On servers consume power, idle servers consume less power, off servers consume no power. In this talk we consider a server farm with an unbounded number of servers. Upon arrival, a costumer is assigned to an idle server, if any, otherwise a server is switched on instantly and the customer is assigned to that server. There is a lump cost for switching on an off server. Upon customer departure, the system manager has to decide whether to switch off the available server, or leave him idle. He has to make a trade-off between the energy cost due to idle servers and switch on cost of off servers so as to minimise the expected discounted and average cost.

In [1] it is shown that there exists an optimal switching curve determining the region where it is optimal to idle a server upon becoming available, in the case that the service rates are assumed bounded. We will extend this result to the unbounded service rate case. Furthermore, we will provide an efficient algorithm for computing the optimal threshold, both for the discounted and average cost optimality criteria. It turns out that for reasonable parameter assumptions, there even exists a strongly Blackwell optimal switching curve. This will be shown be exploiting the fact that the number of customers in the system behaves as an M/M/∞-queue, independent of the idling policy.

References