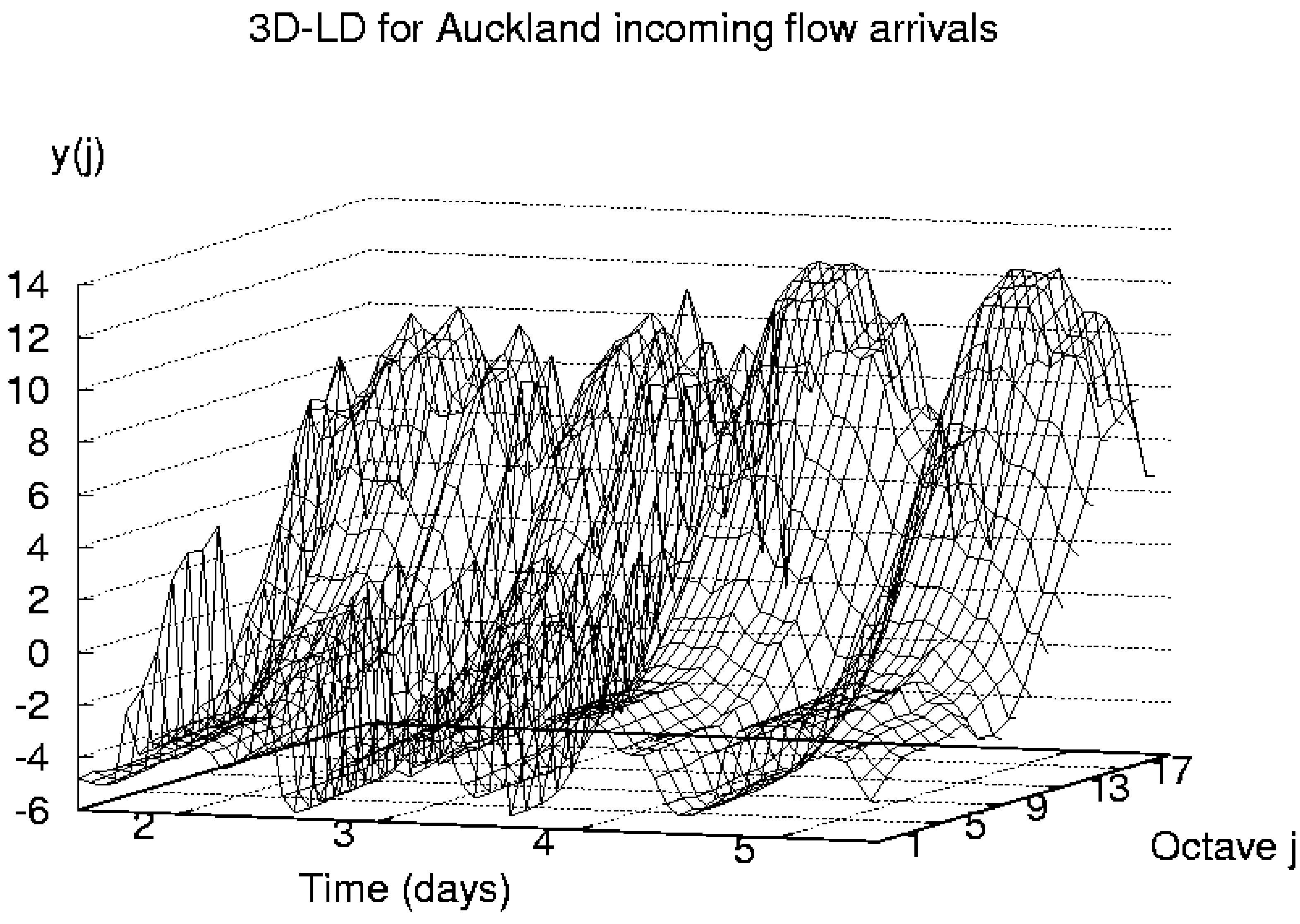
$$\mu_j = \sum_k |d(j,k)|^2$$

- **Scaling requires alignment of $\log_2(\mu_j)$ together with confidence intervals!**

Blindly trusting the LD



Don't trust the LD



High-order scaling

Identifying scaling components

- **Two scaling regimes have been identified in network traffic:**
  - **multiscaling at small timescales (?)**
  - **LRD at large timescales**
- **Physical root of multiscaling might be TCP cascade behavior, LRD might be caused by heavy-tails coupled with ON/OFF behavior**
- **Question: how can we "identify" different scaling components in some particular signal ?**

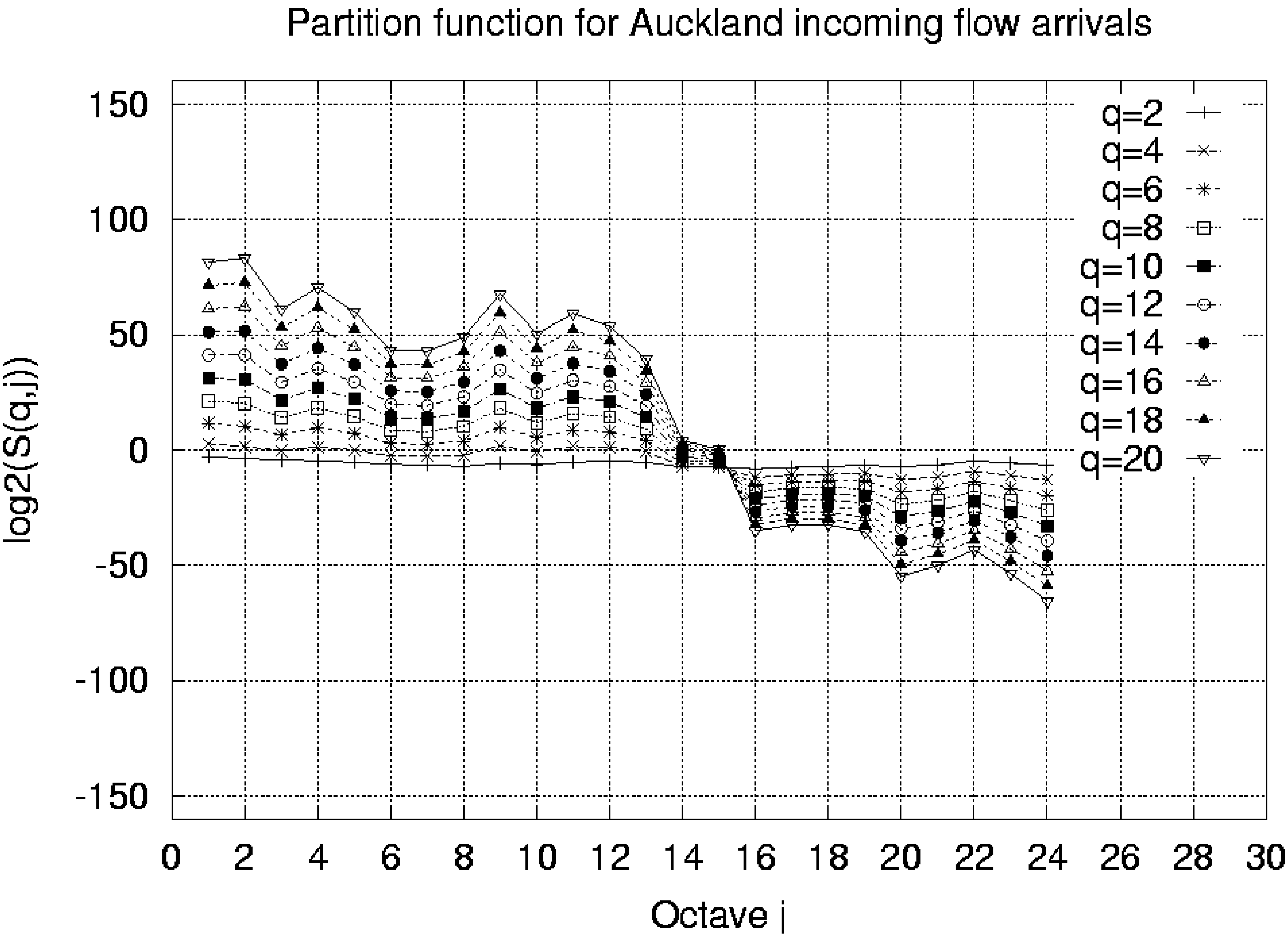
Identifying scaling components

- **Networking literature today mainly relies on LD and MD estimators (Abry et al.) to "identify" scaling**
- **Very few studies have tried to check that scaling is an invariant property of the time series, not just a bias occuring from subsets of the trace**
- **Can we do better than literature by leveraging more information contained in the *d(j,k)*?**

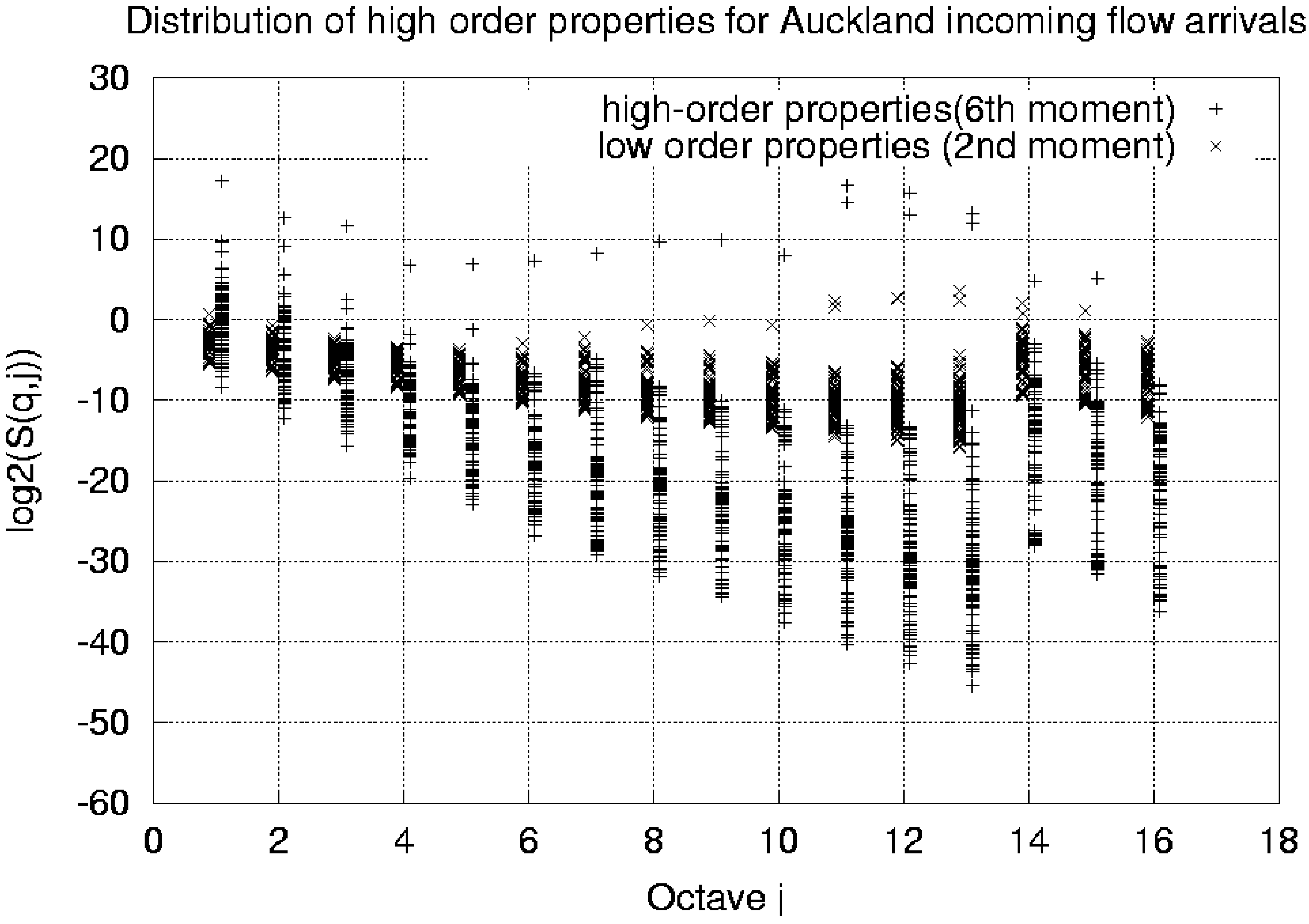
High-order scaling

- **Distinguishing between scaling behaviors can be done by studying how moments *q* scale with timescales *j***
- **Partition function $S(q,j) = \sum_k |2^{-j/2}\, d(j,k)|^q$**
- **$M(j) = \langle \ln(S(q+1,j)) - \ln(S(q,j)) / \ln(S(2,j)) - \ln(S(1,j)) \rangle_q$**
  - **no strict scaling: $M(j) < 1$**
  - **monoscaling: $M(j) \sim 1$**
  - **multiscaling: $M(j) > 1$**

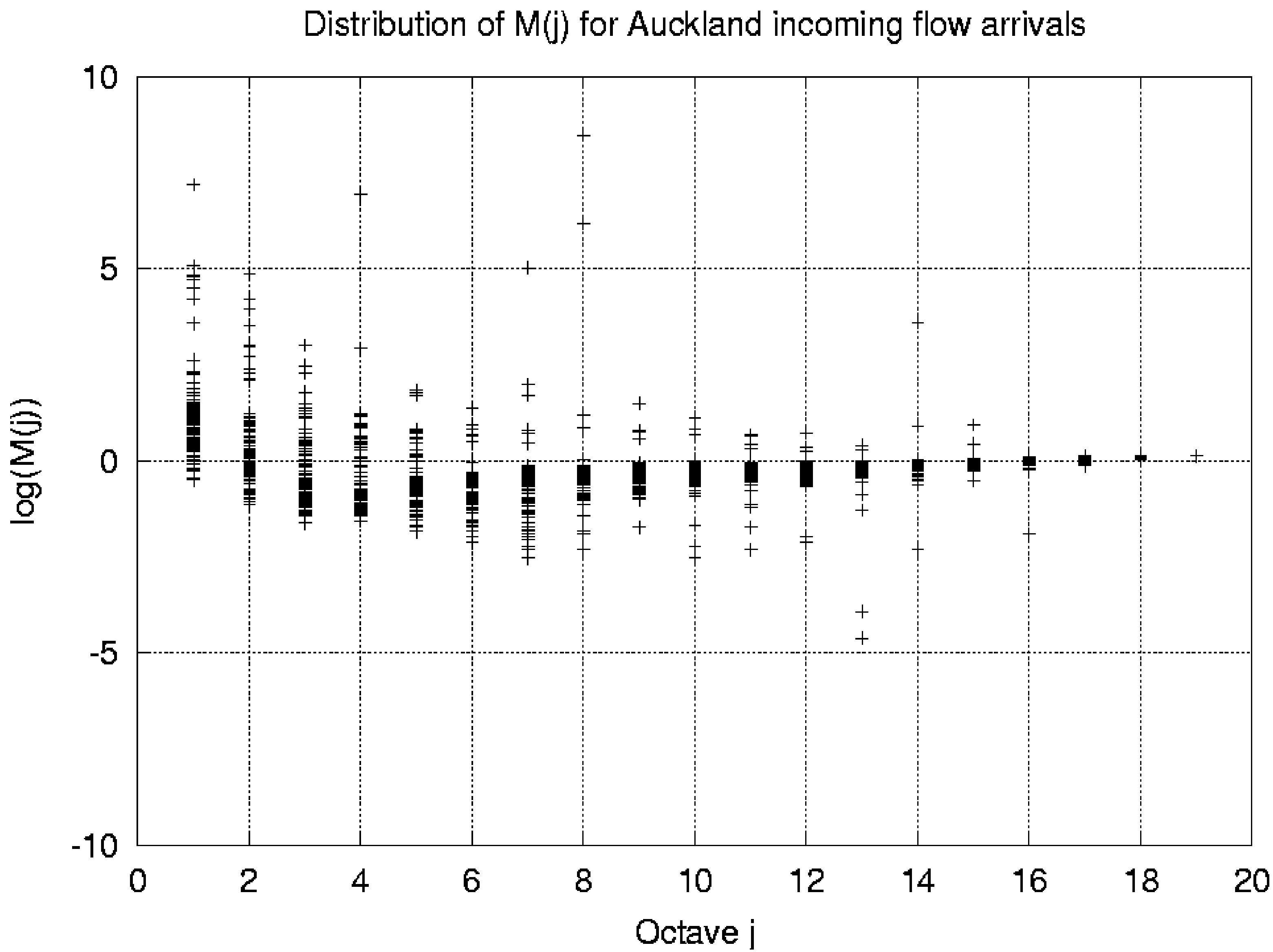
Blindly trusting S(q,j)



Don't trust S(q,j) for complex signals



Bias from non-stationarity



Components of TCP flow arrivals

- **Checking for stationarity and carefully looking at behavior of high-order moments leads to:**
  - **typically no high-order scaling (multiscaling) in TCP flow arrivals**
  - **no strict scaling at small timescales and LRD at large ones**
  - **high-order scaling only under pathological behavior (still unexplained)**

Conclusions

- **There are enough available tools today to analyze complex signals with wild scaling behaviors, people just fail to use them!**
- **Open issues in signal processing:**
  - **lack in ability to focus on a limited range of timescales**
  - **not much is available to check bias from non-stationarity and verify model assumptions, beyond simple statistical tests**