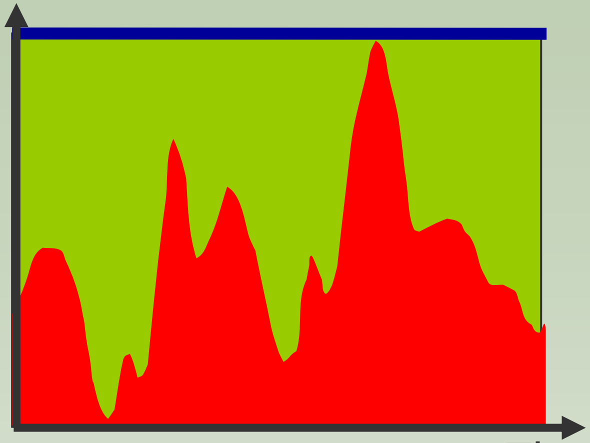
Real-Time Measurement of End-to-End Available Bandwidth using Kalman Filtering

IEEE/IFIP Network Operations & Management Symposium
NOMS 2006, Vancouver, April 2006

Svante Ekelin[1,4], Martin Nilsson[2], Erik Hartikainen[1,3], Andreas Johnsson[4], Jan-Erik Mångs[1], Bob Melander[1,4], Mats Björkman[4]

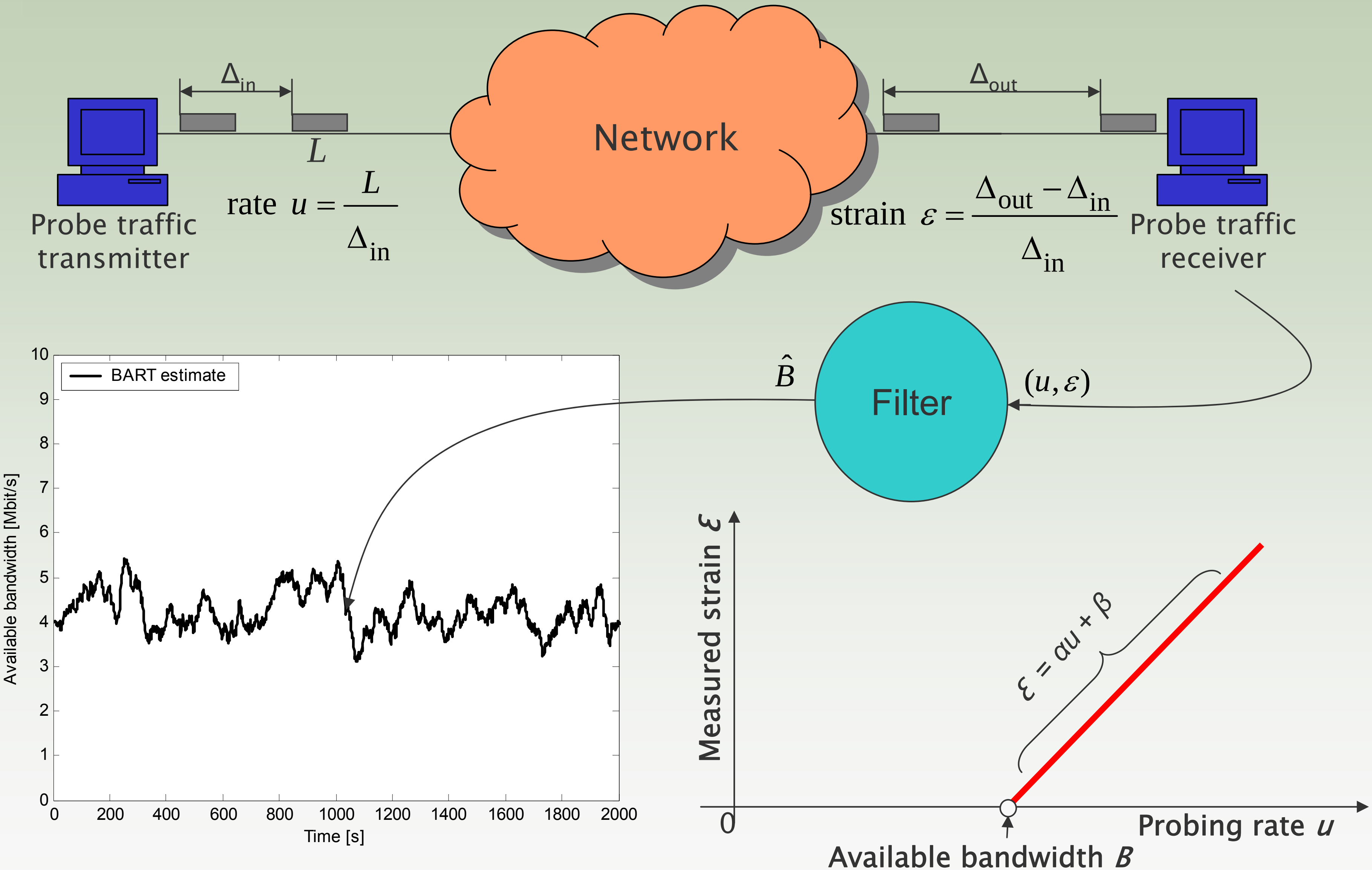
[1] *Ericsson Research*, [2] *Swedish Institute of Computer Science*, [3] *Linköping University*, [4] *Mälardalen University*

outline



- overview
- objective
- challenges
- approach : recursive estimation
- BART – Bandwidth Available in Real-Time
- results
- outlook

executive summary



objectives & challenges

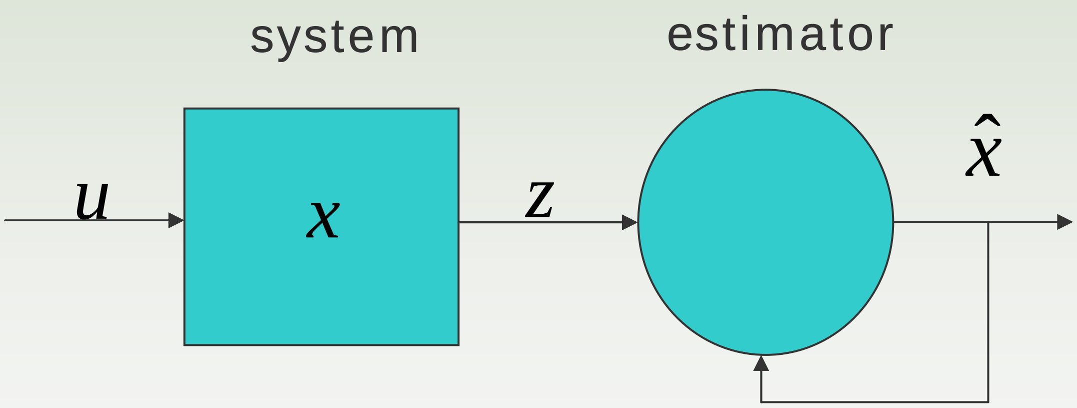
- **objective: estimate available bandwidth in real-time**

- **challenge #1 : no direct access to network nodes**
- **solution : active probing**
  - Send probe packets at various rates over the path, and look for signs of congestion. The threshold rate is the available bandwidth.
    - only requires control over end systems

- **challenge #2 : target is moving**
  - available bandwidth is not constant, but changes during the time it takes to collect the information. We want real-time estimation.
- **solution : ?**

approach: recursive estimation

- **mental framework**
  - We want to estimate the state *x* of a system without having direct access. However, we can observe an output *z*, which depends on the state, and we can influence the system with a control input *u*.
- **estimator** algorithm depends on
  - ***system model*** describing how the system state *x* evolves in time
  - ***measurement model*** describing how *z* depends on *x*



estimator calculates:

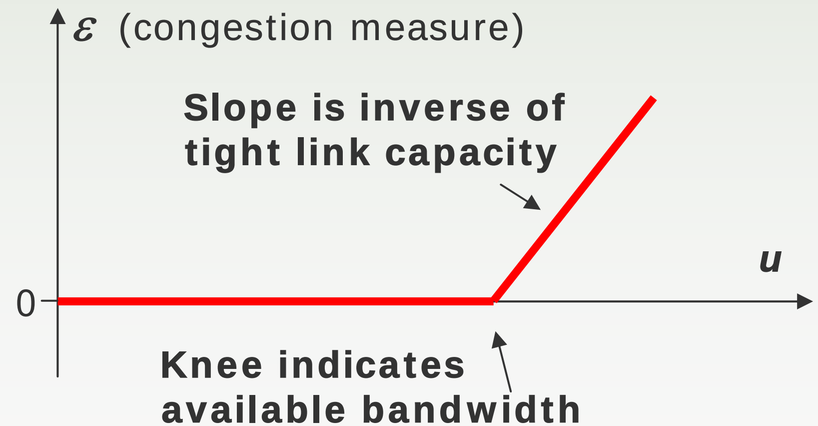
$$\hat{x}_k = g(\hat{x}_{k-1}, z_k)$$

modeling our problem

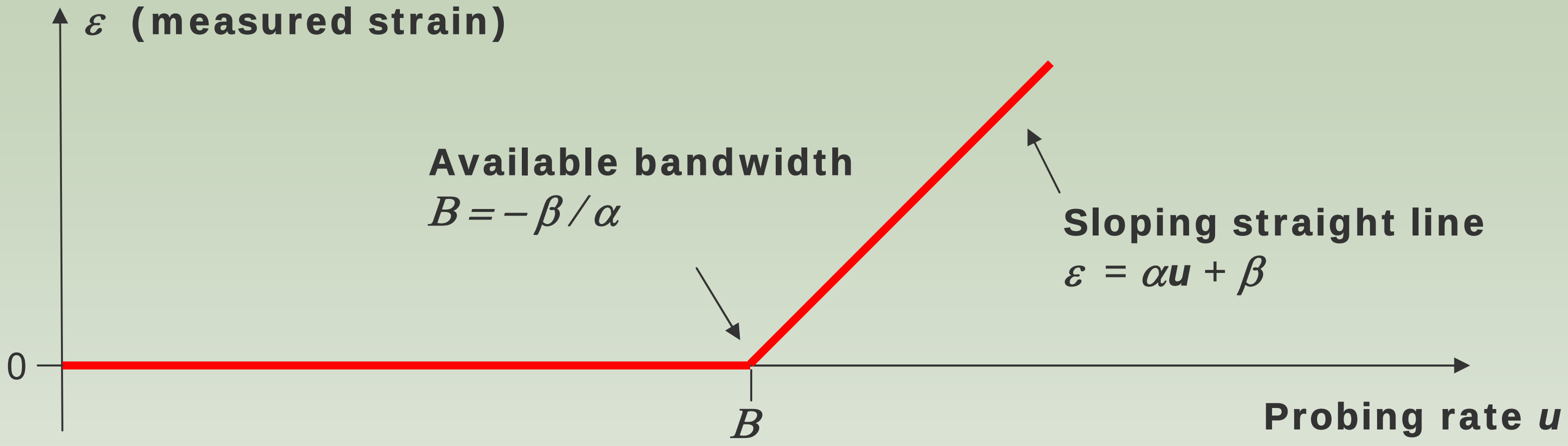
- **the system to be estimated is the network path**

- system state potentially has very high dimensionality, if taking into account detailed static and dynamic characteristics $\boldsymbol{s}_i$ and $\boldsymbol{d}_i(t)$ for hop $i = 1,\ldots,N$
- but we want to avoid unnecessary complexity
- from previous bandwidth estimation methodology work:

In a fluid model of traffic flow, one expects a piecewise linear dependence of the strain ε on the probing rate u. Further, only the tight hop contributes as long as there are not multiple congested hops. (Melander, 2000)



the BART approach



- this curve describes dominating contribution to strain
- the relevant aspects of the system state for our purpose are captured by the sloping straight line

BART model:

- system described by two-dimensional state vector $x = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$
- system evolution model : $x_k = x_{k-1} + w_{k-1}$
- measurement model : $\varepsilon_k = Hx_k + v_k$ where $H = [u \quad 1]$.

(*w* and *v* are noise terms; for the sake of linearity we only consider the region *u*>*B*)

Kalman filtering

- **Since we managed to express the model linearly, we can apply the Kalman filter!** (Kalman, 1960)

- **The Kalman filter is a special type of recursive estimator, with a host of desirable properties :**

  - easy to implement in a few lines of code
  - light-weight w.r.t. memory and CPU
  - tried and tested in many other engineering fields
    - trajectory tracking, computer graphics, analog signal control (phase-locked loop)...
  - proven to be optimal estimator under specific conditions
  - often well-performing even if these conditions are broken

the heart of the estimator: Kalman filter equations

”prediction”

$$\hat{x}_k^- = A\hat{x}_{k-1}$$

$$P_k^- = AP_{k-1}A^T + Q$$

”correction”

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-)$$

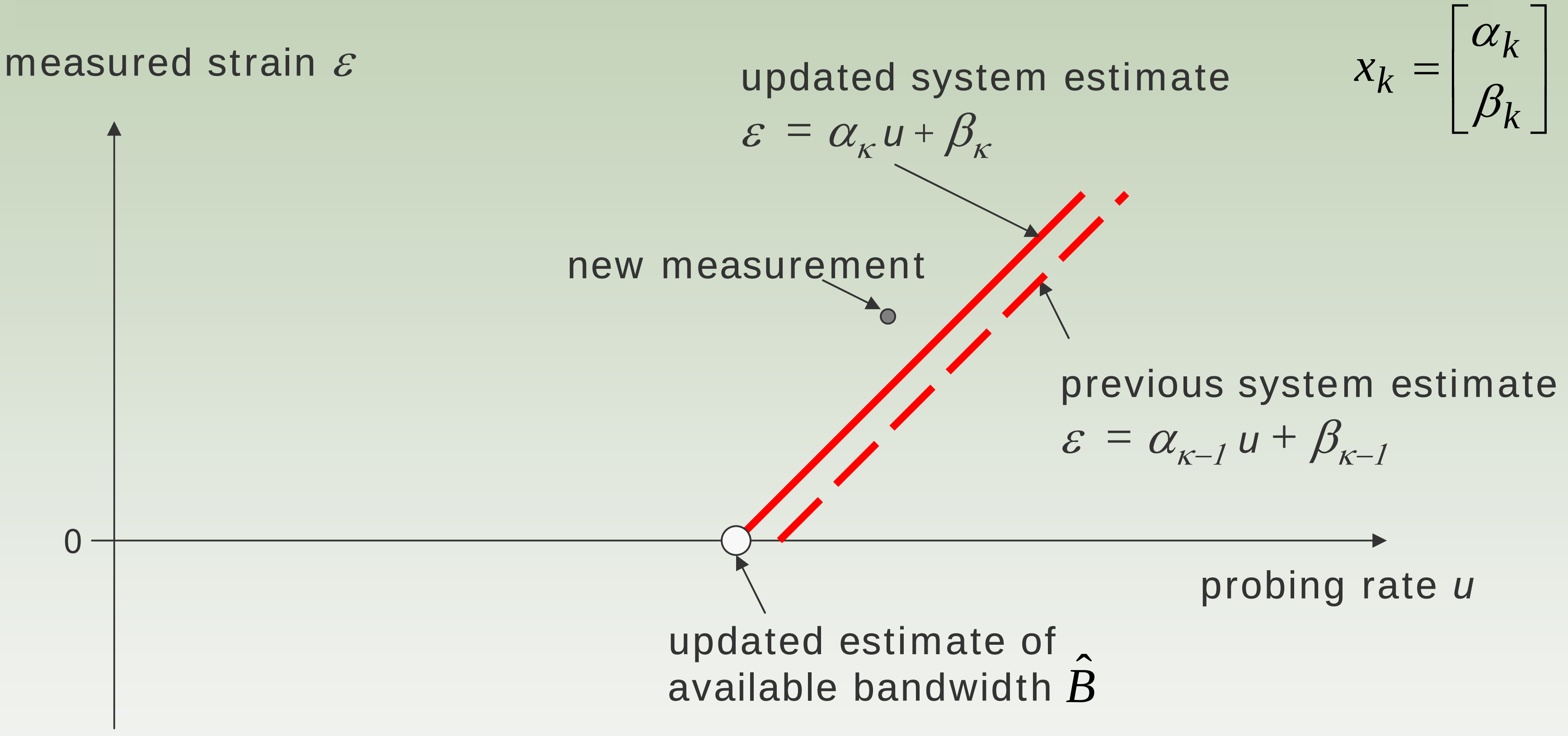
$$P_k = (I - K_k H)P_k^-$$

$$k \leftarrow k+1$$

where the Kalman gain $K_k = P_k^- H^T (HP_k^- H^T + R)^{-1}$

increases with Q and decreases with R

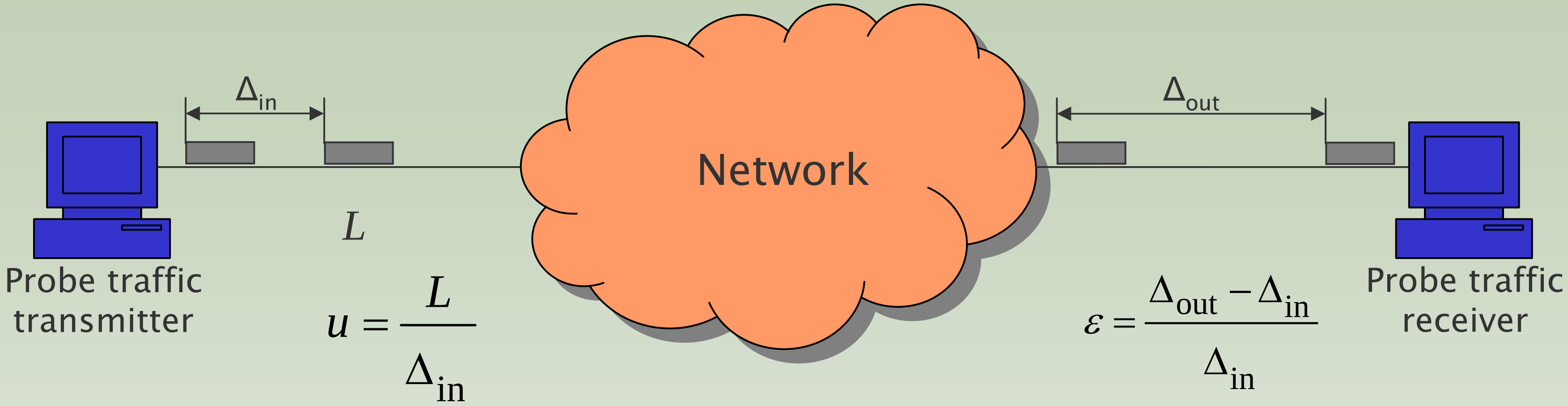
estimate update in BART



BART modeling summary

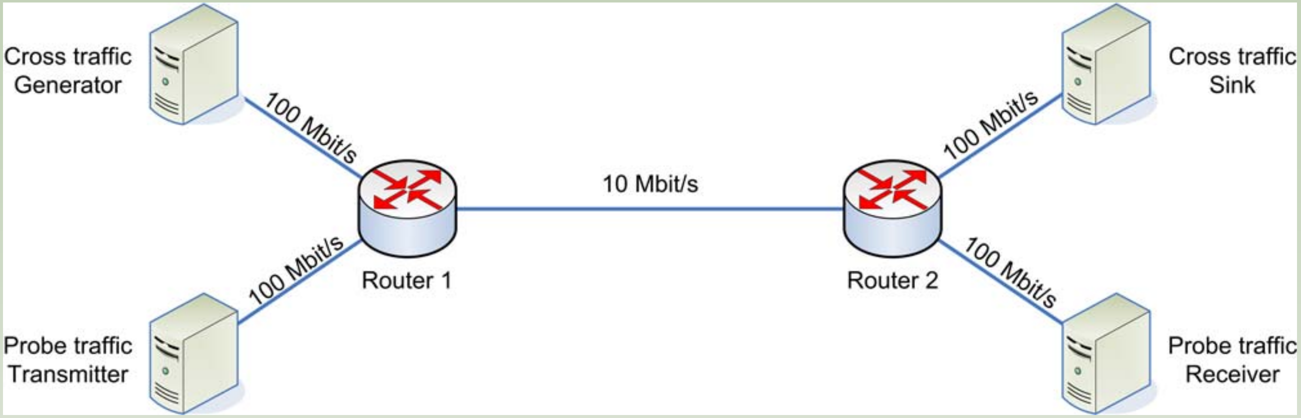
- **follows the spirit of Occam's Razor:**
  - avoid making more assumptions than needed
  - a model should be as simple as possible, but no simpler
- **describes the system state by the two most critical quantities**
  - bottleneck available bandwidth
  - bottleneck capacity
    - (or rather, a transformation of these, to obtain linearity)
- **formulates the models in a linear fashion**
  - enables Kalman filtering

BART execution summary



- **samples the path with a sequence of probe-packet pairs (or train) at a rate *u***
- **randomizes *u* for each sampling**
- **receiver computes the average strain, inputs this to the Kalman filter, which produces an updated estimate**

empirical validation

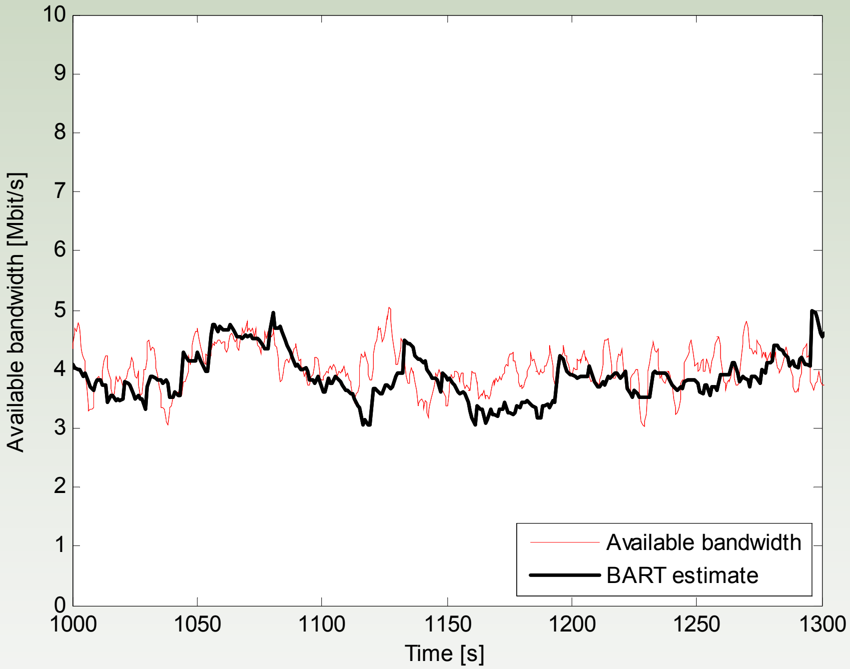


- **A BART implementation was tested in a lab network**
  - 16 probe packet pairs organized as a train of 17 packets (1500B)
  - one sample per second, uniformly chosen between 1 and 20 Mbps
    - 0.2 Mbit/s probe traffic overhead on average

- **Various traffic cases were emulated using a packet generator**
  - high/low user traffic aggregation
  - different distributions for inter-arrival time
    - exponential
    - Pareto

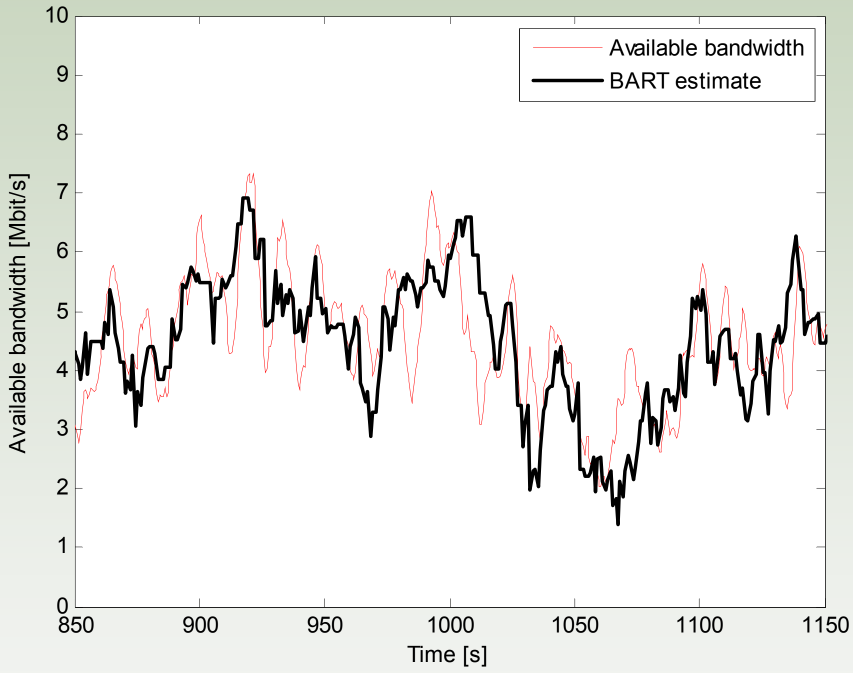
- **True available bandwidth was recorded using `tcpdump` for reference**

results – 100 users, Pareto

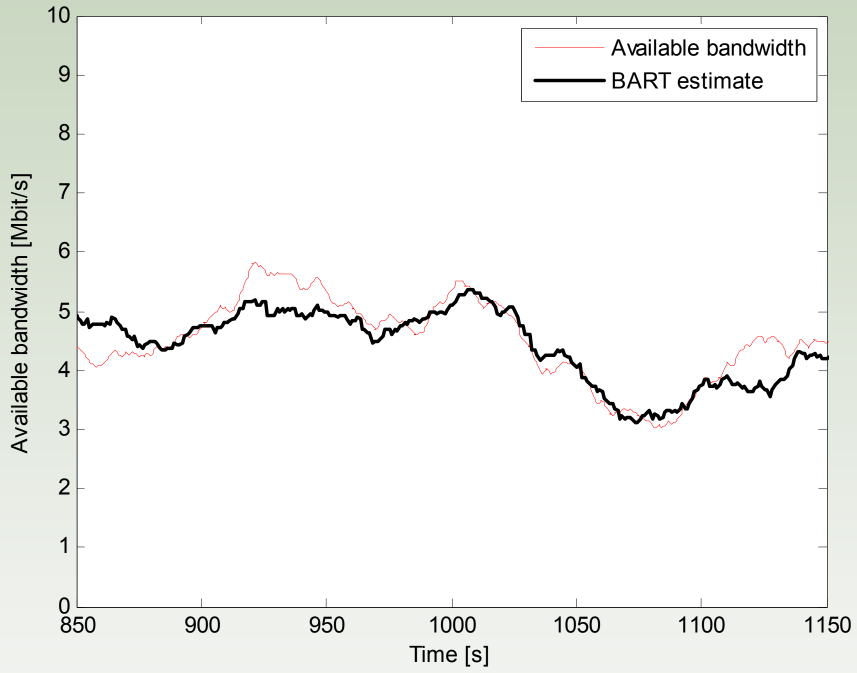


results – 10 users, Pareto

4 s



32 s



final slide...

- **summary and conclusions**
  - We found a way of modeling the problem such that Kalman filtering can be applied to available bandwidth estimation.
  - This approach allows for an updated estimate for each sampling, i.e. real-time characteristics.
  - Validation experiments show reasonable accuracy.
- **under study**
  - performance in a high-speed multihop network
    - Is effect from other (non-tight) hops non-negligible?
    - Is effect from detailed structure of cross traffic non-negligible?
  - Is there need for more elaborate modeling, e.g. higher dimensionality system state vector?
  - Is the model ”simpler than possible”...?
    We think not, but experience will tell.

Thank you!

svante.ekelin@ericsson.com