UPPSALA
UNIVERSITET

Dealing with Uncertain Arrival Rates in Resource Allocation

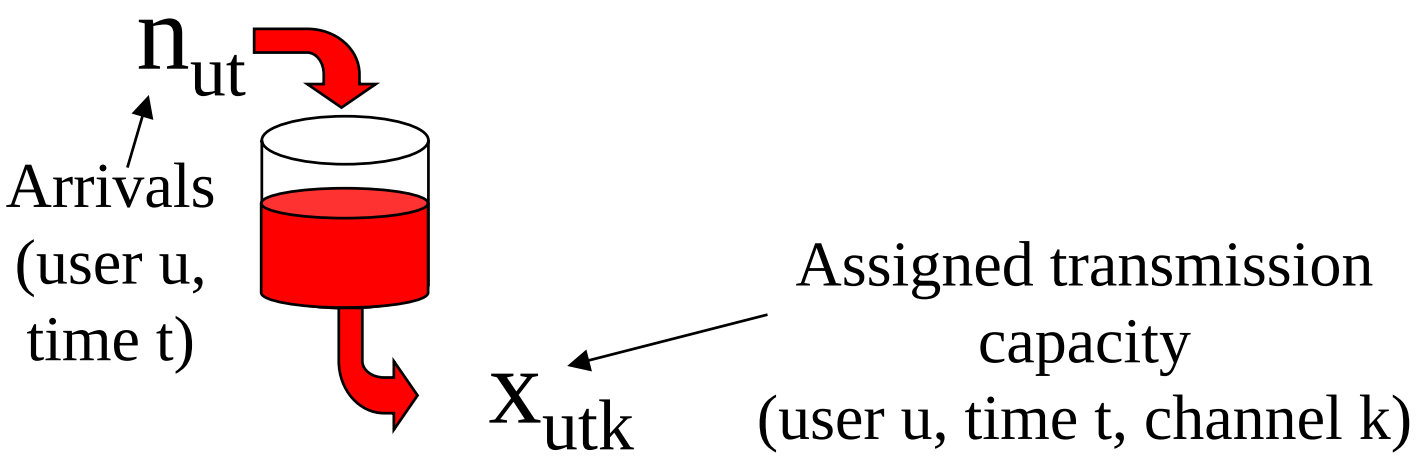
With a discussion on criteria in scheduling

Mathias Johansson
Signals & Systems Group

UPPSALA
UNIVERSITET

Introduction to scheduling

- Cellular downlink scenario
- Wish to allocate transmission channels to users so as to satisfy their individual service preferences
- Each user has a buffer at the BS in which data arrive with unknown rates



Mathias
Johansson
Signals &
Systems Group

UPPSALA UNIVERSITET

Traffic prediction and adaptation

- In theory, better performance can be obtained by scheduling over several time slots
  - Particularly with QoS constraints
- Requires channel prediction over longer periods
  - Should average criterion over pdf for channel to account for higher uncertainty
- Requires prediction of arrival rates
  - ”Always data to send” unrealistic assumption

Mathias Johansson
Signals & Systems Group

UPPSALA UNIVERSITET

Scheduling under uncertainty – criteria

- Minimize the expected total buffer contents after the scheduled horizon
  - Gives maximum expected throughput
- Use constraints to satisfy delay or rate requirements

$$\langle L \rangle = \sum_{u=1}^{U} \sum_{n_u=0}^{\infty} \sum_{x_{ut}=0}^{\infty} p(n_u|I) p(x_{ut}|I) g\left(S_u + n_u - \sum_{t=1}^{T} x_{ut} \right)$$

$$g(x) = x \text{ if } x > 0, \text{ otherwise } g(x) = 0$$

Mathias Johansson
Signals & Systems Group

UPPSALA UNIVERSITET

But is sum throughput a good criterion?

If it weren't for unfairness, we might be tempted to say yes.

But what if:

- user 1, average throughput=10, but can send 100 the next slot,
- user 2, averages 1000, can send 500 the next slot.

Who'd you prioritize?

Mathias Johansson
Signals & Systems Group

UPPSALA
UNIVERSITET

Performance is relative!

- D. Bernoulli (1738) noted that in human society, the utility resulting from a small increase in wealth seems to be proportional to the amount already possessed
  - uniquely determines a logarithmic criterion (which is what prop. fair uses)

$$y = b \log\left(\frac{\alpha + a}{\alpha}\right)$$

Change in wealth

Arbitrary constant

Previous wealth

Mathias
Johansson
Signals &
Systems Group

UPPSALA
UNIVERSITET

Another reasonable criterion?

- Queue stability is sometimes advanced as an important property
  - An algorithm which empties all buffers whenever possible is called stable
  - Algorihtms with this property are also called "throughput optimal"
- Several variants exist, e.g. M-LWDF
  - $\max c_u S_u^k$

Transmission rate

Buffer size

Arbitrary constant

Mathias Johansson
Signals & Systems Group

UPPSALA UNIVERSITET

Another reasonable criterion?

- M-LWDF is *ad hoc* but approximates a minimization of the buffer levels squared.
  - Does not maximize throughput!
- When possible, it empties buffers but otherwise it can be disastrous
  - A user flooding its buffer will get all resources
- PF is not ”stable”!

  ...but should we care???

Mathias Johansson
Signals & Systems Group

UPPSALA UNIVERSITET

What about delays and fairness? More opinions...

Neither PF nor max expected throughput is optimal in terms of fairness or delays

- Fairness is not a fundamental criterion to maximize
  - Communication systems should transfer information. Fairness is just a constraint.
- How to compromise between delay and throughput?
  - Open question, DEEA is an alternative
  - Trade-off depends on application (traffic class)

Mathias Johansson
Signals & Systems Group

UPPSALA UNIVERSITET

Traffic prediction 1: Maximum entropy approach

- We wish to determine p(n|I) given some information I
- If we keep record of average arrival rates <n>, what is the best inference concerning n we can do?
- Use Maximum Entropy Principle to assign a probability distribution for n
  - n positive with known mean $\Rightarrow$ Exponential distribution for n
  - ME distribution is least biased possible

Mathias Johansson
Signals & Systems Group

UPPSALA UNIVERSITET

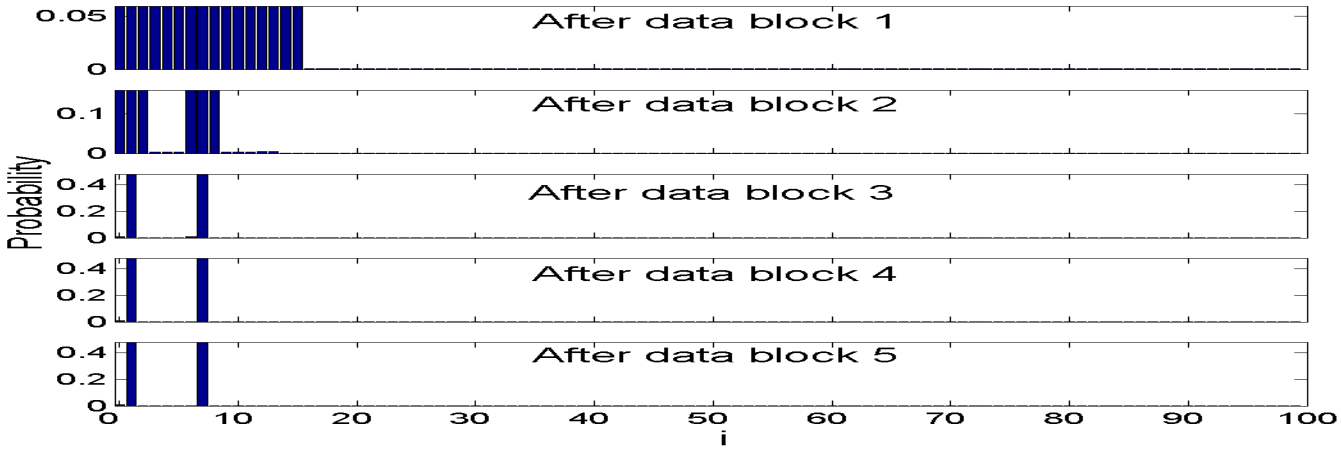
Traffic prediction 2: Adaptive inference

- We wish to determine p(n|I) given
  I = past arrival statistics
  and learn patterns adaptively
- Imagine using histograms
  - Too few observations in comparison to possible inflow sizes
- Instead, partition the inflow-axis into a number of ’bins’
  - Count arrivals within each bin
  - Adapt the bin size to obtain high resolution at intervals of high intensity and lower elsewhere

Mathias Johansson
Signals & Systems Group

UPPSALA
UNIVERSITET

Traffic prediction – Bin probabilities



Using K bins and letting

- m_k = past number of arrivals of size within bin k
- M = total number of observations

we have (after some calculations...)

$$p(n \in \text{bin } k \mid m_k, M, I) = \frac{m_k + 1}{M + K}$$

Mathias Johansson
Signals & Systems Group

UPPSALA UNIVERSITET

Traffic prediction – Adaptation

- Based on the bin probabilities, how do we adapt the bin positions and sizes?
  - Wish to have a quantized distribution which is as close to the exact distribution as possible.
- Formally, we wish to maximize the mutual information between the two distributions

Theorem:

Maximizing the mutual information is equivalent to maximizing the entropy of the bin probability distribution.

Mathias Johansson
Signals & Systems Group

UPPSALA UNIVERSITET

Traffic prediction – Adaptation

Proof:

$$
\begin{aligned}
I(k,n) &= \sum_{k=1}^{K} \sum_{n=n_{min}}^{n_{max}} p(nk) \log \frac{p(nk)}{p(n)p(k)} \\
&= \sum_{k=1}^{K} \sum_{n=n_{min}}^{n_{max}} p(nk) \log \frac{p(k \mid n)}{p(k)} \\
&= -\sum_{k=1}^{K} \sum_{n \in \text{bin } k} p(n \mid k)p(k) \log p(k) \\
&= -\sum_{k=1}^{K} p(k) \log p(k)
\end{aligned}
$$

Mathias Johansson
Signals & Systems Group

UPPSALA UNIVERSITET

Traffic prediction

- The optimum bin partition is adapted according to the M most recent arrivals:
  - Assume a uniform probability distribution within each bin,
    p(n) = bin probability / bin width
  - Redistribute the bins so that each bin has equal probability mass (=max entropy)
  - Approximate low-complexity solution requires single sweep over the possible arrival sizes.

Mathias Johansson
Signals & Systems Group

UPPSALA UNIVERSITET

Aside: Generalization to multi-dimensional case

- Interestingly, the approach generalizes to the multi-dimensional case with unknown time dependencies
- Possible to adaptively infer dependencies across time and variables
- A form of optimal self-organizing vector quantization can be achieved!

Mathias Johansson
Signals & Systems Group

UPPSALA UNIVERSITET

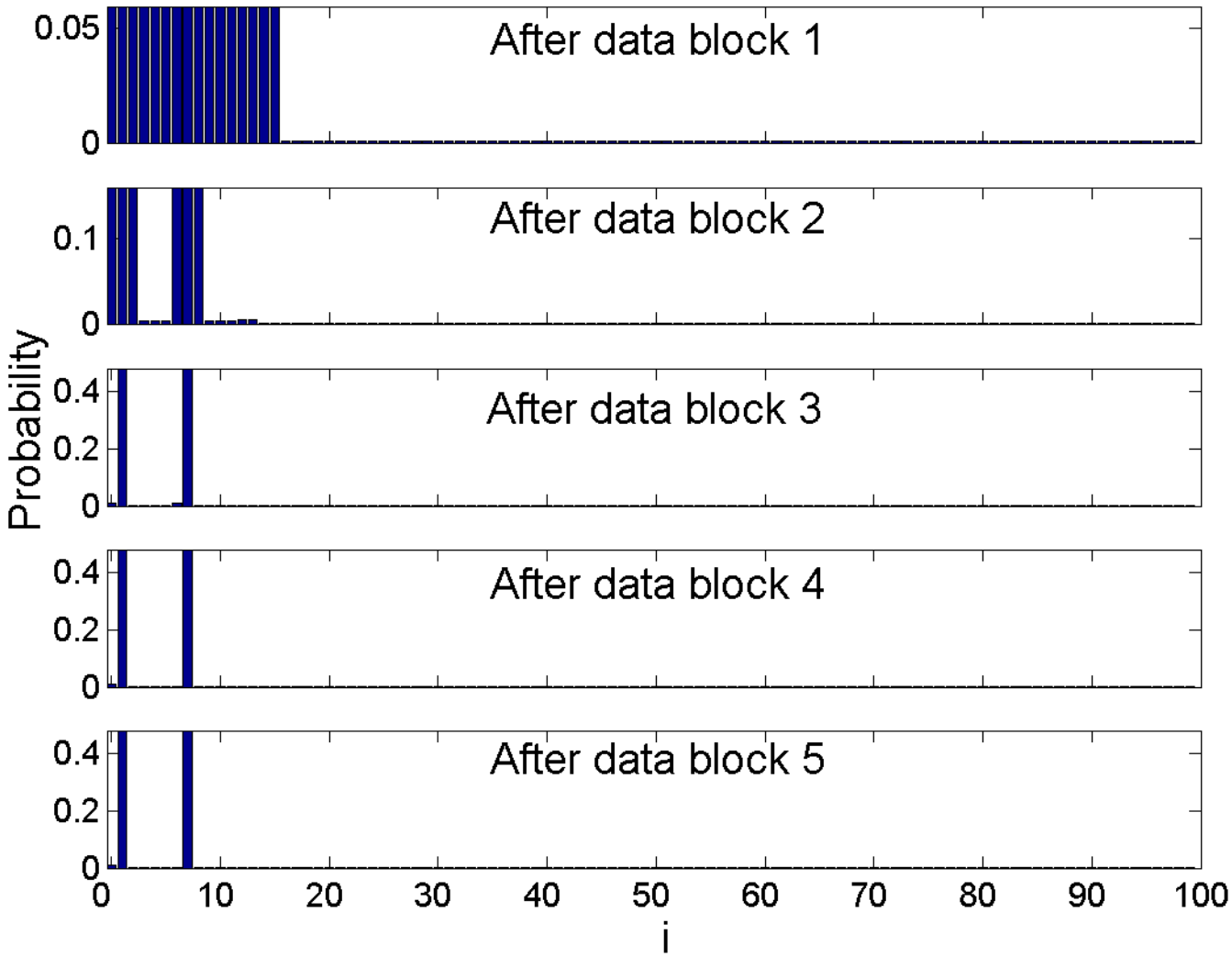
Example – Simulation set-up

- K = 6 bins,
- M = 100 observations between bin updates
- Min arrival rate=0, Max=100 [bit/time unit]
- Arrivals generated as
  - 50% 1-bit packets,
  - 50% 7-bit packets

  (switching between 2 fixed rates with equal frequency)

Mathias Johansson
Signals & Systems Group

Traffic prediction – Results



Mathias
Johansson
Signals &
Systems Group

UPPSALA UNIVERSITET

Summing up

- Presented two means for managing arrival rate uncertainty
  - Max Ent – less computations
  - Self-organizing histogram – flexibility
- The problem in comparing different means of dealing with uncertain rates is that of obtaining meaningful traffic traces