The Art of Modeling

Renato Lo Cigno

Simulation and Performance Evaluation 2017-18

What is a model?

Given a system, a model is a mathematical law (function) that describe some of its properties as a function of one or more free parameters

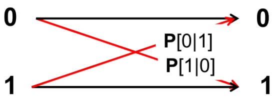
- The Digital Binary Communication Channel (DBCC)
- The bit error probability given the noise on the channel . . . but
  - What is the noise? What modulation is used? What is the "channel"?
- The speed of a car given the power (force, torque) yield by the engine
  - What about frictions, air, gears, . . .
- The number and distribution of arcs in a graph (network) given the "arc generation law"
- The completion time of a job on a specific computer
- The time spent in a bank given the operation I have to do (and the other customers?)

- Deterministic
  - Simple (or complex) equations, e.g., $a = \frac{F}{m}$, $v(t) = \int_t \frac{F(t)}{m}\,dt$
- Stochastic
  - Random Variables ...
- Static (does not depend on time)
  - Deterministic, Stochastic
- Dynamic (depends on time)
  - Deterministic, Stochastic (differential equations, random processes)

- The system

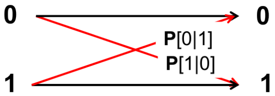
- The system

- The model



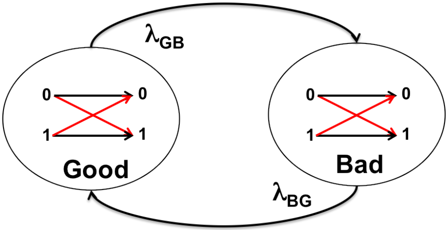
- The system

- The model

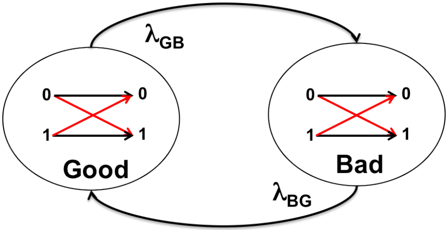


- Characterization only requires **P**[1|0] and **P**[0|1]
- But who give us these parameters?

- DBCC can be easily extended with a Markov Chain to model more complex, non-stationary scenarios



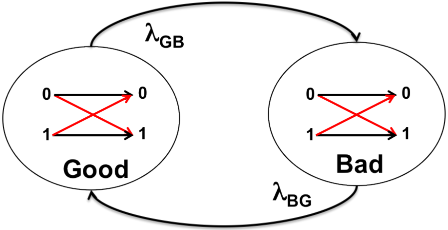
- DBCC can be easily extended with a Markov Chain to model more complex, non-stationary scenarios



- Now we have more parameters to define:

$$\mathbf{P}_G[1|0]; \mathbf{P}_G[0|1]; \quad \mathbf{P}_B[1|0]; \mathbf{P}_B[0|1]; \quad \lambda_{B,G}; \lambda_{B,G}$$

- DBCC can be easily extended with a Markov Chain to model more complex, non-stationary scenarios



- Now we have more parameters to define:

$$\mathbf{P}_G[1|0]; \mathbf{P}_G[0|1]; \quad \mathbf{P}_B[1|0]; \mathbf{P}_B[0|1]; \quad \lambda_{B,G}; \lambda_{B,G}$$

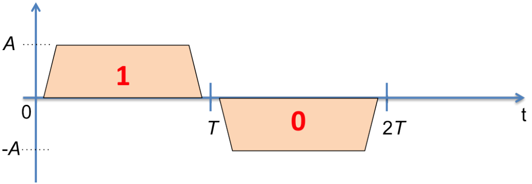
- Yet we do not know how to set these parameters

Bit Error Rate (BER)

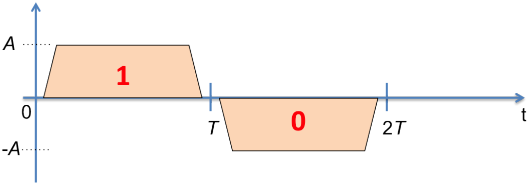
- A pretty simple concept, we need it to tune our DBCCs ...which is what we use as Computer Scientist to design protocols, networks, distributed applications
- It depends on many characteristics of the transmission system
  - Modulation scheme (amplitude, phase, frequency, No. of bits/symbol, ...)
  - The transmission means (copper, fiber, wireless, central frequency, ...)
  - Receiver characteristics
  - Presence and characteristics of error correcting codes
  - ...

- A pretty simple concept, we need it to tune our DBCCs ... which is what we use as Computer Scientist to design protocols, networks, distributed applications
- It depends on many characteristics of the transmission system
  - Modulation scheme (amplitude, phase, frequency, No. of bits/symbol, ...)
  - The transmission means (copper, fiber, wireless, central frequency, ...)
  - Receiver characteristics
  - Presence and characteristics of error correcting codes
  - ...
- **Disclaimer:** this is not meant to be a rigorous analysis of Communication Theory!

- PAM: Pulse Amplitude Modulation:
  1 $\rightarrow$ positive amplitude pulse; 0 $\rightarrow$ negative amplitude pulse
- We use a “reasonable” real waveform $w(t)$ (similar to a square wave) of duration T



- PAM: Pulse Amplitude Modulation:
  1 $\rightarrow$ positive amplitude pulse; 0 $\rightarrow$ negative amplitude pulse
- We use a "reasonable" real waveform $w(t)$ (similar to a square wave) of duration T



- The transmitted energy per bit is

$$E_b^T = \int_0^T Aw(t)\,dt = A$$

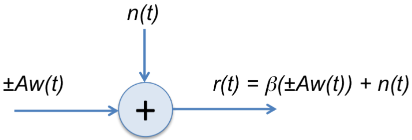
if we assume $w(t)$ energy equal to 1

- Maximum Likelihood Receiver: integrates the received signal over the bit period T and decides based on sign of the integral
  - In practice it evaluates what is the sign of the waveform based on the amount of energy present in the received signal
  - Details are too technical to unfurl here, but in practice we have

$$b_i = \int_{(i-1)T}^{iT} r(t)\, dt$$

where b_i is the i-th bit we decide has been received (1 if $b_i > 0$, 0 if $b_i < 0$), $r(t)$ is the signal received and $w(t)$ is the base waveform

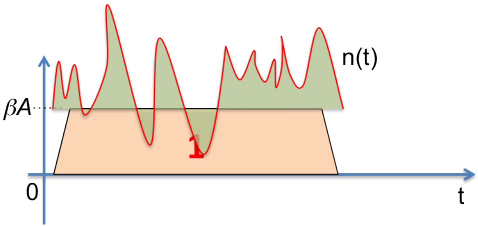
- And the Channel?
- We assume the simplest possible model: only Additive, White (uncorrelated), Gaussian Noise with 0 mean and $\sigma^2 = N_0$; N_0 is called 'spectral noise density'
- An the inevitable attenuation β



- The received useful energy per bit is

$$E_b = \int_0^T \beta A w(t)\, dt = \beta A$$

$$r(t) = \pm\beta A w(t) + n(t)$$



Normalizing so that $t = (i-1)T + t$

$$b_i = \int_0^T \pm\beta A w(t) + n(t)\, dt$$

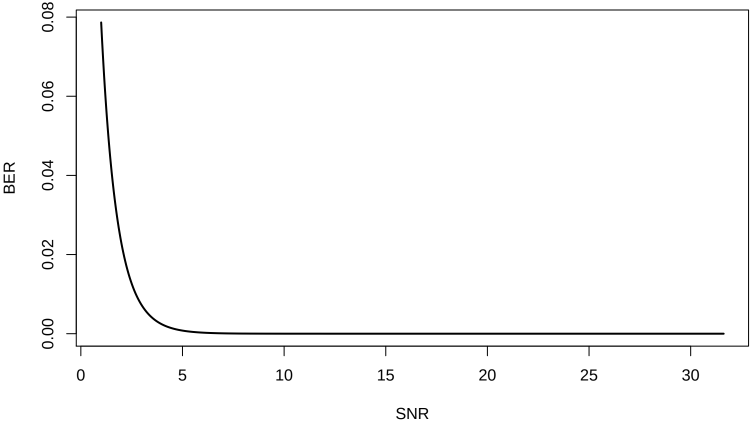
Thanks to the central limit theorem b_i is a Gaussian RV with mean $\pm\beta A = \pm E_b$ and standard deviation $\sigma^2 = N_0$

- Computing the BER reduces to evaluate the probability that a b_i has the wrong sign compared to the transmitted signal, i.e., that a Gaussian RV with $\sigma = N_0$ is larger than $\sqrt{E_b}$

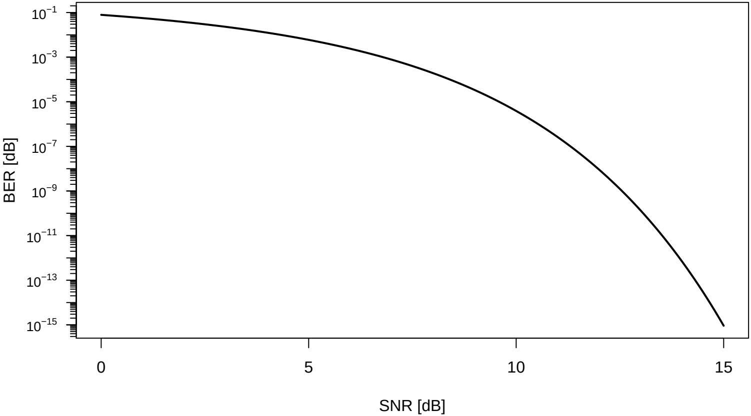
$$\text{BER} = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{E_b/N_0}}^{\infty} e^{-x^2}\, dx = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right)$$

We only have to find a good plot to show its behavior . . .

We only have to find a good plot to show its behavior ...



... not this one ...



. . . much better!

- What do we have to take into account to get a reasonable model?
- The engine power for sure . . . is it enough?
- That in the end is what we mostly know about our car engine . . .
- What is the torque? And what about frictions and air drag?
- Does the gear have influence? And the weight of the car?
- **Let's make some models**

- What do we have to take into account to get a reasonable model?
- The engine power for sure . . . is it enough?
- That in the end is what we mostly know about our car engine . . .
- What is the torque? And what about frictions and air drag?
- Does the gear have influence? And the weight of the car?
- **Let's make some models**

- **Disclaimer:** these are simplifications of Vehicular Technology for Computer Scientists . . .

We want to model the behavior of a vehicle when we go full throttle. We start from high school physics ...

$$\begin{cases} \dot{x} = v \\ \dot{v} = a \end{cases} \tag{1}$$

where x is the position, v is the speed, a is the acceleration

Now we consider three different models for car's acceleration

Model 1

This model assumes constant force (so constant torque) with no RPM limit.

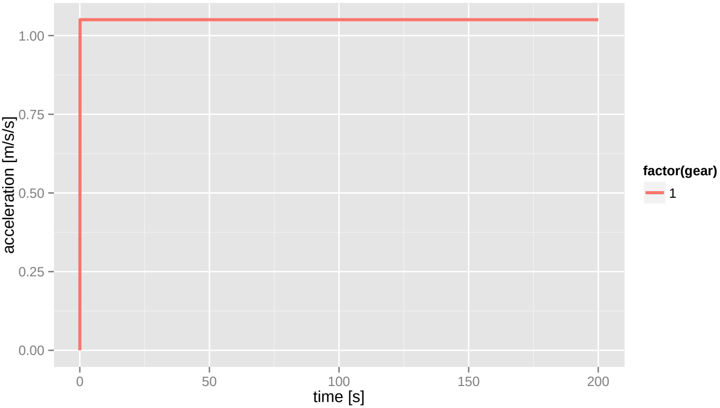
$$\begin{cases} \dot{x} = v \\ \dot{v} = \frac{F_{\text{eng}_1}(r_{\text{gear}})}{m} \end{cases} \tag{2}$$

where F_{eng} is the force generated by the engine, m is the mass of the car, and r_{gear} is the transmission gear ratio. F_{eng} is computed depending on the engine and vehicle parameters. In particular,

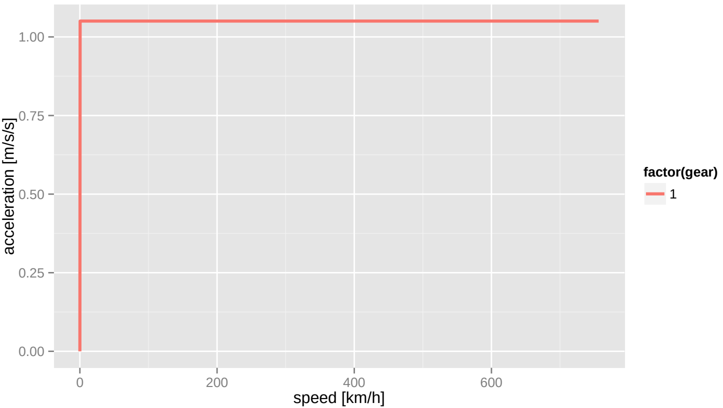
$$F_{\text{eng}_1}(r_{\text{gear}}) = \frac{T \cdot r_{\text{gear}}}{d_{\text{wheel}} \cdot \pi}. \tag{3}$$

T is the torque in Nm, d_{wheel} is the tracting wheels diameter in m. We assume only one gear, and engine RPM limit . . .

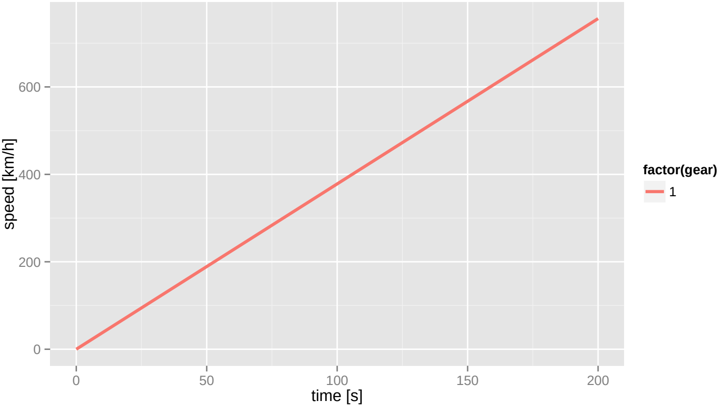
Acceleration versus time



Acceleration versus speed



Speed versus time



This model assumes constant torque, but a maximum number of engine RPM. When we reach this number of RPM, we change gear. In this example, we have four gears.
First we define a function which gives us the engine RPM as function of the speed:

$$RPM(v) = \frac{60 \cdot r_{\text{gear}} \cdot v}{d_{\text{wheel}} \cdot \pi} \tag{4}$$

$$r_{\text{gear}}(v) = \begin{cases} r_1 & \text{if} \quad 0 \leq v < v_1 \\ r_2 & \text{if} \quad v_1 \leq v < v_2 \\ r_3 & \text{if} \quad v_2 \leq v < v_3 \\ r_4 & \text{if} \quad v_3 \leq v \end{cases} \tag{5}$$

$$F_{\text{eng}} = \frac{T \cdot r_{\text{gear}}(v)}{d_{\text{wheel}} \cdot \pi}. \tag{6}$$

$$F_{\text{eng}_2}(v) = \begin{cases} F_{\text{eng}_1}(r_1) & \text{if} \quad 0 \le v < v_1 \\ F_{\text{eng}_1}(r_2) & \text{if} \quad v_1 \le v < v_2 \\ F_{\text{eng}_1}(r_3) & \text{if} \quad v_2 \le v < v_3 \\ F_{\text{eng}_1}(r_4) & \text{if} \quad v_3 \le v < v_4 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

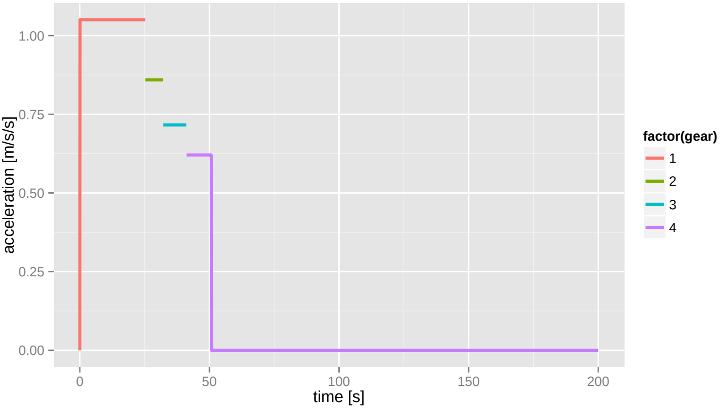
To compute v_i, we can use the following formula which computes the speed of the vehicle given the RPMs and the gear ratio r_i:

$$v_i = \frac{d_{\text{wheel}} \cdot \pi}{60 \cdot r_i \cdot RPM_{\text{max}}} \tag{8}$$

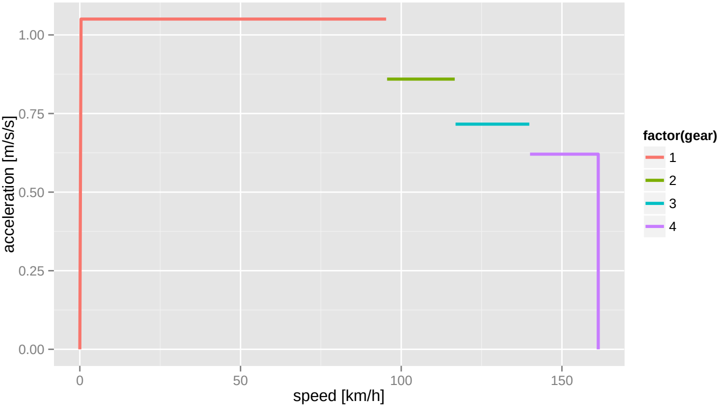
The model now becomes

$$\begin{cases} \dot{x} = v \\ \dot{v} = \dfrac{F_{\text{eng}_2}(v)}{m} \end{cases} \tag{9}$$

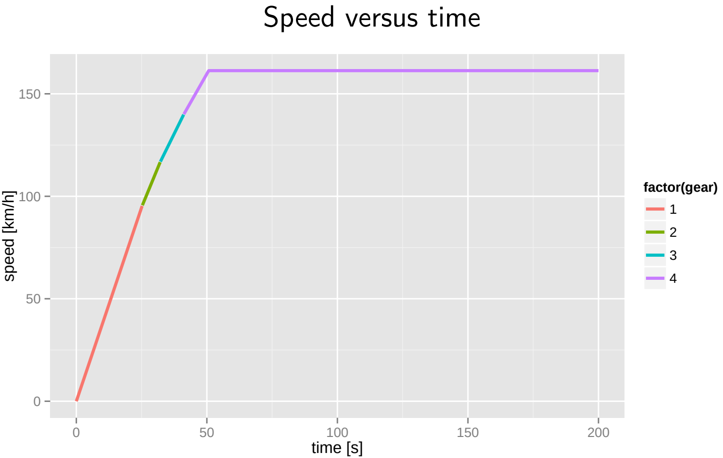
Acceleration versus time



Acceleration versus speed



Model 2


Speed versus time
150
100
50
0
speed [km/h]
0
50
100
150
200
time [s]
factor(gear)
1
2
3
4

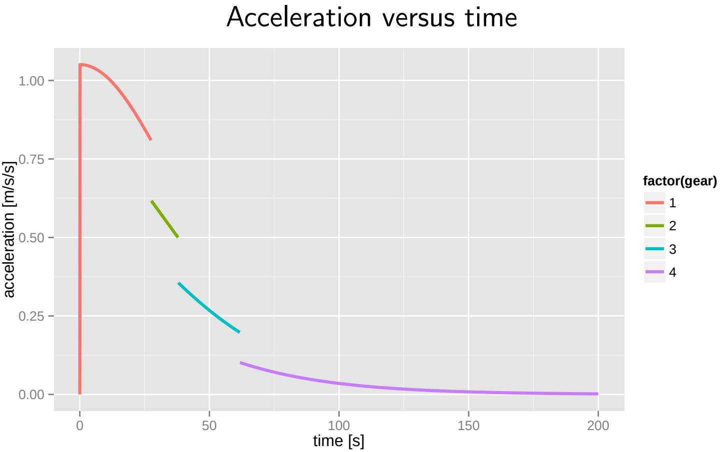
This model assumes the limited RPM engine model, gears, plus air friction

$$\begin{cases} \dot{x} = v \\ \dot{v} = \frac{F_{\mathrm{eng}_2}(v) - F_{\mathrm{air}}(v)}{m} \end{cases} \tag{10}$$

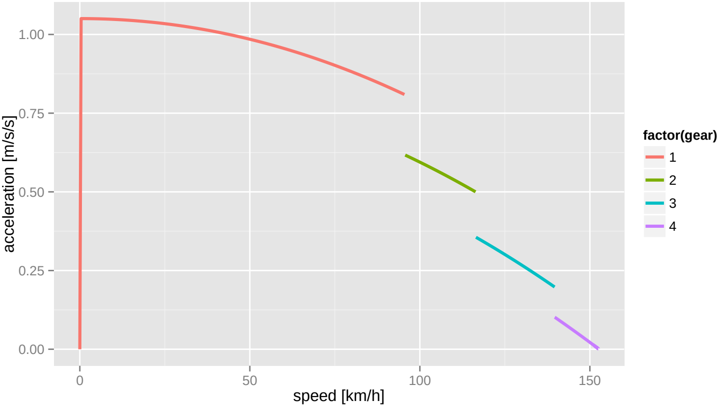
where $F_{\mathrm{air}}(v)$ is the force due to air friction and is defined as

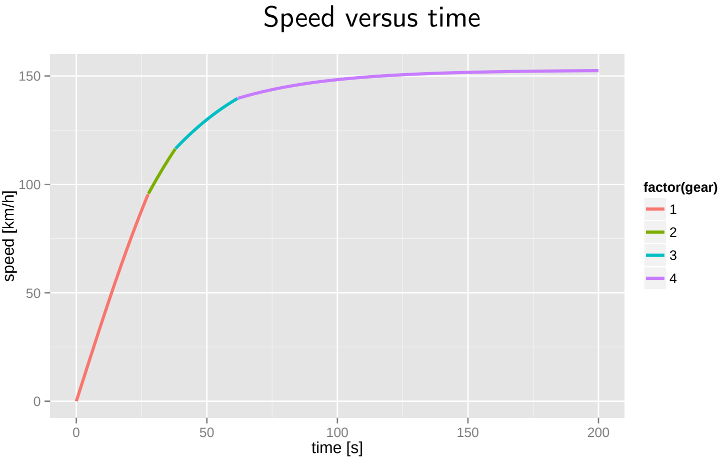
$$F_{\mathrm{air}}(v) = \frac{1}{2} c_{\mathrm{air}} A_L \rho_a v^2 \tag{11}$$

where c_{air} is the drag coefficient, A_L is the maximum vehicle cross section area, ρ_a is the air density, and v the vehicle's speed


Acceleration versus time
acceleration [m/s/s]
1.00
0.75
0.50
0.25
0.00
0
50
100
150
200
time [s]
factor(gear)
1
2
3
4

Acceleration versus speed




Speed versus time
speed [km/h]
150
100
50
0
0
50
100
150
200
time [s]
factor(gear)
1
2
3
4

- The BER model is a static stochastic model
- The Car model is a dynamic (differential equations) deterministic model
- The DBCC model is stochastic, and either static or dynamic depending if there is a single error probability model or if we use a Markov Chain to embed different models ...

- Markov Models are one of the most powerful (yet simple) technique to design models

- We have seen that a Markov Chain (DT or CT) is a simple time-varying SP
- It is a suitable means to model dynamic systems with non-deterministic behavior
- We have to identify a set of variables that represent the state of the system
- We have to identify a set of transition probabilities (rates) that govern the evolution of the system . . .

- We have seen that a Markov Chain (DT or CT) is a simple time-varying SP
- It is a suitable means to model dynamic systems with non-deterministic behavior
- We have to identify a set of variables that represent the state of the system
- We have to identify a set of transition probabilities (rates) that govern the evolution of the system ...
- ... We have to find a method to solve it ...
- ... Or we have to simulate it

- Time is slotted: natural modeling with DT
- Note that slots need not be of the same length, they can depend, e.g., on the state

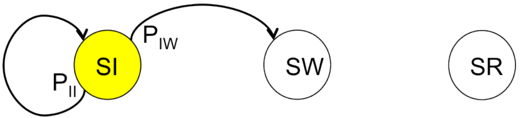
UNIVERSITY OF TRENTO
Department of Information Engineering and Computer Science

- Time is slotted: natural modeling with DT
- Note that slots need not be of the same length, they can depend, e.g., on the state
- The protocol can only be in 3 states:
  - **I**dle: there is nothing to transmit, you can sleep
  - **W**ait: one packet is in transmission, waiting for the acknowledgement
  - **R**e-transmit: a packet has not been ack-ed, we have to re-transmit it
  - $S = \{I, W, R\}$

The States of the Model

- States alone are not enough
- We need the transition probabilities

Transition probabilities from State *I*

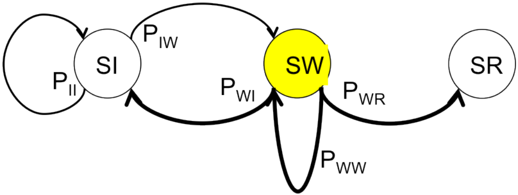


$\mathbf{P}_{II}$ Probability that when Idle no packets arrive

$\mathbf{P}_{IW}$ Probability that when Idle one or more packets arrive

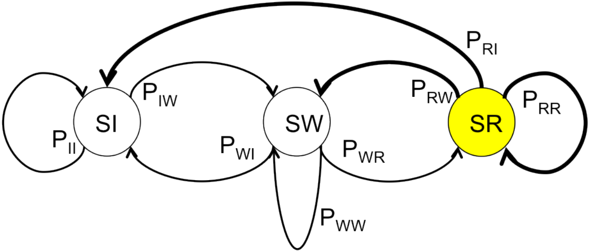
UNIVERSITY OF TRENTO
Department of Information Engineering and Computer Science

Transition probabilities from State *W*



$\mathbf{P}_{WI}$	Probability that the transmission is successful and there are no other packets to transmit
$\mathbf{P}_{WR}$	Probability that the transmission fails the packet must be re-transmitted
$\mathbf{P}_{WW}$	Probability that the transmission is successful and there are other packets to transmit

Transition probabilities from State *R*

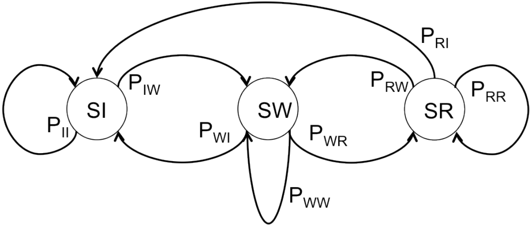


P$_{RI}$ Probability that the re-transmission is successful and there are no other packets to transmit

P$_{RW}$ Probability that the transmission is successful and there are other packets to transmit

P$_{RR}$ Probability that the transmission fails the packet must be re-transmitted (again)

DTMC of the Model



- The slot times include the transmission time and its Ack
- We have external events (arrival of packets from the upper protocol layers that drive the model
- We have complex transitions that account for external arrivals and loss/error probabilities
- We have self-transitions that tells us, e.g., the distribution of the number of re-transmissions per packet