Combining age-based and channel-aware scheduling in wireless systems

Samuli Aalto and Pasi Lassila

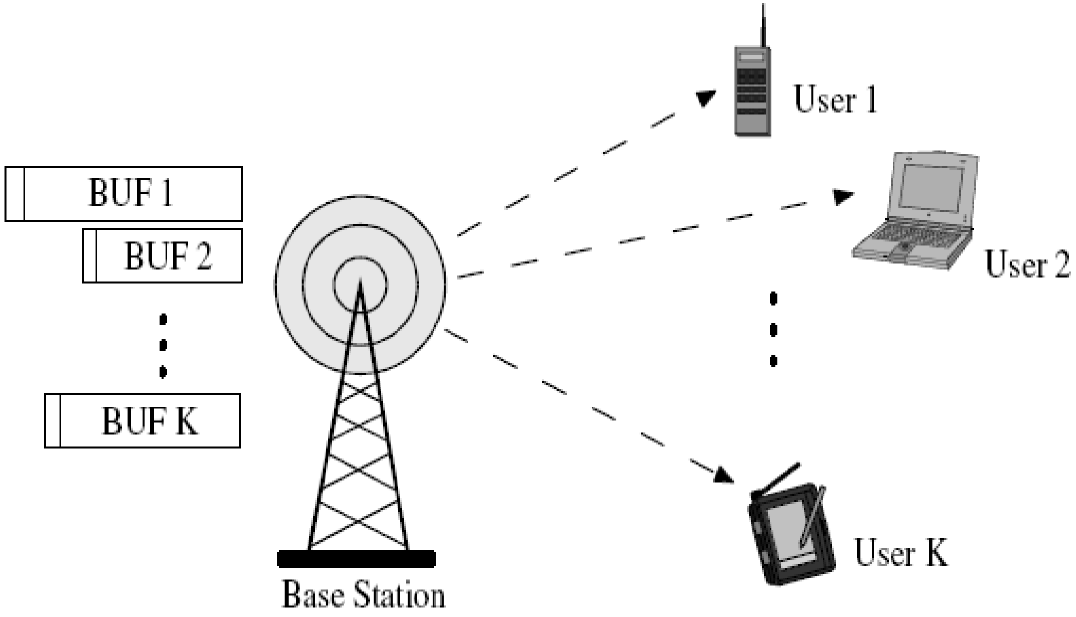
Networking Laboratory

Helsinki University of Technology

Email: {Samuli.Aalto, Pasi.Lassila}@tkk.fi

HSDPA/HDR systems

- Downlink transmissions
  - BS transmits to exactly one user in a time slot with full power



- Scheduling:
  - BS decides allocation of time slots for different users’ traffic
  - For example: round robin scheduling

Flow-level perspective

- We consider elastic traffic
  - Corresponds to users performing web surfing consisting of file transfers
  - Elastic means that applications tolerate variations in instantaneous rates
  - In the dynamic setting, file transfers (or flows) arrive randomly (Poisson arrivals) and have random sizes (typically heavy tailed)
- Performance expressed as mean file transfer delay or throughput
  - Users only care about the **total** time to transmit/receive the complete file
- Connection back to time-slot level
  - To transmit a typical file requires many time slots
  - Different traffic model from “packet level” approaches with for example i.i.d. arrivals per time slot, c.f., cμ-rule (Stolyar et al.)

Opportunistic scheduling / size-based scheduling

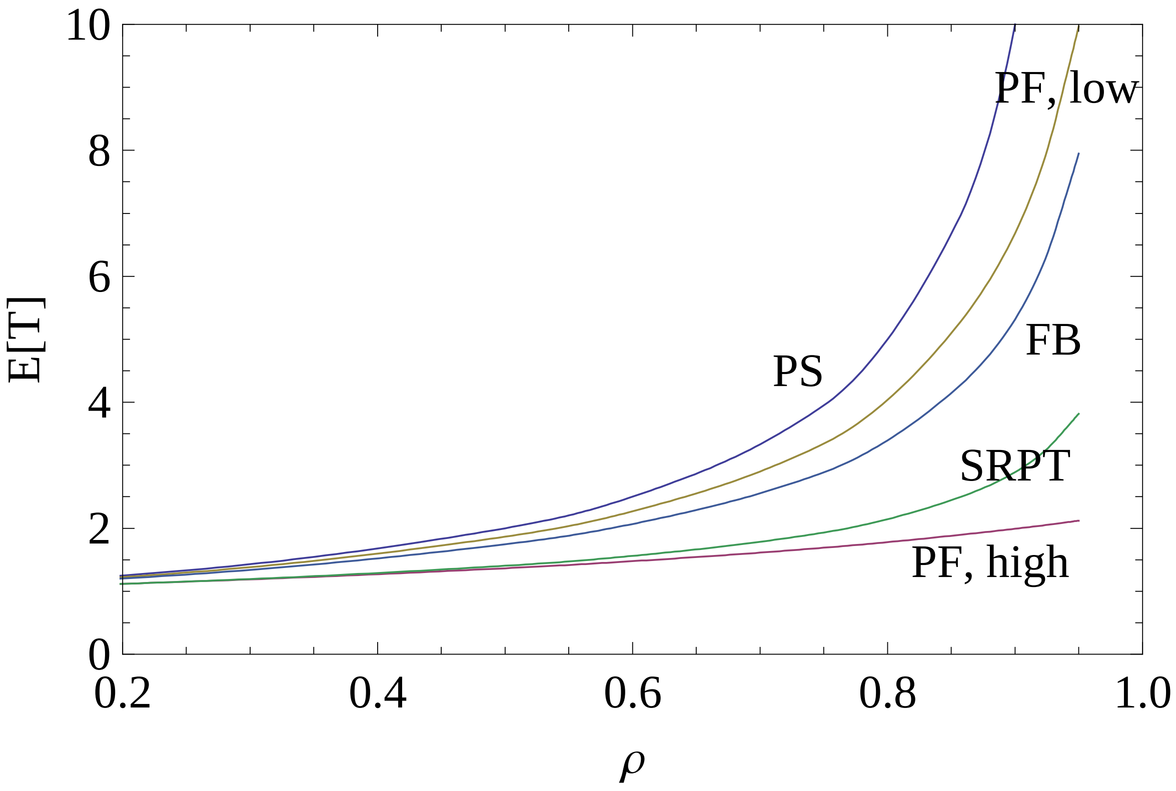
- Fast fading: the rate (or SNR) changes randomly in each time slot (mobility)

- Opportunistic channel aware scheduling
  - Idea is to exploit the channel variations between users and give the time slot to users in a good state (with high rate)
  - “Capacity increases“ due to scheduling gain

- “Standard” age-based schedulers in fast fading environment
  - Idea is to get rid of small flows as quickly as possible to minimize flow delay
  - Depending on what information is available, we have different policies SRPT, FB (LAS), PS
  - Standard approach would utilize knowledge of file sizes (bits) and **mean** rate of the users (not the instantaneous rates)

PF compared with “standard” SRPT and FB

- Age-based schedulers (SPRT, FB) perform better than simple PS, but gain from opportunistic scheduling can be (PF) much greater



Parameters

- Rayleigh fast fading channel
- Rate is linear in SNR up to maximum rate
- Symmetric users

Combining size/rate information

- Problem: how to combine instantaneous size and rate information?
  - Difficult problem, “optimal solution” is not known
- Our approach and results
  - Possible to derive many heuristics that combine size/rate information
  - We assume that sizes obey a continuous distribution (DHR type) but the possible channel rates form a discrete set
  - Analytical results comparing the optimal policy and some heuristics in a simple static setting
  - Simulations under heavy traffic to explore tradeoffs

Different ways to utilize size/rate info

- Assumption: there is a discrete set of possible user rates
  - Also, all users can achieve the maximum rate!

- Two classes of policies
  - Priority policies
  - Index policies

- Priority policies
  - Absolute priority on highest rate
  - Apply size information to break ties
  - Greedy approach for utilizing the channel

- Index policies
  - Single index value that combines rate and size information

Priority policies

- Give absolute priority to highest instantaneous rate
  - Idea: utilize the channel maximally
- If multiple flows have same highest rate, various policies are possible depending on size information available
  - SRPT-P
    - serve flow with least amount of bits left
    - aims for maximum efficiency
  - FB-P
    - serve flow with least amount of bits served
    - same as SRPT-P but with only knowledge of attained service (in bits)
  - RR-P :
    - serve flow with smallest throughput (attained service / time in system)
    - aims for increased fairness

Index policies

- PF (Relatively best)
  - Select user *k* with highest R_k / γ_k
  - R_k = instantaneous rate of user *k* and γ_k = throughput of user *k*
- RB (Relatively best)
  - Select user *k* with highest $R_k/\mathrm{E}[R_k]$; blind policy with respect to size info
- TAOS2 (Hu et al., Computer Networks, 2004)
  - Optimal one step decision rule for improving basic SRPT policy
  - M = nof jobs in the system
  - X_k = remaining number of bits for user *k* (SRPT-like information)
  - Users are ranked in ascending order of $X_k/\mathrm{E}[R_k]$ (basic SRPT)
  - I_k = rank of user *k*, select user k^* so that

$$k^* = \arg\min_k \left(-\left(M - I_k + 1\right) \frac{R_k}{\mathrm{E}\left[R_k\right]} \right)$$

- FB-TAOS2
  - Replace X_k with attained service A_k in ranking (i.e., served bits thus far)

Analytical study of optimal policy

- Assumptions
  - take time slot length $\Delta = 1$
  - 2 possible rates (r^{min}, r^{max}), p = prob. that rate is r^{min} (symmetric case)
  - 2 jobs with given size (size / r^{min} = integer), no new jobs arrive
  - Simple discrete time decision problem

- Performance
  - Total time to serve both jobs until completion (total completion time)

- Objective
  - Compare optimal policy with standard PF policy and those schedulers that use SRPT like information
  - Reference schedulers: PF and two “best” schedulers, SRPT-2L and TAOS2
  - Optimal policy can be solved via dynamic programming

Optimal policy

- Optimal policy can be solved by dynamic programming

serve 1 with rate r_1 serve 2 with rate r_2

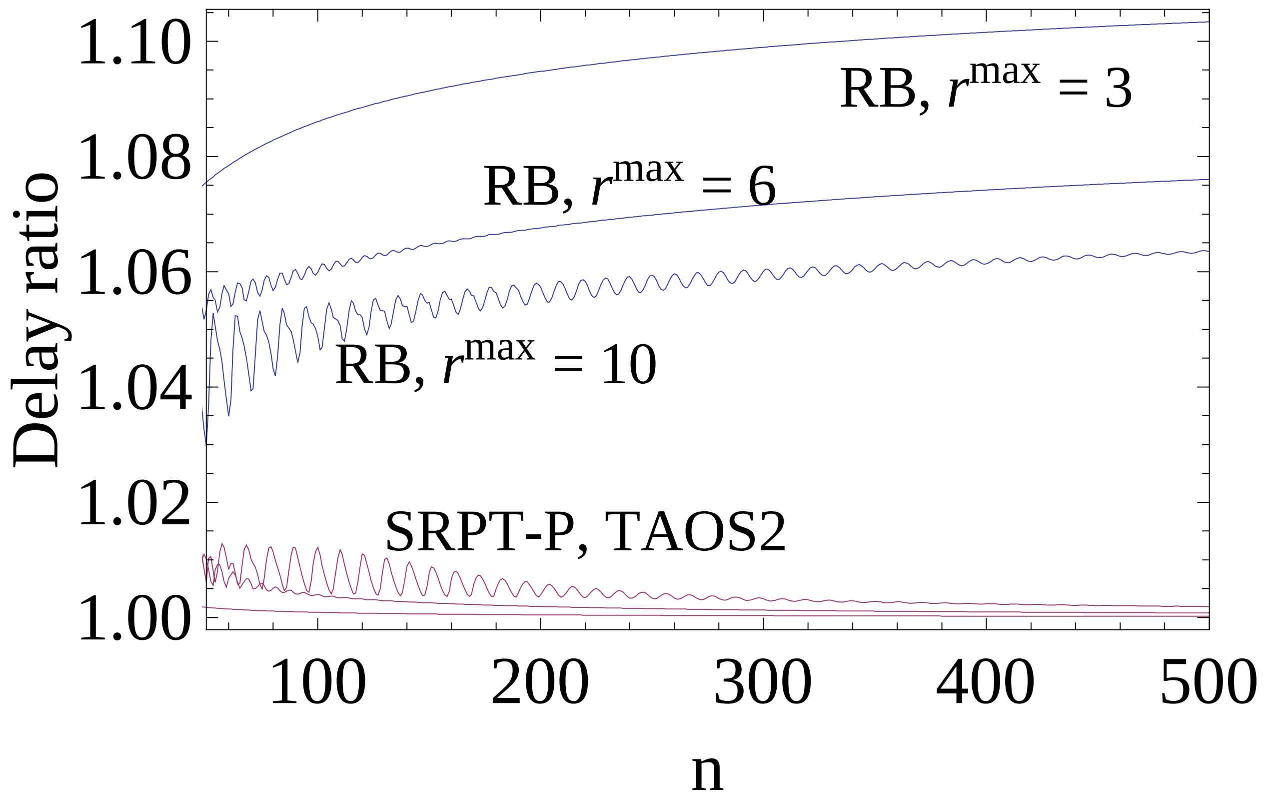
$$v(n_1, n_2, r_1, r_2) = 2\Delta + \mathrm{Min}\ [(1), (2)]$$

$$\begin{aligned}
(1)\quad & p^2 v(n_1 - r_1\Delta, n_2, r^{\min}, r^{\min}) + p(1-p)v(n_1 - r_1\Delta, n_2, r^{\min}, r^{\max}) + \\
& p(1-p)v(n_1 - r_1\Delta, n_2, r^{\max}, r^{\min}) + (1-p)^2 v(n_1 - r_1\Delta, n_2, r^{\max}, r^{\max}) \\
(2)\quad & p^2 v(n_1, n_2 - r_2\Delta, r^{\min}, r^{\min}) + p(1-p)v(n_{11}, n_2 - r_2\Delta, r^{\min}, r^{\max}) + \\
& p(1-p)v(n_1, n_2 - r_2\Delta, r^{\max}, r^{\min}) + (1-p)^2 v(n_1, n_2 - r_2\Delta, r^{\max}, r^{\max})
\end{aligned}$$

  - Easy to compute v(n,0,*,*) for all 4 combinations of channel states and then just iterate the above

- Similar analysis possible also for PF, SRPT-2L and TAOS2

Comparisons with optimal policy



Parameters

- r^{min} = 1
- p = 0.5 (symmetric channels)
- channel initial state = {1,1}

– Conclusion: TAOS2 and SRPT-2L very close to optimal while improvement over PF can be 5-10%

Dynamic simulations

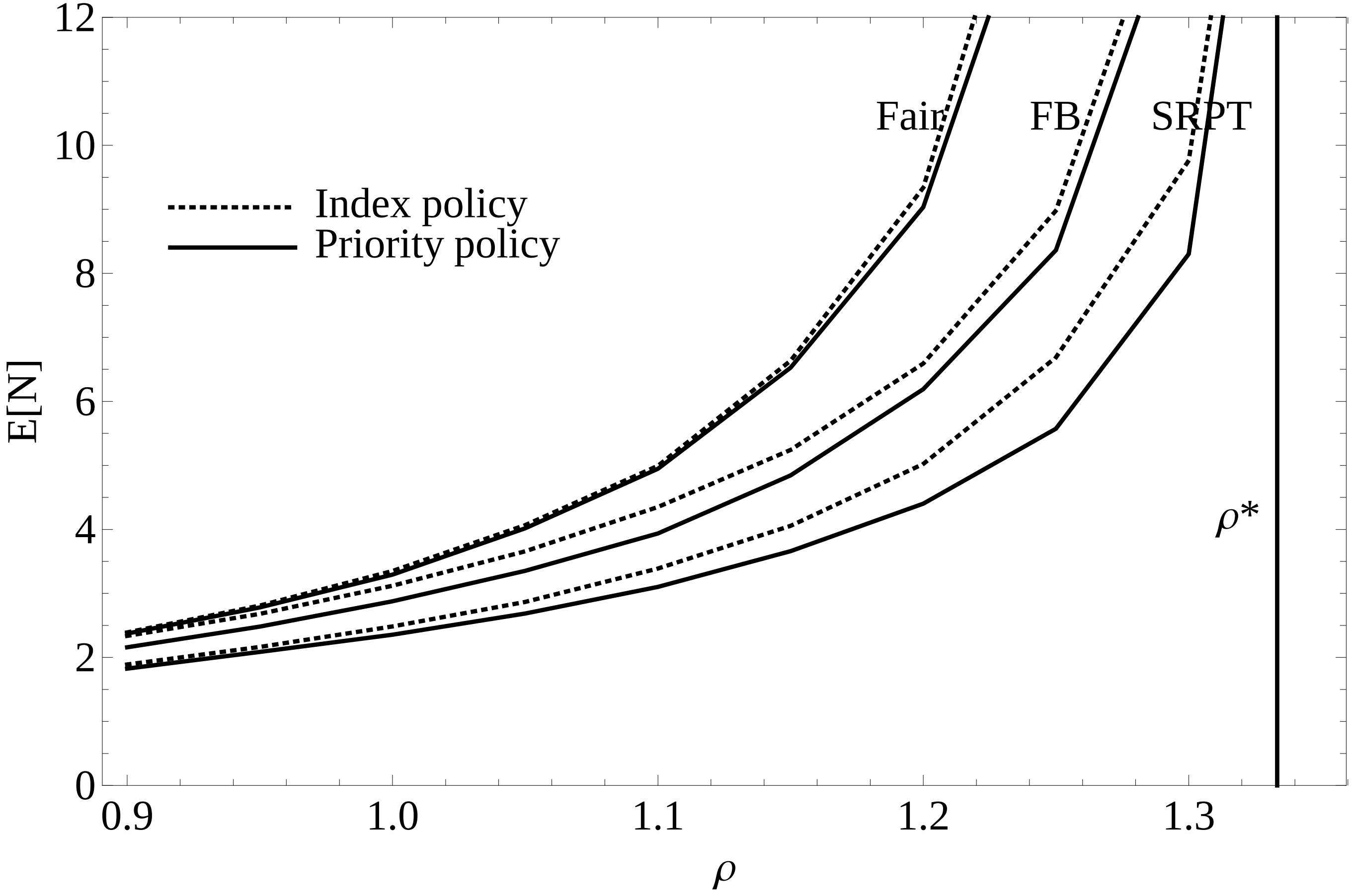
- Idea is to study the heavy traffic behavior of the policies under different settings for the user rates
    - In the setting where Δ is very small compared with time scale of arrivals and departures
    - Poisson arrivals, Pareto(2.0) file sizes
    - We fix λ=1 and vary the service times to get different loads

- Symmetric (= all users are identical) vs. asymmetric settings
    - 2 user classes, equal arrival rates in both classes, $\lambda_1 = \lambda_2 = 0.5$
    - symmetric/asymmetric achieved via parameterization of rates

- Different rate scenarios in an i.i.d. channel
    - Case1: only 2 possible rates, small difference
    - Case 2: only 2 possible rates, large difference
    - Case 3: same set of rates as in HSDPA (11 rates)

Symmetric case, 2 rates, low variability

- Parameters: $r_1 = 1$, $r_2 = 2$, $p_{11} = 0.5$, $p_{12} = 0.5$

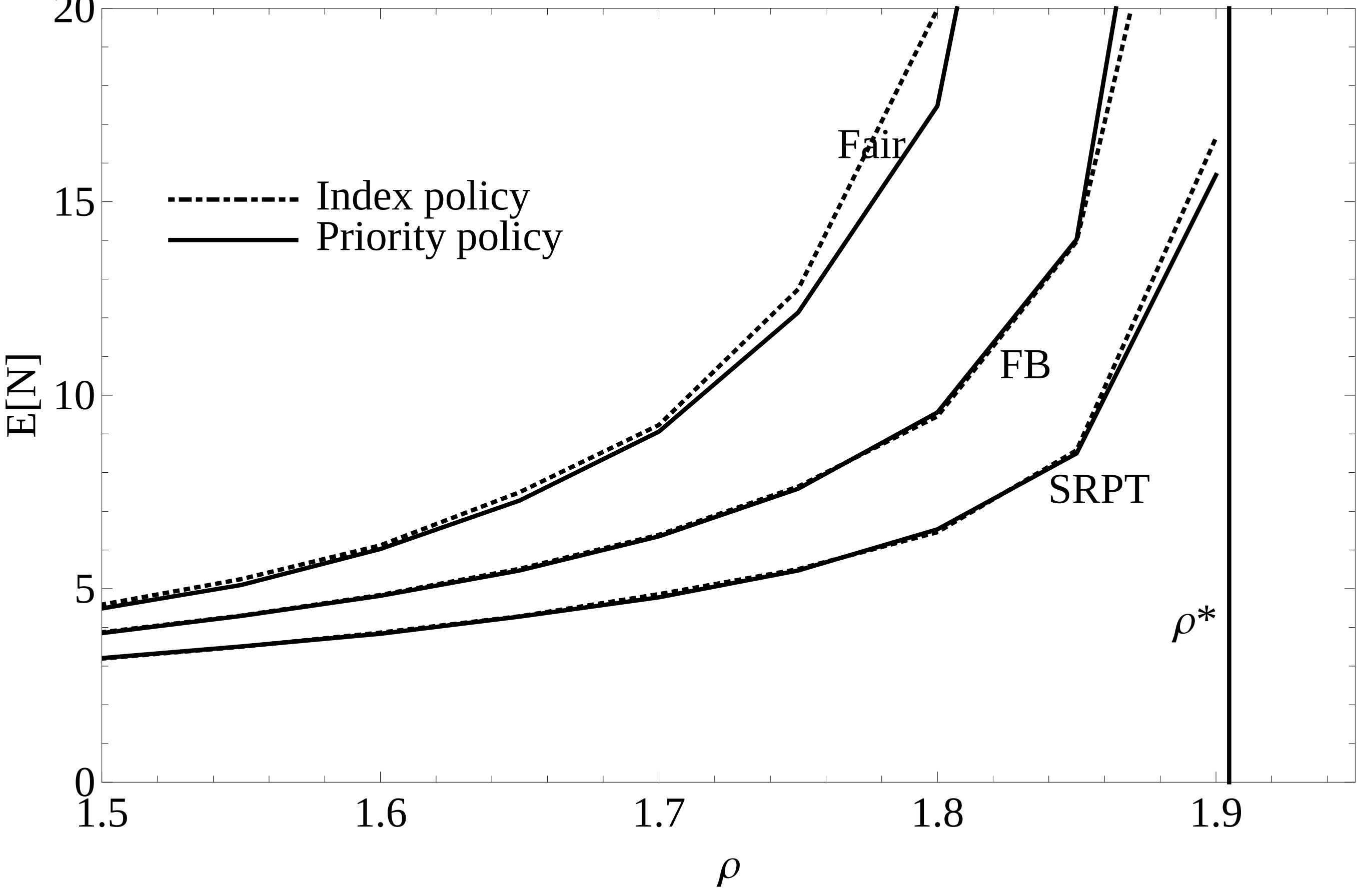


Comments

- All other policies better than PF
- SRPT and FB like policies separate nicely
- Corresponding P-policy better than its TAOS variant

Symmetric case, 2 rates, high variability

- Parameters: $r_1 = 1$, $r_2 = 20$, $p_{11} = 0.5$, $p_{12} = 0.5$

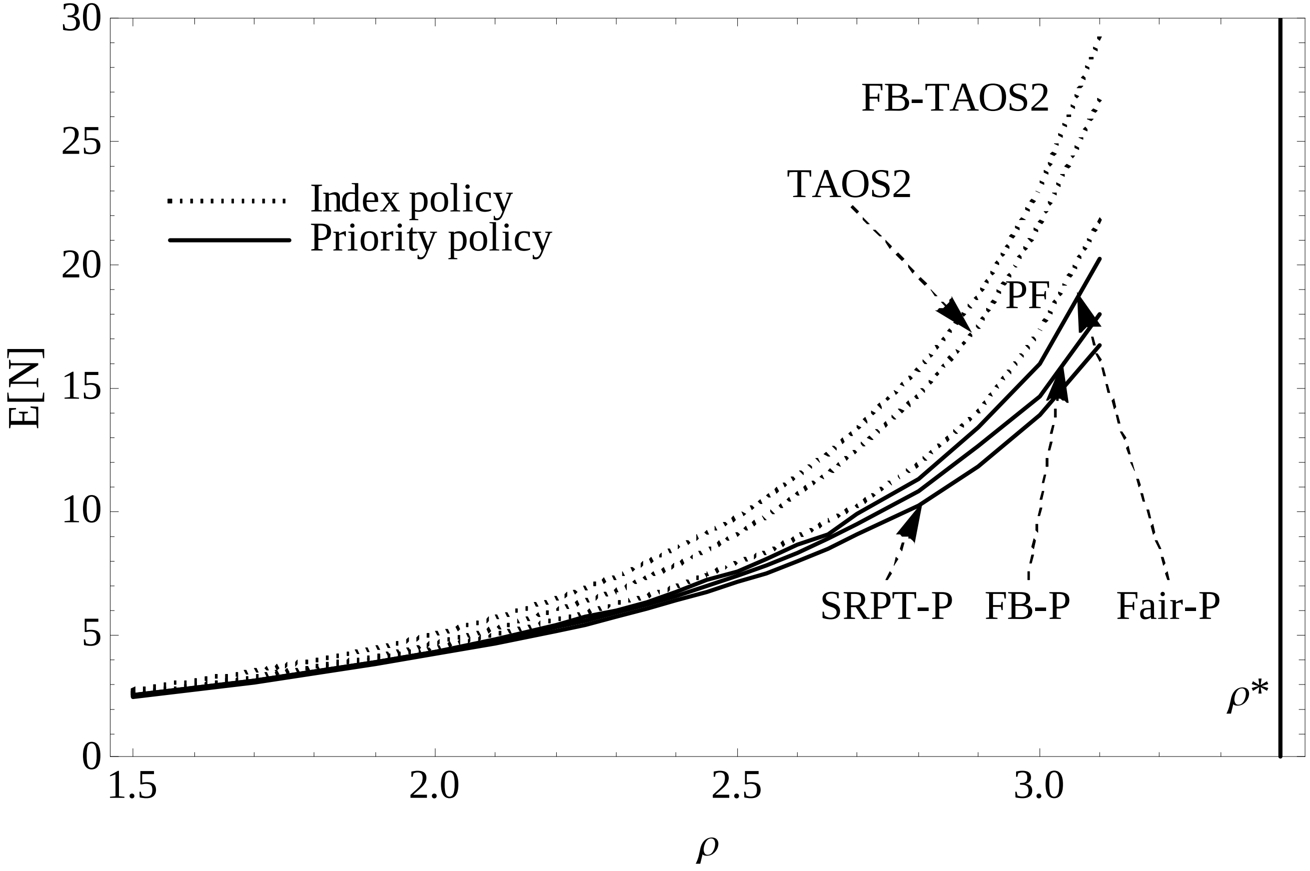


Comments

- All policies better than PF
- SRPT- and FB-like policies separate nicely
- SRPT-P and TAOS2 nearly same
- FB-P and FB-TAOS2 nearly same
- Capacity limit is higher due to higher scheduling gain

Symmetric case, HSDPA rates

- Parameters: uniform distribution for rates



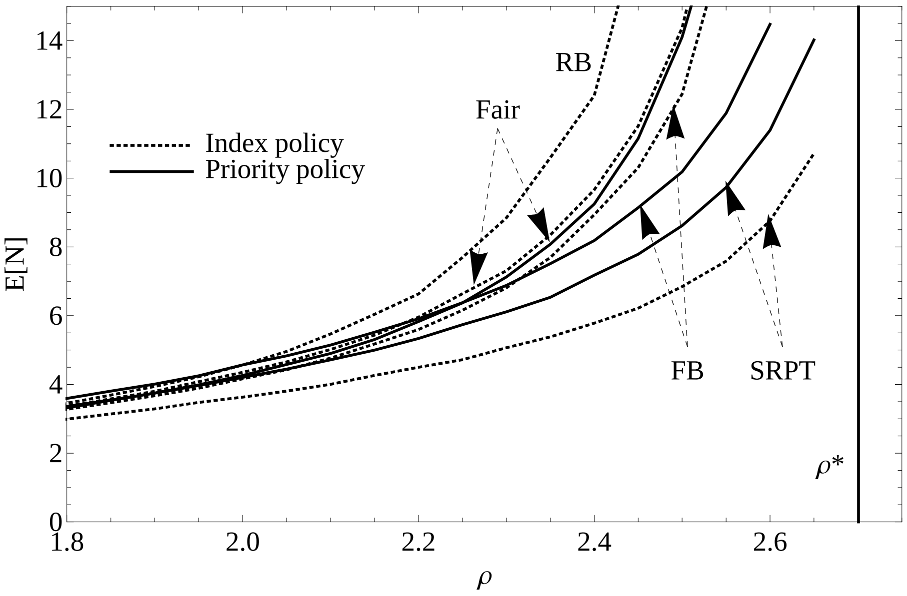
Comments

- Both TAOS policies worse than PF
- P-policies better than PF
- P-policies sorted as expected (SRPT, FB, RR)
- Capacity limit is higher due to higher scheduling gain

Asymmetric cases

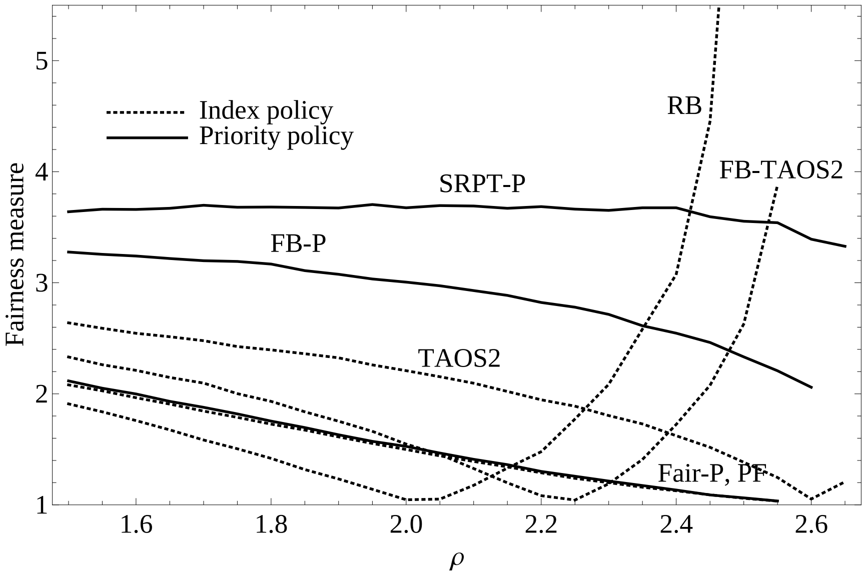
- Much more complex dynamics
  - The different approaches are not systematically anymore better than others (as in symmetric case)
  - Degree of asymmetry is also one arbitrary “parameter”…
- Some observations
  - Depending on the load, one method might be better/worse than another
  - In terms of fairness, the “relative” policies behave differently for low loads and high loads (non-monotonous behavior)

Asymmetric case: 2 rates, high asymmetry



Comments

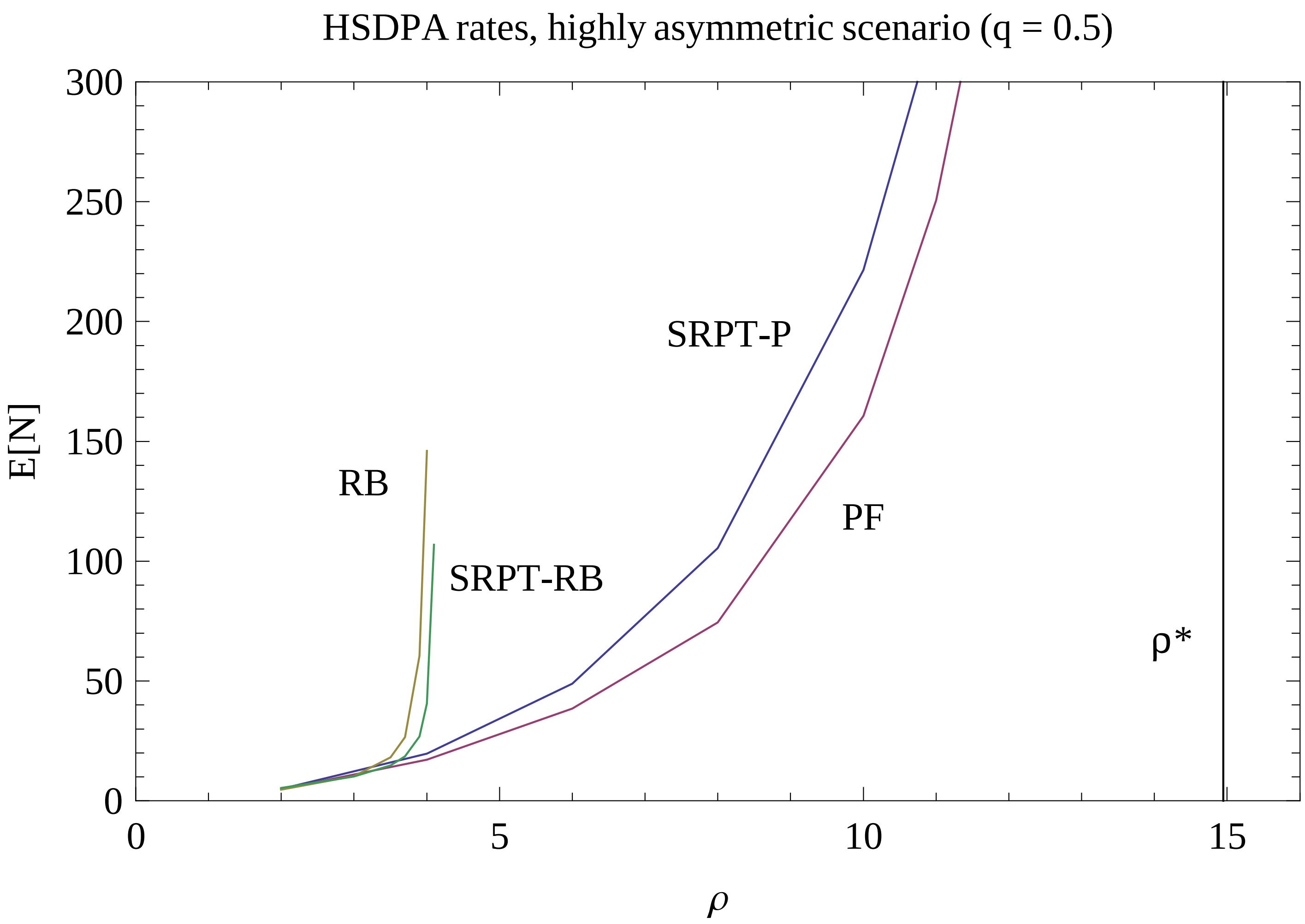
- RB and FB-TAOS2 start rising uncontrollably
- P-policies are quite good



Comments

- Fairness of RB and FB-TAOS2 (and TAOS2) have non-monotonous behavior
- P-policies and PF are quite stable

Asymmetric case: HSDPA rates, high asymmetry



Conclusions

- Flow-level age-based scheduling can give significant performance benefits even in a time varying channel setting

- Approaches in the
  - rate information: priority vs. relative
  - size information: SRPT vs. FB

- Analysis of optimal policy is difficult
  - Based on current analysis SRPT-2L and TAOS2 are quite close to optimal

- Dynamic studies with different policies
  - Symmetric case: tradeoffs are easier to understand
  - Asymmetric case: many interesting phenomena occur as a function of load when comparing the policies, drawing conclusions more difficult