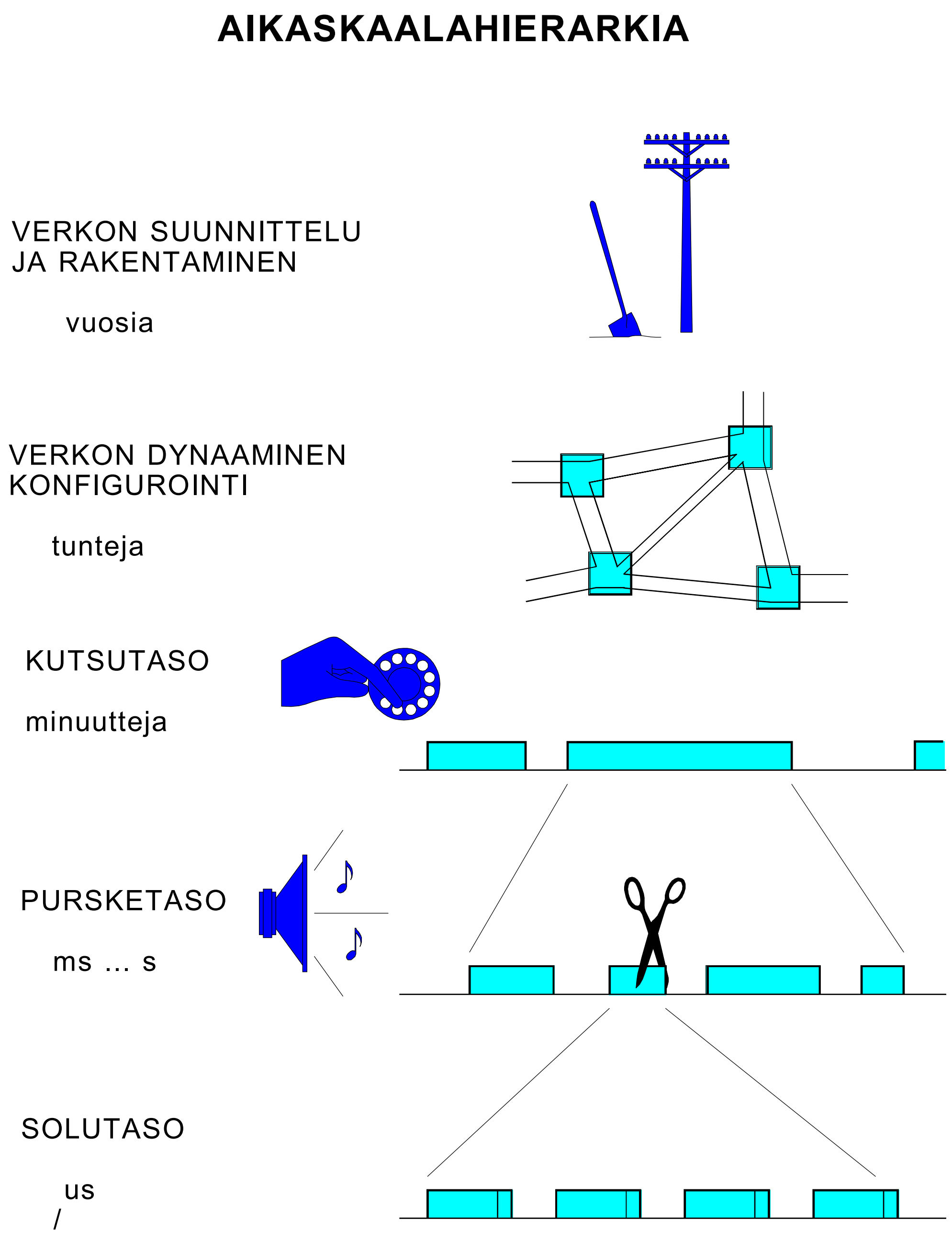
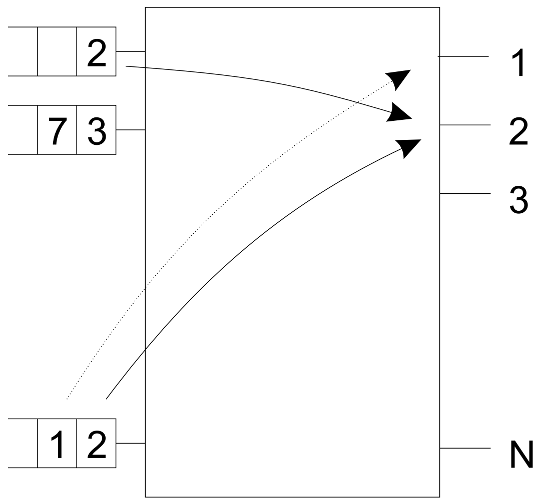
Time Scale Hierarchy of Traffic Problems

- The relevant time scales span a vast range of over 13 decades!
- Each time scale poses of its particular type of problems.
- In the following, we will deal with the three lowest layers, starting from the cell or packet level, going through the burst level up to call of flow level.



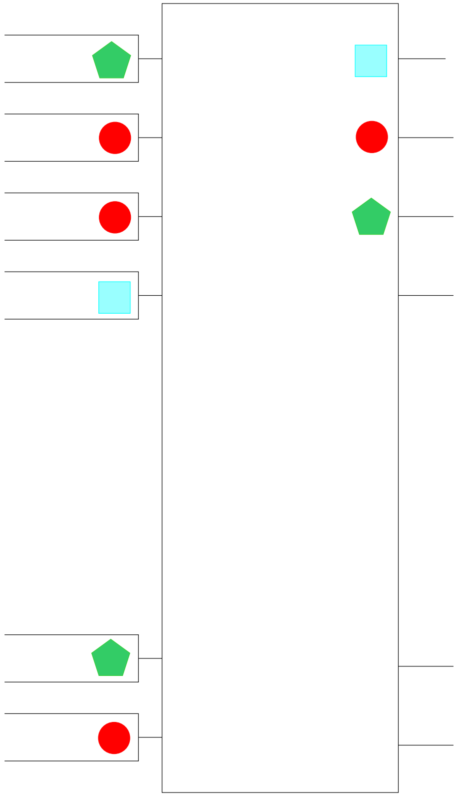
HOL blocking in an input buffered ATM switch



- HOL = Head of Line
- Only one of the cells in HOL position heading to the same output port can be sent, the others have to wait
- Discrete time (slotted time) system
- Time slot = transmission time of a cell
- The destination addresses distributed evenly among the output ports
- Very large offered traffic; input buffers always full
- Throughput per output port p = probability that a randomly chosen slot on the output line is occupied by a cell

- HOL blocking: output port 1 free, but the cell in the lowest input queue destined for output port 1 is blocked by another cell, which cannot be sent because of the contention for output port 2

HOL queues



- HOL queue i comprises of the cells in the HOL position which are destined for output port i.
- In each slot, precisely one cell is sent from each non-empty HOL queue,
  - of course, no cell can be sent from an empty queue.
- In each time slot, $N \cdot p$ cells on the average depart from the switch;
  - when $N \to \infty$, the number of departing cells equals, in relative terms, more and more exactly $N \cdot p$.
- The same number of cells are transferred to the HOL position, i.e. arrive as new cells to the HOL queues.

Distribution of the number of cells arriving at a HOL queue

- Denote $M = N \cdot p$, i.e. $N = M/p$.
- Each of these M cells is intended for output i with the probability $1/N$.
- The number of cells joining queue i obeys the distribution $\text{Bin}(1/N, M)$=$\text{Bin}(p/M, M)$.
- As the size of the switch grows, $N \to \infty$, then also $M \to \infty$.
- At this limit, the number of arriving cells is distributed as Poisson(p).

- Each HOL queue has the same queue length distribution as a continuous time $M/D/1$ queue with load p:
  - the queue length of a continuous time queue can be determined at embedding points separated by one service time D (the distribution at the embedding points is the same as at a random point of time)
  - the queue at the embedding points obeys the rule given before for the discrete time queue: if the queue at the embedding point is non-empty, then precisely one customer will depart until the next embedding point; and if the queue is empty, then no customer will depart before the next embedding point
  - the number of new arrivals from a Poisson process between the embedding points is Poisson distributed with mean $\rho = \lambda D$ (here denoted by p)

Maximum throughput limited by the HOL blocking

- Since each HOL queue behaves as an $M/D/1$ queue with load p, the mean queue length of each is given (by the PK formula)

$$\text{the mean length of a HOL queue} = p + \frac{1}{2}\frac{p^2}{1-p}$$

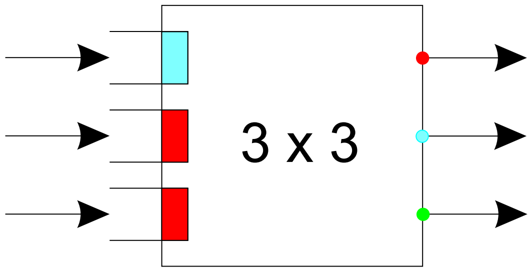
- There are N HOL queues in total (one per each output port).
- In a heavily overloaded system, none of the input buffers is empty; thus the N HOL queues together always fill all the N HOL places. It follows that

$$\boxed{\text{The mean queue lengths of each HOL queue equals 1}}$$

- From this condition one can solve p

$$\begin{aligned} p + \tfrac{1}{2}\tfrac{p^2}{1-p} = 1 \;&\Rightarrow\; \tfrac{1}{2}p^2 = (1-p)^2 \\ &\Rightarrow\; \tfrac{1}{\sqrt{2}}p = 1 - p \\ &\Rightarrow\; p = \tfrac{\sqrt{2}}{1+\sqrt{2}} = \underline{2-\sqrt{2}} \approx \underline{0.586} \end{aligned}$$

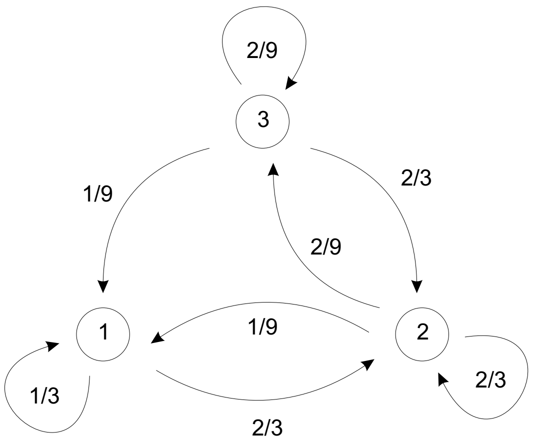
HOL blocking in a finite 3x3 switch



- We can define three states of the HOL cells ('colour' denotes the output port):
  1. all cells have the same colour
  2. cells are of two different colours
  3. cells are of three different colours (all cells have different colour)

- In state 1, only one HOL cell can be forwarded; it is replaced by a new cell, which is of the same colour as the others with the probability 1/3 and of different colour with the probability 2/3.
- In state 2, two cells will be forwarded; they are replaced with two new ones, which have the same colour with the remaining cell with the probability 1/9; all have different colour with the probability 2/9; otherwise, with the probability 6/9, after the replacement the HOL cells are again of two different colours.
- In state 3, all three cells are forwarded and replaced by new ones; these have the same colour with probability 1/9, all have different colour with the probability 2/9, and with the probability 6/9 they are of two different colours.

Throughput of a 3x3 switch (continued)



- The state of the HOL cells constitutes a Markov chain with the state transition diagram shown in the figure.
- The transition probability matrix is

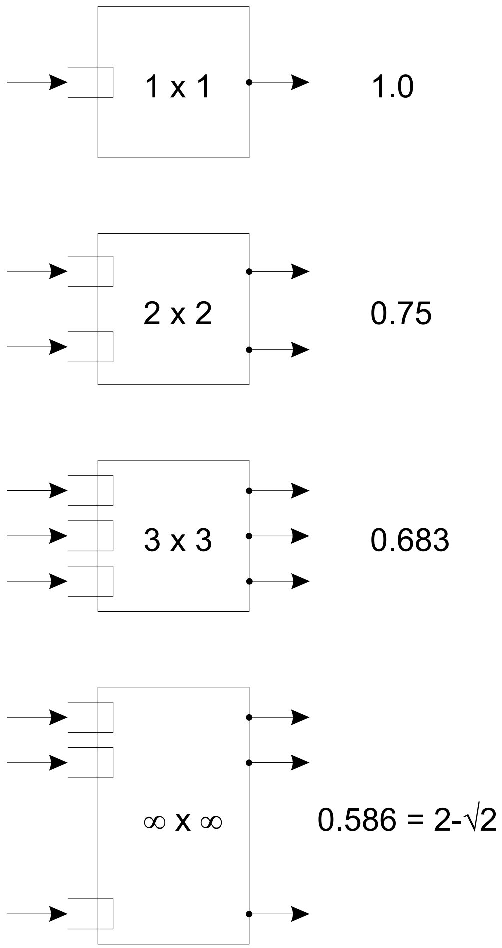
$$\mathbf{P} = \begin{pmatrix} \frac{3}{9} & \frac{6}{9} & 0 \\ \frac{1}{9} & \frac{6}{9} & \frac{2}{9} \\ \frac{1}{9} & \frac{6}{9} & \frac{2}{9} \end{pmatrix}$$

- The equilibrium probability vector $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3)$ can be solved from the balance equation of the Markov chain $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$, i.e.

$$\begin{cases} 9\,\pi_1 = 3\,\pi_1 + \pi_2 + \pi_3 \\ 9\,\pi_2 = 6\,\pi_1 + 6\,\pi_2 + 6\,\pi_3 \\ 9\,\pi_3 = 0 + 2\,\pi_2 + 2\,\pi_3 \end{cases}$$

- The normalized solution is $\boldsymbol{\pi} = (\frac{3}{21}, \frac{14}{21}, \frac{4}{21})$.
- The throughput per output port is $p = \frac{1}{3}(\pi_1 \cdot 1 + \pi_2 \cdot 2 + \pi_3 \cdot 3) = \frac{43}{63} \approx 0.683$.

The throughput limited by the HOL blocking for switches of different sizes



- HOL = Head of Line
- Only one of the cells in HOL position heading to the same output port can be sent, the others have to wait
- Discrete time (slotted time) system
- Time slot = transmission time of a cell
- The destination addresses distributed evenly among the output ports
- Very large offered traffic; input buffers always full
- Throughput per output port p = probability that a randomly chosen slot on the output line is occupied