Automata & Queueing Systems

Francesco Casciola

June 13, 2019

Contents

1 Introduction and State Automata

A system with time-driven dynamics is a kind of system in which, even though events might occur, after the occurrences the system doesn't stay in the same state, but variates as the time goes on.

A system with event-driven dynamics is a kind of system whose state variates only with the occurrence of certain events and it is constant in the time between an event and the next one. This produces a piecewise constant function in the time.

Discrete Event System Dynamical system with discrete states and event-driven dynamics.

State Automaton It's a model through which Discrete Event Systems can be represented and it's identified as a 5-tuple $(\mathcal{E}, \mathcal{X}, \Gamma, f, x_0)$ where:

- $\mathcal{E}$ is a discrete set of events.
- $\mathcal{X}$ is a discrete set of states.
- Γ is a function taking values $\mathcal{X} \to 2^{\mathcal{E}}$, where $2^{\mathcal{E}}$ is the 'power set' of the set $\mathcal{E}$ and it represents the set of all the possible subsets of $\mathcal{E}$:

  $$\text{e.g. : } \mathcal{E} = a, b \Rightarrow 2^{\mathcal{E}} = \{\emptyset, \{a\}, \{b\}, \{a, b\}\} \qquad ; \qquad dim(2^{\mathcal{E}}) = 2^{dim(\mathcal{E})}$$

  $\Gamma(x)$ represents the set of events that are possible in the state x.
- f is a function taking values in $\mathcal{X} \times \mathcal{E} \to \mathcal{X}$ and defines the state transitions, such that $x' = f(x, e)$ is the next state when event $e \in \Gamma(x)$ occurs in the current state $x \in \mathcal{X}$.
- $x_0 \in \mathcal{X}$ is the initial state.

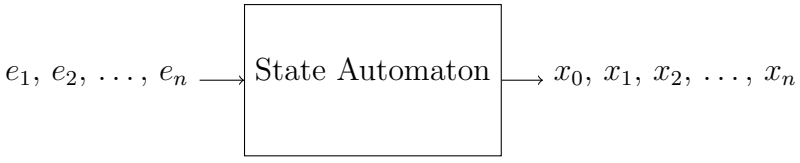

Figure 1: State automaton block diagram.

Please notice the difference between the model and and the actual system: models introduce a certain degree of approximation with respect to the real system.

The concept of feasibility When thinking about a real system there are many events that are possible and many others that aren't. When modelling a system, the events that are extremely improbable (at a level that they can be considered impossible) must be excluded, but there are state-related events that are actually impossible. For instance, a machine which is not working cannot complete a job. A sequence of events $(e_1, e_2, \ldots, e_n)$ is **feasible** (could occur in reality) only if all the events of the sequence occur in states in which they are possible. In other words, the following conditions must hold:

$$e_k \in \Gamma(x_{k-1}), \quad k = 1, 2, \ldots, n$$

$$x_k = f(x_{k-1}, e_k)$$

Unfeasible sequences aren't always obvious, which means that to detect some of them it's either necessary to have a deep knowledge of the system or to run a huge amount of model simulations (assuming that the model is realistic enough).

State Automaton with outputs It's a model through which Discrete Event Systems can be represented and it's identified as a 7-tuple $(\mathcal{E}, \mathcal{X}, \Gamma, f, x_0, \mathcal{Y}, g)$ where:

- $(\mathcal{E}, \mathcal{X}, \Gamma, f, x_0)$ is a state automaton.
- $\mathcal{Y}$ is a discrete set of outputs.
- g is a function taking values in $\mathcal{X} \to \mathcal{Y}$, such that $y = g(x)$ where $y \in \mathcal{Y}$ is the output corresponding to state $x \in \mathcal{X}$.

2 Timed Automata

Concept of time in DES In a real system, when an event occurs, it's also possible to know the time instant when this happens. When trying to introduce the concept of time in a State Automaton model, it's important to remember:

- Time cannot be given as an input to the state automaton, since the inputs are independent of the system while time instants in which events occur depend on the system itself.

This can be demonstrated by considering the execution of jobs on some elements by a given machine as regulated by two different disciplines (while having the time as input): First-In-First-Out (FIFO) and Round-Robin (RR). The FIFO discipline is self-explanatory. The RR is based on the concept of 'time slice', which is the maximum time the machine dedicates to a certain element that needs processing before switching to the next one. In the case in which the time needed to complete the job on the first element is higher than the time slice, the partially processed element will be put back in the queue in last position. Both the disciplines allow the machine to complete the job on all the elements, but, even if their times of arrival are the same, the times in which they are accepted in the system and the ones in which the processing on each single element terminates are different between the two disciplines. This means that the time instants in which the events occur cannot be given as input to the system, since they depend on it.

Timed Automaton A solution to the problem presented in the past paragraph is to use as inputs, instead of the time instants, the duration of processes, at the end of which the events occur. This way, when the system enters a state x in which a given event e is possible, in the model it's possible to start a process of a set duration. This process represents the *event's lifetime* and when the lifetime depletes, the event occurs. Finally, the time instant when e occurs is obtained as sum of the time instants in which the system enters the state x and the lifetime of the event e. This allows us to define the **timed automaton** as a $6-$tuple $(\mathcal{E}, \mathcal{X}, \Gamma, f, x_0, V)$, where:

- $(\mathcal{E}, \mathcal{X}, \Gamma, f, x_0)$ is a state automaton.
- V is the *clock structure* which is an array of 'clock sequences' of the various events:

$$V = \{V_e : e \in \mathcal{E}\}, \quad V_e = \{v_{e,1}, v_{e,2}, v_{e,3}, \dots\}$$

Where V_e is the clock sequence of event e and $v_{e,i}$ is the lifetime of the i-th occurrence of event e. Please note that lifetimes must be $v_{e,i} \geq 0$ for the model to be representative of a real system (since event lifetimes cannot be negative).

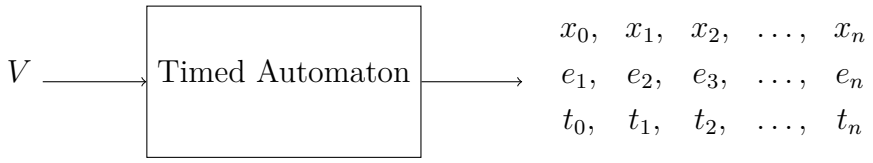

Figure 2: Timed automaton block diagram.

Residual Times According to the definition of a Timed Automaton when entering a state in which an event is possible a process with a certain lifetime starts. Let's consider the situation in which there are two lifetimes, related to two events e_1 and e_2, where $V_{e1} < V_{e2}$ that start when the system enters the state x_{k-1}. The event e_1 will occur first and the system will enter state x_k. If event e_2 is still possible in state x_k then, instead of starting a new process with its lifetime, the 'residual lifetime' $y_{e2} = V_{e2} - V_{e1}$ is employed. If event e_2 is not possible there are two options: either dropping the current lifetime in order to start a new process when the event becomes possible again or keeping the residual lifetime in order to reuse it when the system enters a state in which e_2 is possible again. The choice between these options, in the model, depends on the behaviour of the system.

Notation for Timed Automata The 'score' of an event e at the time t, denoted $n_e(t)$, is the number of lifetimes of the event e used in the interval $[t_0, t]$ (where t_0 is the initial time of the system). From now on the following notation will be adopted:

- With respect to event occurrences:
  - k is the event index ($k = 1, 2, 3, \ldots$).
  - e_k is the k-th event.
  - x_k is the state reached after e_k occurs.
  - t_k is the time when the e_k occurs.
  - $n_{e,k}$ is score of the event e after the k-th event.
  - $y_{e,k}$ is the residual lifetime of e after the k-th event.
- With respect to time:
  - t is the continuous time.
  - $n_e(t)$ is score of the event e after the time t.
  - $x(t)$ is the state of the system at time t.

General algorithm for event timing The algorithm works only when the following assumptions hold:

1. When an event e doesn't occur and it's not possible in the next state, its residual lifetime is ignored and the next time event e becomes possible a new total lifetime is taken from the corresponding clock sequence.

2. When event e occurs, the next time it becomes possible a new total lifetime is taken from the corresponding clock sequence.

3. If the event e doesn't occur and it's still possible in the next state, then its residual lifetime is used.

Under these assumptions, the algorithm is composed by the following steps:

0. **Initialization:** for all the events $e \in \mathcal{E}$, if $e \in \Gamma(x_0)$ we consider $y_{e,0} = v_{e,1}$ and $n_{e,0} = 1$. If $e \notin \Gamma(x_0)$, $y_{e,0}$ is undefined and $n_{e,0} = 0$.

1. **Selection of the next event:** the next event is the one with the smallest residual lifetime.

$$e_k = \underset{e \in \Gamma(x_{k-1})}{\arg\min} \; (y_{e,k-1}) = \arg(y^*_{k-1})$$

2. **Determination of the time instant of the next event**

$$t_k = t_{k-1} + y^*_{k-1}$$

3. **State update**

$$x_k = f(x_{k-1}, e_k)$$

4. **Score update:** for all the events $e \in \mathcal{E}$.

$$n_{e,k} = \begin{cases} n_{e,k-1} + 1 & \text{if a new total lifetime is used (asm. 1, asm 2)} \\ n_{e,k-1} & \text{if the residual lifetime is used (asm. 3)} \end{cases}$$

5. **Update of the residual lifetimes:** for all the events $e \in \mathcal{E}$.

$$y_{e,k} = \begin{cases} v_{e,n_{e,k}} & \text{if } [(e \notin \Gamma(x_{k-1}) \wedge e \in \Gamma(x_k)) \vee (e = e_k \wedge e \in \Gamma(x_k))] \\ y_{e,k-1} - y^*_{k-1} & \text{if } [e \in \Gamma(x_{k-1}) \wedge e \neq e_k \wedge e \in \Gamma(x_k)] \end{cases}$$

6. **Increase by one the value of variable k and go to step 1.**

3 Stochastic Timed Automata

Concept of ubiquitous uncertainty In real systems there might be a degree of uncertainty which must be accounted for in the system's models. Let's suppose having two machines which operate in parallel: how can one determine if they are both available, which one will start working when a piece arrives? Here's a list of the elements in the Timed automaton which are subject to uncertainty:

- f: The example just described is a case of uncertainty in the state transition function.
- x_0: Let's consider a shop, if it opens at a given time t_0 and there are some people waiting for it to open, how long is the queue (state) at t_0?
- V: The processing times can vary depending on the request, it's not always possible to know them in advance.

The need to introduce elements of uncertainty in the model bring to the definition of the next kind of state automaton.

Stochastic Timed Automaton It's a model through which Discrete Event Systems can be represented, when the elements of uncertainty must be taken into account, and it's identified as a 6-tuple $(\mathcal{E}, \mathcal{X}, \Gamma, P, p_0, F)$ where:

- $(\mathcal{E}, \mathcal{X}, \Gamma)$ are the same as for the timed automaton.
- p is a set of transition probabilities from a state to another. It substitutes f and it's defined as follows:

$$\mathsf{p}(x' \mid x, e) = P\left(X_{k+1} = x' \mid X_k = x,\, E_{k+1} = e\right), \qquad \forall e \in \Gamma(x), \quad \forall x, x' \in \mathcal{X}$$

  This set of probabilities generalises the deterministic case, in fact if we have:

$$x' = f(x, e) \qquad \Rightarrow \qquad P(x' \mid x, e) = 1$$

- p_0 is a discrete random variable which defines the initial state probabilities:

$$\mathsf{p}_0(x) = P(x_0 = x), \quad \forall x \in \mathcal{X}$$

- F is the stochastic clock structure and F_e are the cumulative distribution functions of the lifetimes of event e:

$$F = \{F_e : e \in \mathcal{E}\}, \quad F_e(t) = P(V_{e,i} \leq t)$$

We consider stochastic clock structure satisfying three assumptions:

- The random variables $V_{e,i}$ are independent;
- The lifetimes $V_{e,i}$ of the same event are identically distributed;
- Lifetimes of *different* events are independent.

Exponential Distribution When computing probabilities (like the probability of reaching a given state within a certain amount of time), since the lifetimes are random variables, it's normally possible to know only the distributions of the total lifetimes, but not the one of the ones of the residual lifetimes. An exception is the one of exponential distribution which has some helpful properties that will be observed soon. The exponential distribution $X \sim \text{Exp}(1/\lambda)$ (where $1/\lambda$ is the location parameter and λ is the rate) is defined as follows:

$$F_X(t) = P\left(X \leq t\right) = \begin{cases} 1 - e^{-\lambda t} & \text{if } t \geq 0 \\ 0 & \text{otherwise} \end{cases} \qquad f_X(t) = \frac{dF_X(t)}{dt} = \begin{cases} \lambda e^{-\lambda t} & \text{if } t \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Where $F_X(t)$ is the CDF and $f_X(t)$ is the pdf. The aforementioned properties are the following:

- **Memoryless property**: If the time X between the occurrences of a given event is modelled through an exponential distribution and at the time t the event hasn't occurred yet, the probability of the occurrence of the event (computed at a time $s > t$) does not depend t. In formulae:

$$P(X > t + s \mid X > t) = P(X > s)$$

Proof:

$$P(X > t + s \mid X > t) = \frac{P(X > t + s \ , \ X > t)}{P(X > t)} = \frac{P(X > t + s)}{P(X > t)} =$$

$$= \frac{1 - P(X \leq t + s)}{1 - P(X \leq t)} = \frac{1 - F_X(t + s)}{1 - F_X(t)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} =$$

$$= 1 - P(X \leq s) = P(X > s)$$

■

- **Extended memoryless property**: The memoryless property can be defined in a more generic way by considering, instead of the time t, a generic time distribution Y, with support in $[0, \infty)$, independent from X:

$$P(X > Y + s \mid X > Y) = P(X > s)$$

Proof:

$$P(X > Y + s \mid X > Y) = \frac{P(X > Y + s\ ,\ X > Y)}{P(X > Y)} = \frac{P(X > Y + s)}{P(X > Y)}$$

Let's compute the numerator: $P(X > Y + s)$ is equal to the highlighted area A in Figure 3 where the value of X is greater than $Y + s$.

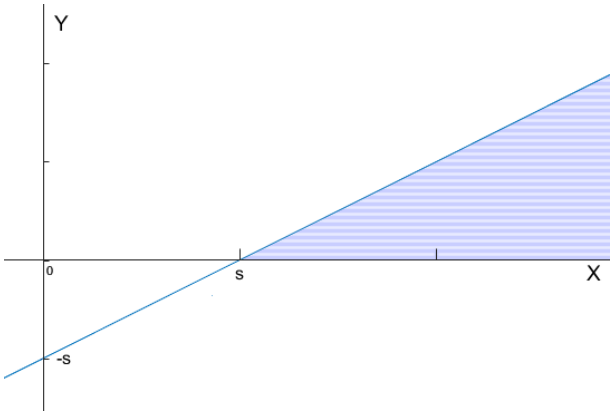

Figure 3: The highlighted area A is the one for which $X - Y > s$

This consideration allows us to proceed as follows:

$$P(X > Y + s) = P(X, Y \in A) = \int_0^{+\infty} dy \int_{y+s}^{+\infty} f_Y(y) \lambda e^{-\lambda x} dx =$$

$$= \int_0^{+\infty} f_Y(y) dy \int_{y+s}^{+\infty} \lambda e^{-\lambda x} dx = \int_0^{+\infty} f_Y(y) \left[-e^{-\lambda x} \right]_{y+s}^{+\infty} dy =$$

$$= \int_0^{+\infty} f_Y(y) e^{-\lambda y} e^{-\lambda s} dy = e^{-\lambda s} \int_0^{+\infty} f_Y(y) e^{-\lambda y} dy$$

Since the s in the lower limit of integration in the innermost integral produces only a term $e^{-\lambda s}$ which can be put outside the integral, it's clear that

$$\int_0^{+\infty} f_Y(y) e^{-\lambda y} dy = P(X > Y)$$

Now it is easy to compute $P(X > Y + s \mid X > Y)$:

$$\frac{P(X > Y + s)}{P(X > Y)} = \frac{e^{-\lambda s} \int_0^{+\infty} f_Y(y) e^{-\lambda y} dy}{\int_0^{+\infty} f_Y(y) e^{-\lambda y} dy} = e^{-\lambda s} = 1 - P(X \leq s) = P(X > s)$$

■

- **Superposition property**: Let's consider the case where it's required to know, for independent events with exponentially distributed lifetimes $\left(X_i \sim Exp\left(\frac{1}{\lambda_i}\right)\right)$, the one that occurs first. The random variable that must be taken into account is:

$$X = \min_{i=1,2,\ldots,n} \{X_i\}$$

The random variable X is exponentially distributed with rate

$$\lambda' = \sum_{i=1}^{n} \lambda_i$$

Proof:

$$P(X \leq t) = 1 - P(X > t) = 1 - P(\min_{i=1,\ldots,n} X_i > t)$$

From this last result, considering both the independence of the distributions and the fact that if the value $\min_i \{X_i\}$ is greater than t then all the X_i are, it's possible to proceed with the computation as follows:

$$1 - P(\min_{i=1,\ldots,n} X_i > t) = 1 - \prod_{i=1}^{n} P(X_i > t) = 1 - \prod_{i=1}^{n} 1 - P(X_i \leq t) =$$

$$= 1 - \prod_{i=1}^{n} e^{-\lambda_i t} = 1 - e^{-\sum_i \lambda_i t} = 1 - e^{-t\lambda'}$$

Where $\lambda' = \sum_{i=1}^{n} \lambda_i$.

■

In addition to these properties, a 'useful computation' (for the next topics) will be executed. Given two exponentially distributed random variables $X \sim Exp\left(\frac{1}{\lambda}\right), Y \sim Exp\left(\frac{1}{\mu}\right)$, let's compute the probability $P(X \leq Y + s)$:

$$P(X \leq Y + s) = \int_0^{+\infty} dy \int_0^{y+s} \lambda e^{-\lambda x} \mu e^{-\mu y} dx = \int_0^{+\infty} \mu e^{-\mu y} dy \int_0^{y+s} \lambda e^{-\lambda x} dx =$$

$$= \int_0^{+\infty} \mu e^{-\mu y} \left[-e^{-\lambda x}\right]_0^{y+s} dy = \int_0^{+\infty} \mu e^{-\mu y} \left[1 - e^{-\lambda(y+s)}\right] dy =$$

$$= \int_0^{+\infty} \mu e^{-\mu y} dy - e^{-\lambda s} \int_0^{+\infty} \mu e^{-\mu y} e^{-\lambda y} dy$$

$$= \left[-e^{-\mu y} + \frac{\mu e^{-\lambda s}}{\lambda + \mu} e^{-(\lambda+\mu)y}\right]_0^{+\infty} = 1 - \frac{\mu e^{-\lambda s}}{\lambda + \mu}$$

Setting $s = 0$ yields a closed-form formula for computing $P(X \leq Y)$ (skipping the entire chain of integration) which corresponds to the probability of lifetime X being less than lifetime Y.

$$P(X \leq Y) = \frac{\lambda}{\lambda + \mu} \tag{1}$$

In the next chapter this last result will be used a lot, so it's important to keep it in mind.

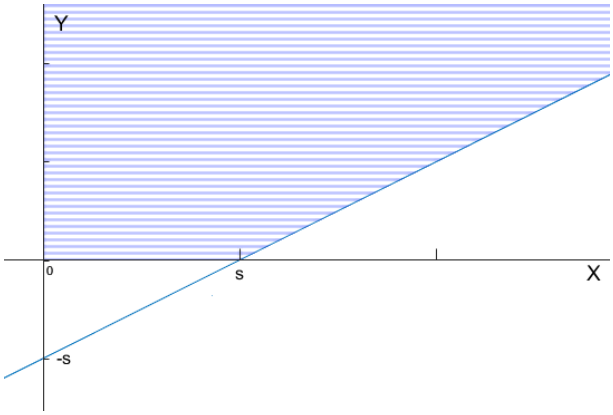

Figure 4: The highlighted area is the one for which $X - Y \leq s$

4 Stochastic Timed Automata With Poisson Clock Structure

Poisson Counting Processes It's a process which counts the occurrences of an event which is always possible. In particular, the 'interarrival times' T_i between any two occurrences of the same event are *i.i.d.* and their distribution is:

$$T_i \sim Exp\left(\frac{1}{\lambda}\right), \quad \lambda > 0$$

A Poisson counting process is defined as a discrete random variable $N_e(t, t+s)$ (it is a random variable because the interarrival times are, indeed, random variables) representing the number of occurrences of the event e over the interval $(t, t+s]$. Given that the interarrival times are exponentially distributed, the probability mass function (pmf) of the Poisson counting process is:

$$P\left(N_e(t, t+s) = n\right), \qquad n = 0, 1, 2, \dots$$

$$P\left(N_e(t, t+s) = n\right) = \frac{(\lambda s)^n}{n!} e^{-\lambda s}$$

Quite evidently, the pmf depends only on s (due to the memoryless property), so:

$$P\left(N_e(s) = n\right) = \frac{(\lambda s)^n}{n!} e^{-\lambda s} \tag{2}$$

Finally, it possible to compute the probability that at least n occurrences of the event e fall in the interval s:

$$P\left(N_e(s) \geq n\right) = 1 - \sum_{m=0}^{n-1} P\left(N_e(s) = m\right) = 1 - \sum_{m=0}^{n-1} \frac{(\lambda s)^m}{m!} e^{-\lambda s}$$

The first equality arises from the fact that the complement of $(n, n+1, n+2, \dots)$ is $(n-1, n-2, \dots)$. This is equivalent to comparing s to the sum of n interarrival times T_i, coming from the same exponential distribution. Thus:

$$P(T_1 + \cdots + T_n \leq s) = 1 - \sum_{m=0}^{n-1} \frac{(\lambda s)^m}{m!} e^{-\lambda s}$$

Stochastic Timed Automaton With Poisson Clock Structure It's a stochastic timed automaton $(\mathcal{E}, \mathcal{X}, \Gamma, \mathsf{p}, \mathsf{p}_0, F)$ where $F = \{F_e : e \in \mathcal{E}\}$, **with *i.i.d.* events' lifetimes** and:

$$F_e(t) = 1 - e^{-\lambda_e t}, \quad t \geq 0, \quad \lambda_e > 0$$

It's important to notice that there is no constraint about whether the events are always possible or not, this means that the 'Poisson clock structure' doesn't refer to a Poisson counting process but to the fact that all the events have exponential interarrival times. *Moreover, for stochastic timed automata with Poisson clock structure, the residual lifetimes of the events follow the same distribution of the corresponding total lifetimes.* This last property can be proved through induction. The actual demonstration can be considered as homework by the reader.

In stochastic timed automata with Poisson clock structure it's quite straightforward (compared to other models) to compute useful probabilities via closed-form formulae, like:

1. $P(E_{k+1} = e \mid X_k = x)$ which is the probability that the next event E_{k+1} (uppercase, since it's a random variable) will be e, given that the current state is x:

$$P(E_{k+1} = e \mid X_k = x) = P\left(Y_{e,k} < \min_{\left[\begin{smallmatrix} e' \in \Gamma(x) \\ e' \neq e \end{smallmatrix}\right]} \{Y_{e',k}\} \right)$$

   Where the Y random variables are the residual lifetimes of the feasible events in state x. Since the residual lifetimes follow exponential distribution, it's possible to use Equation 1 and the superposition property to proceed with the computation:

$$P\left(Y_{e,k} < \min_{\left[\begin{smallmatrix} e' \in \Gamma(x) \\ e' \neq e \end{smallmatrix}\right]} \{Y_{e',k}\} \right) = \frac{\lambda_e}{\lambda_e + (\Lambda(x) - \lambda_e)} = \frac{\lambda_e}{\Lambda(x)} \tag{3}$$

   Where $(\Lambda(x) - \lambda_e)$ is the rate of $Y_{e',k}$ since $\Lambda(x)$ is the sum of the rates of all the feasible events $e \neq e'$:

$$\Lambda(x) = \sum_{e \in \Gamma(x)} \lambda_e$$

2. $P(X_{k+1} = x' \mid X_k = x)$ which is the probability that the next state is x', given that the current state is x:

$$P(X_{k+1} = x' \mid X_k = x) \overset{\text{total probability rule}}{=}$$

$$= \sum_{e \in \Gamma(x)} \left[P\left(X_{k+1} = x' \mid X_k = x,\, E_{k+1} = e\right) \cdot P\left(E_{k+1} = e \mid X_k = x\right) \right] =$$

$$= \sum_{e \in \Gamma(x)} \left[P\left(X_{k+1} = x' \mid X_k = x,\, E_{k+1} = e\right) \cdot \frac{\lambda_e}{\Lambda(x)} \right] \triangleq$$

$$\overset{\triangleq}{\underset{\text{for a more lean notation}}{=}} \sum_{e \in \Gamma(x)} \left[\mathsf{p}\left(x' \mid x, e\right) \cdot \frac{\lambda_e}{\Lambda(x)} \right]$$

These two results are quite good, but there are still more things that can be done. By defining $\mathcal{E} = \{1, 2, \ldots, m\}$ and $\mathcal{X} = \{1, 2, \ldots, n\}$ it's possible to define the following matrices and vector:

$$P_E = \begin{bmatrix} P\left(E_{k+1} = 1 \mid X_k = 1\right) & P\left(E_{k+1} = 2 \mid X_k = 1\right) & \ldots & P\left(E_{k+1} = m \mid X_k = 1\right) \\ P\left(E_{k+1} = 1 \mid X_k = 2\right) & P\left(E_{k+1} = 2 \mid X_k = 2\right) & \ldots & P\left(E_{k+1} = m \mid X_k = 2\right) \\ \vdots & \vdots & \ddots & \vdots \\ P\left(E_{k+1} = 1 \mid X_k = n\right) & P\left(E_{k+1} = 2 \mid X_k = n\right) & \ldots & P\left(E_{k+1} = m \mid X_k = n\right) \end{bmatrix}$$

$$P_X = \begin{bmatrix} P\left(X_{k+1} = 1 \mid X_k = 1\right) & P\left(X_{k+1} = 2 \mid X_k = 1\right) & \ldots & P\left(X_{k+1} = n \mid X_k = 1\right) \\ P\left(X_{k+1} = 1 \mid X_k = 2\right) & P\left(X_{k+1} = 2 \mid X_k = 2\right) & \ldots & P\left(X_{k+1} = n \mid X_k = 2\right) \\ \vdots & \vdots & \ddots & \vdots \\ P\left(X_{k+1} = 1 \mid X_k = n\right) & P\left(X_{k+1} = 2 \mid X_k = n\right) & \ldots & P\left(X_{k+1} = n \mid X_k = n\right) \end{bmatrix}$$

$$\Pi_X(k) = \begin{bmatrix} P(X_k = 1) & P(X_k = 2) & \cdots & P(X_k = n) \end{bmatrix}$$

Thus, it's possible to redefine $P(X_{k+1} = x' \mid X_k = x)$ as the probability that the $k+1$-th state is j as:

$$P\left(X_{k+1} = j\right) = \sum_{i \,\in\, \mathcal{X}} \underbrace{P\left(X_{k+1} = j \mid X_k = i\right)}_{(i,j)-\text{th entry of } P_X} \underbrace{P\left(X_k = 1\right)}_{i-th \text{ entry of } \Pi_X(k)}$$

$P(X_{k+1} = j)$ will be the j-th entry of the vector $\Pi_X(k+1)$ and of the product $\Pi_X(k) \cdot P_X$. So assuming that $\Pi_X(0)$ is known:

$$\begin{aligned}
\Pi_X(0) &= \left[\mathsf{p}_0(1), \quad \mathsf{p}_0(2), \quad \ldots, \quad \mathsf{p}_0(n)\right] \\
\Pi_X(1) &= \Pi_X(0) \cdot P_X \\
\Pi_X(2) &= \Pi_X(1) \cdot P_X &&= \Pi_X(0) \cdot P_X^2 \\
\Pi_X(3) &= \Pi_X(2) \cdot P_X &&= \Pi_X(0) \cdot P_X^3 \\
&\vdots \\
\Pi_X(k+1) &= \Pi_X(k) \cdot P_X &&= \Pi_X(0) P_X^k
\end{aligned}$$

Likewise, it's possible to obtain the same result for the events:

$$P\left(E_{k+1} = j\right) = \sum_{i \,\in\, \mathcal{X}} P\left(E_{k+1} = j \mid X_k = i\right) P\left(X_k = i\right)$$

$$\Pi_E(k) = \left[P\left(E_k = 1\right), \quad P\left(E_k = 2\right), \quad \ldots, \quad P\left(E_k = m\right)\right]$$

$$\Pi_E(k+1) = \Pi_X(k) P_E = \Pi_X(0) P_X^k P_E$$

With this result, it's finally possible to say that *when using stochastic timed automata with Poisson clock structure, it's enough to know the matrices P_X and P_E and the vector $\Pi_X(0)$ to find all the future state and event probabilities.*

Distribution of State Holding Times The state holding time $V(x)$ is a continuous random variable characterising the time spent by the system in state x. Notice that, while the system is in state x, there might occur events which do not trigger a state transition: this implies that $V(x)$ keeps increasing until the system leaves state x. Let's compute the CDF of the state holding time, for a stochastic timed automaton with Poisson clock structure, and show that it is exponentially distributed with rate:

$$\sum_{e \,\in\, \Gamma(x)} \lambda_e \left[1 - \mathsf{p}\left(x \mid x, e\right)\right]$$

Proof: The cdf of $V(x)$ is:

$$P(V(x) \leq t) = 1 - P(V(x) > t)$$

Consider $P(V(x) > t)$ only:

$$P(V(x) > t) = P\left(\textit{no state transitions over the } (0,t] \textit{ interval} \mid X(0) = x\right) =$$

$$= P\left(\bigcap_{e \in \Gamma(x)} \textit{no state transition triggered by event } e \textit{ over } (0,t] \mid X(0) = x\right) = {}^{1}$$

$$= \prod_{e \in \Gamma(x)} P\left(\textit{no state transition triggered by event } e \textit{ over } (0,t] \mid X(0) = x\right) = {}^{2}$$

$$= \prod_{e \in \Gamma(x)} P\left(\begin{array}{l} \bigcup_{n=0}^{+\infty} \textit{event } e \textit{ occurs exactly } n \textit{ times over } (0,t] \\ \\ \textit{and never triggers a state transition} \mid X(0) = x \end{array}\right) = {}^{3}$$

$$= \prod_{e \in \Gamma(x)} \sum_{n=0}^{+\infty} P\left(\begin{array}{l} \textit{event } e \textit{ occurs exactly } n \textit{ times over } (0,t] \\ \\ \textit{and never triggers a state transition} \mid X(0) = x \end{array}\right) =$$

In this last probability the occurrences of an *i.i.d* set of exponentially distributed random variables are counted, which is equivalent to a Poisson counting process. Therefore, it's possible to substitute Equation 2 and multiply it by the probability that the system remains in x, given event e, exactly n times:

$$= \prod_{e \in \Gamma(x)} \sum_{n=0}^{+\infty} \frac{(\lambda_e t)^n}{n!} \cdot e^{-\lambda_e t} \cdot \mathsf{p}\left(x \mid x, e\right)^n = \prod_{e \in \Gamma(x)} e^{-\lambda_e t} \sum_{n=0}^{+\infty} \frac{\left[(\lambda_e t) \cdot \mathsf{p}(x \mid x, e)\right]^n}{n!} =$$

Now the last sum can be rewritten as $e^x = \sum_{n=0}^{+\infty} x^n/n!$ (Maclaurin series expansion):

$$= \prod_{e \in \Gamma(x)} \exp(-\lambda_e t) \cdot \exp(\lambda_e t \cdot \mathsf{p}\left(x \mid x, e\right)) = \prod_{e \in \Gamma(x)} \exp(-\lambda_e \left[1 - \mathsf{p}(x \mid x, e)\right] t) =$$

$$= \exp\left(-\sum_{e \in \Gamma(x)} \lambda_e \left[1 - \mathsf{p}(x \mid x, e)\right] t\right)$$

Switching to the cdf of interest is straightforward (namely, the complement of $P(V(x)) > t$), thus:

$$P\left(V(x) \leq t\right) = 1 - \exp\left(-\sum_{e \in \Gamma(x)} \lambda_e \left[1 - \mathsf{p}(x \mid x, e)\right] t\right)$$

■

[1] Thanks to the independence of the events' lifetimes in Poisson clock structure

[2] It may occur more than once

[3] Union of disjoint events, you may sum them up

5 Markov Chains

Stochastic Processes An example of a stochastic process was already proposed when talking about 'Poisson counting process', now a slightly more formal definition will be given: *'a stochastic process is a collection* $\{X(t)\}_{t\in T}$ *of random variables indexed by a time index* $t \in T$*'*, where T is a time interval which can be either discrete or continuous. The stochastic processes are actually classified depending on the nature of T:

- **if** T **is discrete** (either finite or not) the process is a *'discrete time stochastic process'* and it's also called *'chain'*.
- **if** T **is continuous** the process is a *'continuous time stochastic process'*.

To characterise stochastic processes it's necessary to provide joint distributions of all the possible n-tuples of the random variables which compose the process. Since this is really hard to realise, in these notes the stochastic processes will be used only when the independence between the random variables holds, therefore with stochastic timed automata with Poisson clock structure. Finally, the independence concept in stochastic processes is defined as follows:

Given $X(t_1), X(t_2), \ldots, X(t_n)$ random variables with $t_1 < t_2 < \cdots < t_n \in T$ and $n \in \mathbb{N}^+$ the process is said *'independent'* if all the $n-$tuples are independent. So, let $x(t)$ be the realisation of the random variable $X(t)$, if the equality:

$$P\left(X(t+s) = \widetilde{x} \mid X(\tau) = x(\tau),\ \forall \tau \leq t\right) = P\left(X(t+s) = \widetilde{x}\right)$$

holds, the process is independent. This means that the history of the past evolution of the system is irrelevant for predicting the future evolution ($t + s$ is called *prediction horizon*).

Continuous Time Homogeneous Markov Chains (CTHMC) are a subset of a kind of stochastic processes called 'Markov processes' for which the definition of process independence is more relaxed than the standard one just given. If the condition:

$$P\left(X(t+s) = \widetilde{x} \mid X(\tau) = x(\tau),\ \forall \tau \leq t\right) = P\left(X(t+s) = \widetilde{x} \mid X(t) = x(t)\right)$$

holds, the process is independent. This definition allows the stochastic processes whose next realisation depends only on the current one to be called independent (*'Markov property'*).

A Continuous Time Homogeneous Markov Chain is a stochastic process with the following properties:

- $T = \mathbb{R}^+ \Rightarrow \{t \in \mathbb{R} : \ t \geq 0\}$ (**Continuous Time**).
- $X(t) \in \mathcal{X} = \{1, 2, \dots\}$ (**Chain**).
- **Markov property**.
- **Homogeneity**: the transition function depends only on the prediction horizon

$$P\left(X(t+s) = j \mid X(t) = i\right) = P\left(X(t'+s) = j \mid X(t') = i\right), \ \forall t \neq t'$$

Thus, it can be rewritten as:

$$\mathsf{p}_{i,j}(s) = P\left(X(t+s) = j \mid X(t) = i\right)$$

Chapman-Kolmogorov Equation Please notice that $x(t) \in \mathcal{X}$ implies that the process' random variables realisations are the states of the system modelled as a CTHMC. So, let's try to compute the probability that, given a current state i, after a time s the state will be j:

$$\mathsf{p}_{i,j}(s) = P\left(X(t+s) = j \mid X(t) = i\right)$$

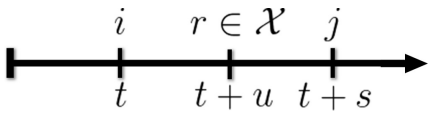

Figure 5: At time t the current state is i. At $t+s$ the next state is j, while at $t+u$ the state is $r \in \mathcal{X}$ which can be i, j or any other state.

In order to execute the computation, let's consider the case in Figure 5:

$$\mathsf{p}_{i,j}(s) = P\left(X(t+s) = j \mid X(t) = i\right) = {}^{4}$$

$$= \sum_{r \in \mathcal{X}} P\left(X(t+s) = j \mid X(t+u) = r,\, X(t) = i\right) \cdot P\left(X(t+u) = r \mid X(t) = i\right) = {}^{5}$$

$$= \sum_{r \in \mathcal{X}} P\left(X(t+s) = j \mid X(t+u) = r\right) \cdot P\left(X(t+u) = r \mid X(t) = i\right) =$$

$$= \sum_{r \in \mathcal{X}} \mathsf{p}_{r,j}(s-u) \cdot \mathsf{p}_{i,r}(u)$$

[5]By applying the total proability rule
[5]By applying the Markov property

The equality:

$$\mathsf{p}_{i,j}(s) = \sum_{r \in \mathcal{X}} \mathsf{p}_{i,r}(u) \cdot \mathsf{p}_{r,j}(s-u) \tag{4}$$

is known as 'Chapman-Kolmogorov equation'. As seen for the stochastic timed automata with Poisson clock structure, also here it's possible to define a matrix for the transition probabilities $\mathsf{p}_{i,j}(s)$. Let's call this matrix $H(s)$:

$$H(s) = \begin{bmatrix} \mathsf{p}_{1,1}(s) & \mathsf{p}_{1,2}(s) & \cdots \\ \mathsf{p}_{2,1}(s) & \mathsf{p}_{2,2}(s) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

This matrix has some important properties:

1. The sum of the elements along any row of $H(s)$ is 1 (being the sum of the probabilities related to all the possible cases starting from the state the row refers to).

2. As a result of the previous property, if $s = 0$ then the probability that the state does not change is 1, therefore $\mathsf{p}_{i,i} = 1$ and $\mathsf{p}_{i,j} = 0,\ j \neq i$:

$$H(0) = I$$

3. Since $\mathsf{p}_{i,j}(s)$ is the generic element of matrix $H(s)$, from Equation 4 one notices that $H(s)$ equals the matrices product $H(u)$ by $H(s-u)$ (matrix form of the Chapman-Kolmogorov equation):

$$H(s) = H(u) \cdot H(s-u)$$

From the last property it's possible to define the derivative of $H(s)$:

$$\frac{dH(s)}{ds} = H(s) \cdot Q, \quad Q \triangleq \lim_{ds \to 0} \frac{H(ds) - I}{ds}$$

Proof: We would like to obtain the difference quotient. Let's consider an infinitesimal difference ds:

$$H(s+ds) = H(s)H(ds) = {}^{6}$$

$$\frac{H(s+ds) - H(s)}{ds} = \frac{H(s)H(ds) - H(s)}{ds}$$

Taking the limit for $ds \to 0$ yields the difference quotient on the left-hand side:

$$\lim_{ds \to 0} \frac{H(s+ds) - H(s)}{ds} = \lim_{ds \to 0} \frac{H(s)H(ds) - H(ds)}{ds}$$

[6] By subtracting H(s) from both sides and dividing by ds

Thus:

$$\frac{dH(s)}{ds} = H(s) \cdot \lim_{ds \to 0} \frac{H(ds) - I}{ds}$$

The right-hand side limit has an indeterminate form $0/0$ since $H(0) = I$, then

$$\frac{(I - I)}{ds} \xrightarrow{ds \to 0} \frac{0}{0}$$

Let's *assume that the limit exists* and let's call it Q. Since all the elements involved in the limit are matrices, also Q will be a matrix (as a side note, matrix Q can be estimated on the field by performing measurements on the system). Finally, taking into account property #2 it's possible to define the following Cauchy problem:

$$\begin{cases} \dfrac{dH(s)}{ds} = H(s) \cdot Q \\ H(0) = I \end{cases}$$

Whose solution is:

$$H(s) = e^{Qs}, \quad e^{Qs} = \sum_{n=0}^{+\infty} \frac{(Q \cdot s)^n}{n!}$$

Where e^{Qs} is the matrix exponential. Now we need to validate our initial assumption on the existence of Q. Starting from the initial definition of matrix Q, substitute the matrix exponential:

$$\lim_{ds \to 0} \frac{H(ds) - I}{ds} = \lim_{ds \to 0} \frac{e^{Qds} - I}{ds} \overset{\text{Taylor}}{\underset{1^{st} \text{ order}}{=}}$$

$$\lim_{ds \to 0} \frac{(I + Qds + o(ds)) - I}{ds} = \lim_{ds \to 0} \frac{Q\cancel{ds}}{\cancel{ds}} + \frac{\cancel{o(ds)}}{\cancel{ds}} \overset{o(1)}{\nearrow} = Q$$

Due to the last equality, it is clear that the result is consistent with the initial assumption.

■

5.1 The Q Matrix

From the previous proof, it's hard to obtain any direct information about Q, so this section will be dedicated to the properties of the Q matrix. To start with, this matrix is called '*Transition Rate Matrix*':

$$Q = \begin{bmatrix} q_{1,1} & q_{1,2} & \cdots \\ q_{2,1} & q_{2,2} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

where the generic element $q_{i,j}$ is called transition rate (and has dimention s^{-1}). Since the need of defining Q came from the computation of the derivative of the matrix H, the properties of H have been used to define the ones of Q. Let's start from the property of H about the sum along any row being always 1. Said $\underline{\mathbf{1}}$ the vector $(1, 1, 1, \ldots, 1)^T$:

$$H(s) \cdot \underline{\mathbf{1}} = \underline{\mathbf{1}} \quad \underset{\text{of both the sides}}{\overset{\text{taking the derivative}}{\Rightarrow}} \quad \underbrace{\frac{d\left(H(s)\right)}{ds}}_{H(s)Q} \underline{\mathbf{1}} + \underbrace{\frac{d\underline{\mathbf{1}}}{ds}}_{=O} H(s) = \underbrace{\frac{d\underline{\mathbf{1}}}{ds}}_{=O} \Rightarrow$$

$$\Rightarrow \; H(s)Q \cdot \underline{\mathbf{1}} = O \quad \underset{H(0)=I}{\overset{s \to 0}{\Rightarrow}} \quad Q \cdot \underline{\mathbf{1}} = O$$

Where O is the vector $(0, 0, 0, \ldots, 0)$. From this final result it's clear that:

The sum along every row of Q is always 0. $\quad (\star)$

This property has an important implication: *One of the eigenvalues of Q it's always* 0, since summing to one of the columns of Q all the other columns will produce O. Therefore, the property $(\star)$ implies the linear dependency of one of the columns and, since the number of zero eigenvalues is equal to the dimension of the kernel [2], one of the eigenvalues will always be 0.

So, keeping in mind the property $(\star)$, let's proceed with the definition of the generic coefficient $q_{i,j}$. At the moment, only the case in which $i \neq j$ will be considered. Let's start with the first order Taylor's representation of $H(ds)$, with $ds \to 0$:

$$H(ds) = e^{Qds} \quad \underset{1^{st} \text{ order}}{\overset{\text{Taylor}}{=}} \quad I + Qds + o(ds)$$

So, reducing this last equation to the coefficients of the matrices involved, the result will be (always under the condition $ds \to 0$):

$$\mathsf{p}_{i,j}(ds) = q_{i,j}ds + o(ds)$$

Looking at the equation it's clear that *all the $q_{i,j}$ with $i \neq j$ are non-negative*, since $\mathsf{p}_{i,j}(s) \geq 0$ (being a probability), $ds \geq 0$ (as it represents time, which cannot be negative) and $o(ds)$ goes to zero with $ds \to 0$.

[2] $dim(ker(Q)) = dim(Q) - rank(Q)$

The non-negativity of $q_{i,j}$ and the property $(\star)$, impose that:

All the coefficients $q_{i,i}$ of the main diagonal of Q must be non-positive, and in particular:

$$q_{i,i} \;=\; -\sum_{j \neq i} q_{i,j}$$

Physical Representation Of The Transition Rates On a physical level the elements of the Q matrix have two different meanings, depending on whether they are on the main diagonal or not:

- The elements on the main diagonal define the rate of the distribution of the state holding times (which are exponentially distributed) of the system and, in particular:
$$E\left[V(i)\right] = \frac{1}{-q_{i,i}}$$
- All the other elements are used, together with the ones from the main diagonal, to compute the overall state transition probability, independent from the time, from the generic state i to the generic state $j \neq i$ with the formula:
$$p_{i,j} = -\frac{q_{i,j}}{q_{i,i}}$$

Let's start by computing the CDF of the state holding time of the generic state i, in order to provide a proof for the statement done about elements on the main diagonal:

$$F_i(t) \;\triangleq\; P\left(V(i) \leq t\right)$$

And let's consider a small increment dt of t such that a state transition (only one) occurs in the time interval $(t, t+dt]$:

$$F_i(t+dt) - F_i(t) = P\left(t < V(i) \leq t+dt\right) = P\left(V(i) \leq t+dt \mid V(i) > t\right) \cdot \underbrace{P\left(V(i) > t\right)}_{1-F_i(t)}$$

$$\overset{\text{in particular}}{\Rightarrow} \qquad P\left(V(i) \leq t+dt \mid V(i) > t\right) = \sum_{j \neq i} \mathsf{p}_{i,j}(dt) \quad \overset{\text{Taylor}}{\underset{1^{st}\text{ order}}{=}} \quad \sum_{j \neq i} q_{i,j} dt + o(dt) \Rightarrow$$

$$\Rightarrow F_i(t+dt) - F_i(t) = \left[\sum_{j \neq i} q_{i,j} dt + o(dt)\right]\left[1 - F_i(t)\right]$$

Dividing both the members of the last equation for dt and computing the limit for $dt \to 0$

$$\lim_{dt \to 0} \frac{F_i(t+dt) - F_i(t)}{dt} = \frac{d\left(F_i(t)\right)}{dt} \quad = \quad \lim_{dt \to 0} \frac{\left[\sum_{j \neq i} q_{i,j} dt + o(dt)\right]\left[1 - F_i(t)\right]}{dt} =$$

$$= \lim_{dt \to 0} \left[\sum_{j \neq i} \frac{q_{i,j} dt}{dt} + \frac{o(dt)}{dt}\right]\left[1 - F_i(t)\right] \quad = \quad \sum_{j \neq i} q_{i,j}\left[1 - F_i(t)\right] = -q_{i,i}\left[1 - F_i(t)\right]$$

In order to find $F_i(t)$, let's solve the following Cauchy Problem:

$$\begin{cases} \dfrac{d(F_i(t))}{dt} = -q_{i,i} \cdot [1 - F_i(t)] \\ F_i(0) = 0 \end{cases}$$

Where $F_i(0)$ is the initial value of the generic state holding time $V(i)$. In order to compute the solution let's apply the following substitution:

$$G_i(t) \triangleq 1 - F_i(t) \quad \overset{\text{substituting } G_i(t) \text{ in both the equations}}{\Rightarrow} \quad \begin{cases} \dfrac{d(G_i(t))}{dt} = -\dfrac{d(F_i(t))}{dt} = q_{i,i} \cdot G_i(t) \\ G_i(0) = 1 \end{cases}$$

The solution to this Cauchy problem is:

$$G_i(t) = \exp(q_{i,i} \cdot t) \Rightarrow F_i(t) = 1 - G_i(t) = 1 - \exp(q_{i,i} \cdot t)$$

This means that all the state holding times in CTHMC are exponentially distributed and, moreover that the rate of $F_i(t)$ is negative, namely $-q_{i,i}$. Computing the expected value of the distribution will return the result of the first statement.

■

Now the proof for the statement about the elements outside the main diagonal will be provided. Given two generic states i and j, where $i \neq j$, and a generic time interval $(t, t + dt]$, let's compute the following probability:

$$P\Big(\textit{a transition from } i \textit{ to } j \textit{ occurs in the interval } (t, t + dt] \mid X_k = i\Big) = \tag{5}$$

$$= P\left(V(i) > t\right) P_{i,j}(dt) = \left[1 - P\left(V(i) \leq t\right)\right] \underbrace{P_{i,j}(dt)}_{\substack{\text{Taylor } 1^{\text{st}} \text{ order} \\ \Rightarrow q_{i,j}dt + o(dt)}} \overset{dt \to 0}{=} \left[1 - \left(1 - e^{q_{i,i}t}\right)\right] q_{i,j}dt$$

$$\Rightarrow P\left(V(i) > t\right) P_{i,j}(dt) = \exp(q_{i,i}t) q_{i,j}dt$$

Let's now introduce another probability, which is the overall transition probability from i to j, without taking into account the time:

$$p_{i,j} = P(X_{k+1} = j \mid X_k = i)$$

$p_{i,j}$ can be obtained as sum of the probabilities like 5 computed for all the intervals of the kind $[t_n, t_n + dt_n]$, with $t_0 = 0$, $dt \to 0$, $n \in \mathbb{N}^+$:

$$p_{i,j} = \int_0^{+\infty} P\left(\text{a transition from } i \text{ to } j \text{ occurs in the interval } [t, t + dt] \mid X_k = i\right) =$$

$$= \int_0^{+\infty} e^{q_{i,i}t} q_{i,j}dt = q_{i,j} \left[\frac{e^{q_{i,i}t}}{q_{i,i}}\right]_0^{+\infty} = -\frac{q_{i,j}}{q_{i,i}}$$

■

5.2 Steady State Analysis

Let's consider the 'State Probability Vector' $\Pi(t)$, similar to $\Pi_X(k)$ in *Section 4*:

$$\Pi(t) \triangleq \left[\begin{array}{cccc} \Pi_1(t), & \Pi_2(t), & \Pi_3(t), & \dots \end{array} \right]$$

The 'State Probability' of the generic $j-th$ state at time t will be obtained as:

$$\Pi_j(t) \;=\; P(x(t)=j) \;=\; \sum_{i \in \mathcal{X}} \underbrace{P(x(t)=j \mid x(0)=i)}_{P_{i,j}(t)} \cdot \underbrace{P(x(0)=i)}_{\Pi_i(0)} \;=\; \Pi(0) \cdot H_j(t)$$

Where $H_j(t)$ is the $j-th$ column of the matrix $H(t)$. This means that, given a time instant t, it's possible to compute all the state probabilities in one go by using:

$$\Pi(t) \;=\; \Pi(0) \cdot H(t) \;=\; \Pi(0) \cdot e^{Qt}$$

In the kind of system treated in this section the behaviour of the probabilities in time can be divided in two states: 'Transient State' and 'Steady State':

- **Transient State**: In order to analyse the behaviour of the system's state probabilities during the transient state, the derivative of $\Pi(t)$ must be computed:

  $$\frac{d\Pi(t)}{dt} \;=\; \Pi(0)\frac{dH(t)}{dt} \;=\; \Pi(0)H(t)Q \;=\; \Pi(t)Q$$

  Which means that, to study the behaviour of the system in the transient state, it's enough to find the solution to the following Cauchy problem:

  $$\begin{cases} \dfrac{d\Pi(t)}{dt} = \Pi(t)Q \\ \Pi(0) = \Pi_0 \end{cases}$$

  Where Π_0 is the initial state of the system.

- **Steady State**: A system reaches steady state when every probability in the system starts varying less and instead converges asymptotically to a constant value. So in order to study steady state probabilities, it's necessary to focus on the following limit:

  $$\lim_{t \to \infty} \Pi_i(t)$$

Classification of States in this paragraph some properties of the states, useful for the steady state analysis, will be explained:

- **Reachability**: A generic state j is reachable from state i if

$$\exists\, s : P_{i,j}(s) > 0.$$

  Such concept can be also informally explained with the following definition: '*it must exist a directed path from state i to state j*'.

- **Closure**: A subset $S \subseteq \mathcal{X}$ is '*closed*' if

$$P_{i,j} = 0, \quad \forall i \in S, \quad j \in \mathcal{X} \setminus S.$$

  Informally, it's possible to say that from the subset S it's not possible to reach the states of the subset $\mathcal{X} \setminus S$:

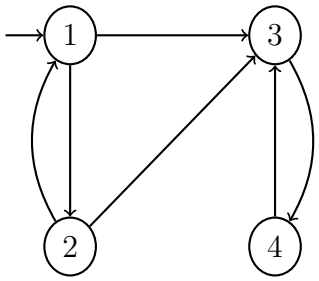

Figure 6: the subset $\{3, 4\}$ is an example of closed set, since once the system enters in state 3 it's not possible that it will return to state 1 or 2 anymore.

- **Irreducibility**: A closed subset $S \subseteq \mathcal{X}$ is called irreducible if every state of S is reachable from other states of S.

- **Recurrence**: Let's define the random variable $T_{i,i}$ as the time that takes to the system to return in state i. Let's also define the probability $\rho_i(t) \triangleq P\left(T_{i,i} < t\right)$. Considering the limit for $t \to \infty$ of $\rho_i(t)$ it's possible to know if the system will ever return to state i.

$$\rho_i \triangleq \lim_{t\to\infty} \rho_i(t) = \begin{cases} 1 & \text{then the state will return (for } t \to \infty) \text{ to the state i} \\ a < 1 & \text{then the state i is transient} \end{cases}$$

  Starting form the assumption that the state i is recurrent, it's possible to realise a further classification. First of all it's important to remember that $\rho_i(t)$ is the CDF of $T_{i,i}$, so it's possible to define also the PDF $f_i(t)$ as $\rho_i(t)$'s derivative.

So, now it's possible to compute the expected value M_i of $T_{i,i}$, as follows:

$$M_i \triangleq E\left[T_{i,i}\right] = \int_0^{+\infty} t \cdot f_i(t)dt = \begin{cases} \infty & \text{then the integral doesn't converge} \\ & \text{and the state } i \text{ is said 'null recurrent'.} \\ a < \infty & \text{then the integral converges and the} \\ & \text{state } i \text{ is said 'positive recurrent'.} \end{cases}$$

In real applications the behaviours of null recurrent and transient states are the same, so the positive recurrent states are the only ones in which the system actually returns to the state i.

From these definition it's possible to define two theorems and an important corollary:

1. If i is a positive recurrent state and j is reachable from i, then j is positive recurrent.
2. If S is a closed, irreducible and finite subset of $\mathcal{X}$, then all the states in S are positive recurrent.

2.1. *An irreducible and finite Markov Chain has only positive recurrent states.*

Steady State Analysis Let's now define the stationary probability vector:

$$\Pi \triangleq \begin{bmatrix} \Pi_1, & \Pi_2, & \Pi_3, & \dots \end{bmatrix} \quad , \quad \Pi_i \triangleq \lim_{t \to \infty} \Pi_i(t)$$

The actual definition of the stationary probability vector brings some problems that need to be solved before actually working with it, such as:

- Existence of the limit.
- Conditions for the independence from the initial state Π_0.
- Consistency of the probability vector, as the limit might exist but the probabilities might not sum up to 1.

A theorem (whose proof is not provided in these notes) states that for Continuous Time Homogeneous Markov Chains these problems, under precise conditions, can all be easily solved.

Theorem:
For a CTHMC, which is irreducible and with all positive recurrent states, the limits

$$\Pi_i = \lim_{t\to\infty} \Pi_i(t)$$

exist, with $\Pi_i > 0,\ \forall i \in \mathcal{X}$ and they are all independent of Π_0. Moreover, the vector Π can be computed by solving the system of linear equations:

$$\begin{cases} \Pi Q = 0 \\ \sum_{i\in\mathcal{X}} \Pi_i = 1 \end{cases}$$

From this last theorem, keeping in mind also the corollary #2.1 of the previous paragraph, the following corollary can be obtained:

Corollary:
The previous theorem holds for irreducible and finite CTHMC.

In both the theorem an the corollary the vector Π can be found by solving the following system:

$$\begin{cases} \Pi Q = 0 & n \text{ equations} \\ \sum_{i\in\mathcal{X}} \Pi_i = 1 & 1 \text{ equation} \end{cases}$$

Where n is the cardinality of Π. This means that there is a redundant equation, which can be found in $\Pi Q = 0$, due to the fact that Q doesn't have full rank since it has an eigenvalue $\lambda = 0$. So, the constraint $\sum_{i\in\mathcal{X}} \Pi_i = 1$ ensures both the consistency of the probability vector and the existence of a unique solution. As a final clarification about the equations in the system, the set of equations given by $\Pi Q = 0$ comes from the definition already treated for the transient state definition:

$$\frac{d\Pi(t)}{dt} = \Pi(t)Q$$

Since for $t \to \infty$ the state probabilities are expected to converge to constant values, their derivatives are expected to converge to zero, so:

$$\frac{d\Pi(t)}{dt} = \Pi(t)Q \quad \overset{t\to\infty}{\Rightarrow} \quad \Pi Q = 0$$

Equivalent CTHMC A Stochastic Timed Automaton with Poisson Clock Structure $(\mathcal{E}, \mathcal{X}, \Gamma, P, p_0, F)$ is 'stochastically equivalent' to a CTHMC $(\mathcal{X}, Q, \Pi_0)$ which has:

- the same distributions for the state holding times:

$$V_i \sim Exp\left(\frac{1}{-q_{i,i}}\right)$$

- the same state transition probabilities:

$$p_{i,j} = \frac{q_{i,j}}{-q_{i,i}}$$

So, in order to find the equivalent Markov Chain for a Stochastic Timed Automaton with Poisson Clock Structure it's enough to:

- Compute the expected value of the state holding time for every state, and compute the $q_{i,i}$ as:

$$q_{i,i} = -\frac{1}{E[V(i)]} = -\sum_{e \in \Gamma(i)} \lambda_e \left[1 - \mathsf{p}(i \mid i, e)\right]$$

- Compute the $p_{i,j}$ according to the frequentist probability in order to find the corresponding $q_{i,j}$ as:

$$q_{i,j} = -q_{i,i} p_{i,j} = \sum_{e \in \Gamma(i)} \lambda_e \mathsf{p}(j \mid i, e)$$

Where λ_e is the rate of the exponential distribution (since all the events lifetimes' distributions are exponential) of the event e.

6 Queueing Systems

Queueing systems are a really important kind of discrete events systems and they can be treated using any of the models defined up to now. When looking for a representation for a queueing system, a really general one could be the following:

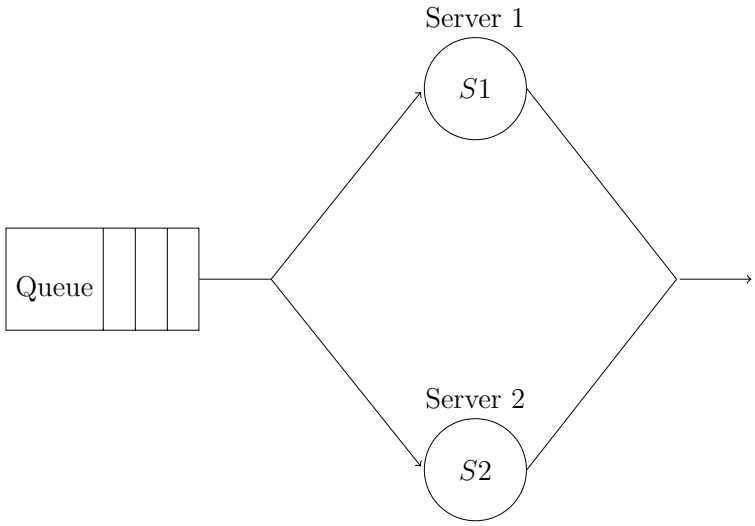

Figure 7: Generic queueing system representation.

In order to define a queueing system, some parameters must be specified:

- **Structural Parameters**:
  - Number of servers.
  - Capacity of the queue.
- **Operating Policies**:
  - Number of accepted customers and the kind of service they need (type).
  - Scheduling Policy of the queue (FIFO, Round-Robin, ...).
  - Conditions to accept new customers in the system.
- **Distributions of interarrival and sevice times**.

The choices made in the definition phase of a queueing system will affect its behaviour, especially the '*effective production rate*'. When producing any kind of product (or providing any kind of service), this is done by expecting a certain demand for that product (or service). The queueing system must be designed in order to meet the demand, in particular the effective production rate μ_{eff} must be greater than the '*demand rate*' d:

$$\mu_{eff} \geq d$$

Kendall's Notation can be used to describe queueing systems where there exists only one kind of customer. The notation is composed by the following set of parameters:

$$\mathbf{A \ / \ B \ / \ m \ / \ K \ / \ n \ / \ D}$$

Where:

- **A**: is the distribution of the interarrival times.
- **B**: is the distribution of the service times.

The possible values for A and B are:

$$\begin{array}{rcc} M & : & \text{Exponential distribution (Memoryless)} \\ U & : & \text{Uniform distribution} \\ G & : & \text{Generic distribution} \\ D & : & \text{Deterministic distribution} \end{array}$$

- **m**: is the number of servers.
- **K**: is the capacity of the system (therefore the capacity of the queue is $K - m$).
- **n**: is the size of the population from which the customers come.
- **D**: is the scheduling policy.

When one or more parameters are not specified it means that the model of the queueing system operates without taking into account the information deriving form them. In fact, in these notes, only systems that can be represented through the notation $A/B/n/K$ are treated.

Queueing systems in steady state Before proceeding with the behaviour of queueing systems in steady state, it's necessary to provide some definitions. Let's consider the generic k-th customer entering a queueing system. The time spent by the customer in the system doesn't necessarily correspond to the time that took the system to process the customer request due to the presence of the queue. So, said Z_k and W_k respectively the 'service time' and the 'waiting time' of the k-th customer in the system, the overall time spent in the system is called 'system time' and it's defined as:

$$S_k \triangleq W_k + Z_k$$

Generally, the distribution of the system time changes accordingly to the number k of customers accepted in the system up to that moment, but if, after a certain amount of customers accepted, the system enters in a steady state, then the distribution of the system times will be the same for all the customers.

System time in steady state: If there exists a random variable S such that:

$$P\left(S \leq t\right) = \lim_{k \to \infty} P\left(S_k \leq t\right) \quad , \quad \forall t$$

Then the random variable S describes the system time of a generic customer while the system is in steady state.

Average amount of customers in steady state If there exists a random variable X such that:

$$P(X = i) = \lim_{t \to \infty} P(X(t) = i), \quad \forall i$$

Then the random variable X describes the number of customers in the system when this is in steady state. Moreover, always in steady state, since the number of customers in the system doesn't depend on the time anymore, then also its expected value $E\left[X(t)\right]$ doesn't:

$$E\left[X(t)\right] = \sum_i i \cdot P\left(X(t) = i\right) \quad \stackrel{t \to \infty}{=} \quad \sum_i i \cdot P\left(X = i\right)$$

From the rightmost result comes out that $E\left[X(t)\right] = E\left[X\right]$. In particular, thorugh these definitions it's possible to find a necessary condition for the system to be in steady state.

Effective rates and necessary condition for steady state Said μ_{eff} the effective production rate and λ_{eff} the effective arrival rate (rate of arrivals accepted in the system) at steady state the following condition must hold:

$$\mu_{eff} = \lambda_{eff}$$

This condition holds also for queueing networks, which are queueing systems composed by others queueing sub-systems. So, considering a system consisting of two stations in series, each of them composed by a machine preceded by a buffer, if the whole system reaches steady state the following conditions will hold:

$$\mu_{eff,1} = \lambda_{eff,1} \quad , \quad \mu_{eff,2} = \lambda_{eff,2}$$

When studying the steady state of a queueing system, an important parameter to take into account would be the 'utilization' U, which is the fraction of time (over all the system activity time) in which the machine is actively working. If $U = 1$ the machine it's always working, of course if $U = 1$ for an observation time $t \to \infty$, then the machine is perfect.

The utilization value is also used to obtain another parameter called *throughput*, which is defined as the mean number of requests served during a time unit:

$$throughput \;\; = \;\; U \cdot m \cdot \mu_{eff}$$

In particular, on a single server system with $U = 1$ the throughput is equal to μ_{eff}.

Little's Law Let's consider a queueing network at steady state and a curve Σ, closed around any fraction of the network (also the whole network can be considered an acceptable fraction). Let's consider only the portion of network within Σ and let's define:

- λ_Σ as the arrival rate for the arrivals which enter the curve Σ (if Σ surrounds the whole network $\lambda_\Sigma = \lambda_{eff}$).
- $E[X_\Sigma]$ as the expected value of the number of customers within Σ (if Σ surrounds the whole network $E[X_\Sigma] = E[X]$).
- $E[S_\Sigma]$ as the expected value of the time spent by a customer in Σ (if Σ surrounds the whole network $E[S_\Sigma] = E[S]$).

Little's Law states that:

$$E[X_\Sigma] = \lambda_\Sigma \cdot E[S_\Sigma]$$

This law comes particularly handy since $E[S_\Sigma]$ it's often hard to compute directly, while instead it's easy to do it for the other two parameters.

PASTA Property It's another property for queueing systems. Let's define:

- $A(t)$ as the occurrence of an arrival in the system at the time t.
- $\alpha_n(t)$ as the posterior probability that the state at time t is n, knowing that at time t an arrival occurs:
$$\alpha_n(t) \;=\; P(X(t) = n \mid A(t))$$
- $\Pi_n(t)$ as the prior probability that the state at time t is n:
$$\Pi_n(t) \;=\; P(X(t) = n)$$

Generally, $\alpha_n(t) \neq \Pi_n(t)$, but PASTA property states that:

If the arrivals are generated by a Poisson process and the lifetimes for arrivals and service terminations are independent then:

$$\alpha_n(t) = \Pi_n(t) \quad , \quad \forall t, n$$

Which means that it's possible to obtain, under these conditions, the posterior probability simply by computing the prior. Moreover, since the result of PASTA property holds $\forall t$, it can also be used when the system is in steady state.

Ergodicity It's a property of stochastic processes $X_h(t)$ in which the ensemble average (average over h) and the time average are the same. In particular, when the system is in steady state, its stochastic behaviour won't depend on time anymore and therefore it can be considered ergodic. So, the only condition for ergodicity is that the system must be able to reach steady state.

The reason why ergodicity is such an interesting property is that it makes possible to compute, when the system is in steady state, the fraction of time spent in a certain state as the prior probability of being in that state and vice versa.