Fluid limits for partial service queues

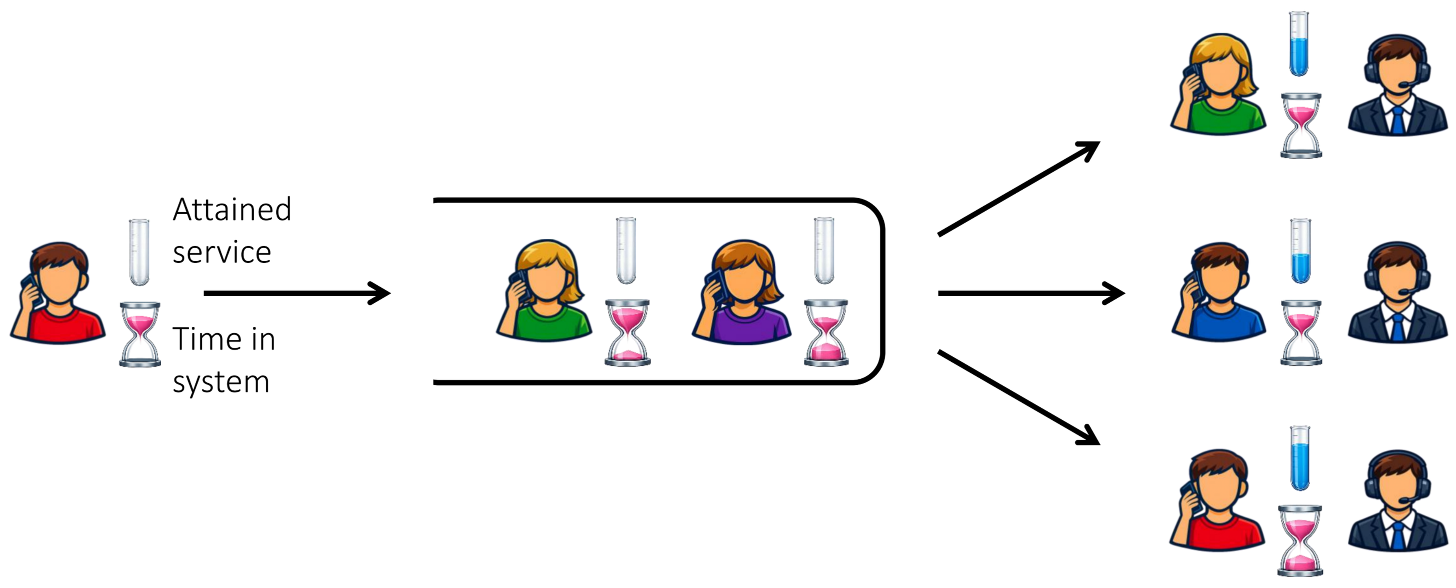
The preemptive LCFS policy and its equivalence to EDF

Diego Goldsztajn Andres Ferragut Fernando Paganini

Universidad ORT Uruguay

Partial service queues

Many server system with reneging customers.



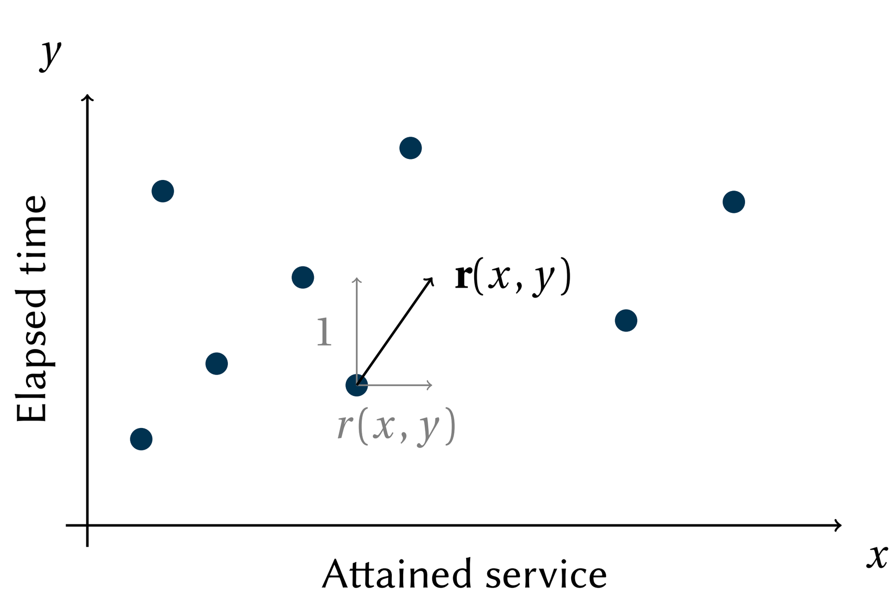
Key differences:

- $\rightsquigarrow$ Users may renege even during service.
- $\rightsquigarrow$ But all work performed is useful (e.g. EV charging).

Assumptions

- $\rightsquigarrow$ Poisson arrivals at rate λ, N servers.
- $\rightsquigarrow$ Service and sojourn times (S,T) with density $g(\sigma,\tau)$.
- $\rightsquigarrow$ We allow for correlated service and sojourns.

System state and dynamics



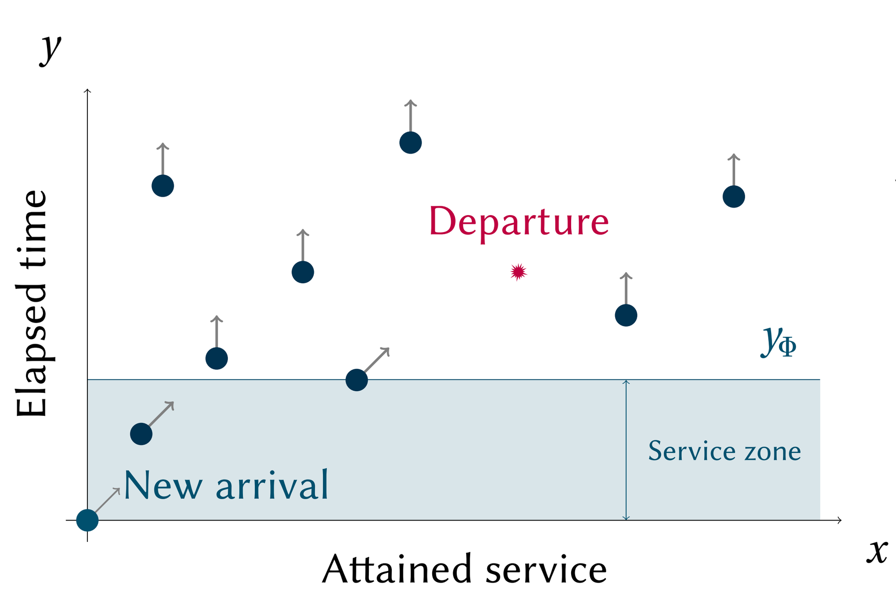
Counting measure:

$$\Phi_t = \sum_i \delta_{(x_i(t), y_i(t))}$$

- $\rightsquigarrow$ x_i = attained service.
- $\rightsquigarrow$ y_i = elapsed time.
- $\rightsquigarrow$ $r(x,y)$ service policy.

$$0 \leqslant r \leqslant 1, \iint r\, d\Phi \leqslant C.$$

LCFS policy



LCFS: serve the tasks with least elapsed time.

- $\rightsquigarrow$ y_Φ: service threshold.
- $\rightsquigarrow$ Age of the C oldest customer.

Hazard rate field $\rightsquigarrow$ controls departures

Let $\bar{G}(\sigma,\tau) = P(S>\sigma, T>\tau)$ and define:

$$\mathbf{h}(x,y) = -\nabla \log \bar{G}(\sigma,\tau),$$

the hazard rate field.

Departure rate at (x,y):

$$\eta(x,y) = \mathbf{h}(x,y)\cdot \mathbf{r}(x,y).$$

Fluid model

Let $f(x,y)$ be a fluid approximation of Φ, then:

$$\frac{\partial f}{\partial t} + \nabla\cdot(\mathbf{r}f) = \lambda\delta_{(0,0)} - f(x,y)\eta(x,y).$$

Advection equation:

- $\rightsquigarrow$ The mass follows the field lines.
- $\rightsquigarrow$ Arrives with no elapsed time or service at rate λ.
- $\rightsquigarrow$ Leaves according to the departure field.

Main Result

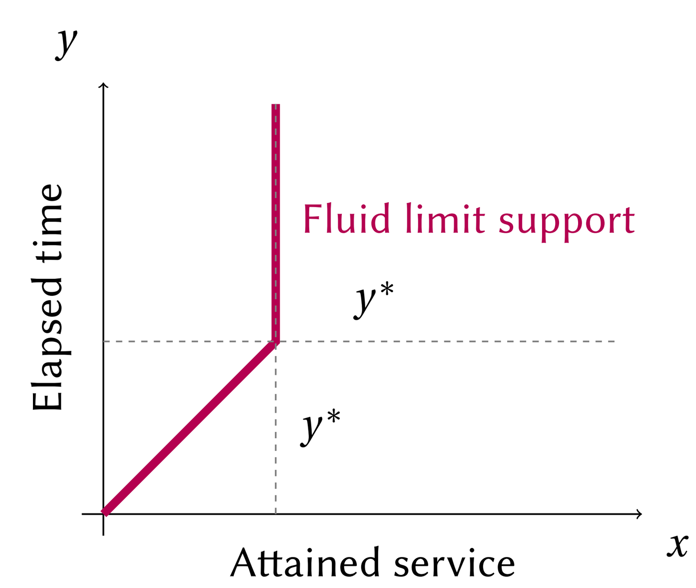
Theorem: In the large scale limit $\lambda \mapsto N\lambda$, $N\to\infty$ the rescaled measure-valued stochastic processes $\Phi_t^{(N)}$ converge weakly to a solution of the advection equation.

We further characterize the equilibrium solution of the advection equation and prove uniqueness under mild hypotheses.

Key proof challenges

- $\rightsquigarrow$ Working in measure-valued space through measure projections against test functions.
- $\rightsquigarrow$ Proving that the martingale compensator involves the hazard rate field.
- $\rightsquigarrow$ Bounding the quadratic variation of the involved martingales.
- $\rightsquigarrow$ Tightness via *Jakubowski criterion*.
- $\rightsquigarrow$ Uniqueness of the weak PDE solution.

Steady-state solution



Degenerate measure in $\mathbb{R}_+^2$

System load:

$$\rho := \lambda E[\min\{S,T\}].$$

Theorem: In overload ($\rho > 1$), y^* is the unique solution of

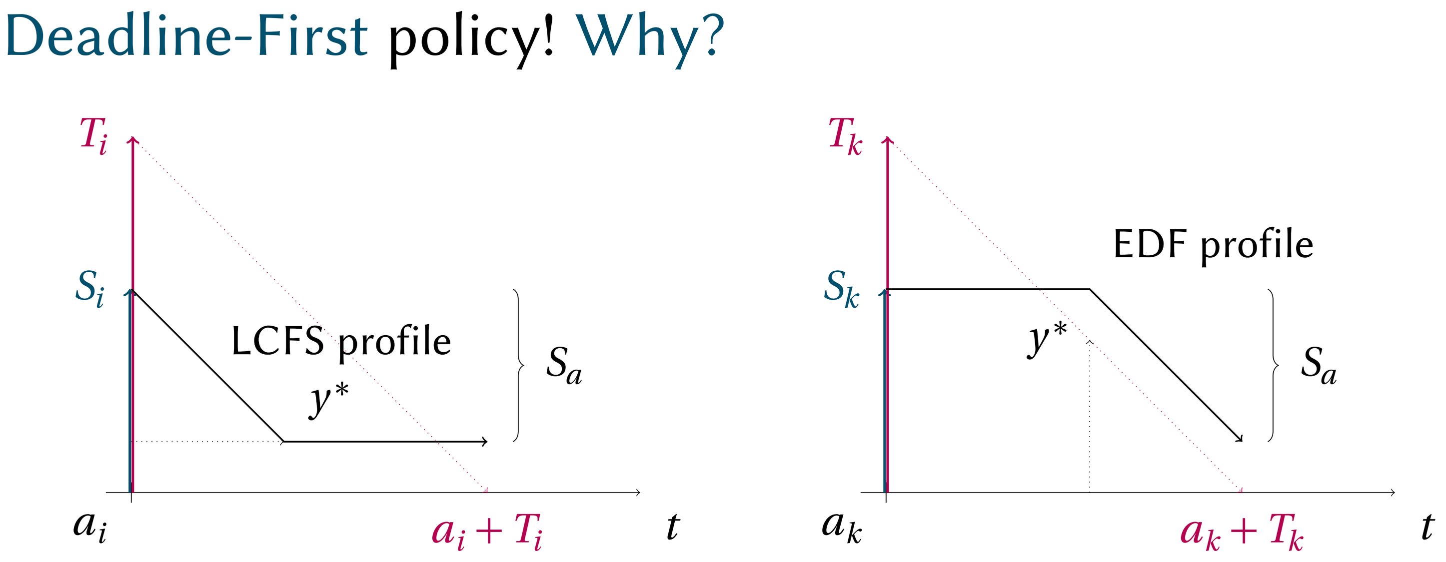
$$\lambda E[\min\{S,T,y^*\}] = 1.$$

Key metric: attained service

Corollary: in the fluid limit, the *attained service* S_a is:

$$S_a = \min\{S,T,y^*\}$$

This is exactly the same attained service of the Earliest-Deadline-First policy! Why?



Key takeaway

Partial service queues model several practical systems. In a large scale limit, the deadline oblivious policy LCFS performs as well as a deadline-aware policy such as EDF. You don't need the deadlines to schedule right!

Contact: ferragut@ort.edu.uy. https://aferragu.github.io