Motivation

- Application at Lucent Technologies: software to monitor distributed IP-based services.
- Goal: characterize various metrics (e.g. e-mail transaction times), locally and aggregated, updated over time.
- Constraint: computing at the node, amount of data transmitted to server.

The Idea

(Absortimate, Update, Aggregate)

- **Approximate** the empirical distribution for each metric & node (agent)
- **Update** each approximation periodically for new data at the node.
- **Aggregate** the ecdfs for relevant groupings of nodes (e.g., regions)
**Update for each agent**

- Objects represent each evolving quantile estimate: `a <- seqQuants(....)`
- OOP-style functions to simulate updating, aggregating: `a$merge(data)` (modifies `a`)
- Using R closures (object contains functions with a shared environment for updates).

**Aggregate agent records**

- R simplifies large-scale simulation studies, with varying statistical assumptions.
- R also helps in the algorithm development in C, by calling an R tracer from C.
Summary

• An example of the productive interaction between applications and research, typical of Bell Labs research (in the old days).

• An interesting algorithmic study to estimate distributions with distributed, ongoing data.

• The productive computing environment centered on R essential for productivity.