Norwegian University of Science and Technology
Department of Telematics

Contact during exam [Faglig kontakt under eksamen]:
Poul E. Heegaard (94321 / 99286858)

EXAM IN COURSE [EKSAMEN I EMNE]
TTM4110 Dependability and Performance with Discrete event Simulation [Pålitelighet og ytelse med simulering]

Wednesday [Onsdag] 2008-12-03
09:00 – 13:00

The English version starts on page 2.

Bokmålsutgaven starter på side 10.

Hjelpemidler:
C - Graham Birtwistle: DEMOS - A system for Discrete Event Modelling on Simula. Formula sheet for TTM4110 Dependability and Performance with Discrete Event Simulation is attached. Predefined simple calculator.
Graham Birtwistle: DEMOS - A system for Discrete Event Modelling on Simula. Formelsamling i fag TTM4110 Pålitelighet og ytelse med simulering er vedlagt. Forhåndsbestemt enkel kalulator.]

Sensur 2009-week 1

English version[1]

In this exam we will study a ticketing system for railway passengers. To avoid queues at the platforms the railway company has removed the control gates where you previously "checked in, checked out". Instead the passengers will "be in, be out" which means that they are detected by their eIdentity (eID) that is assumed to be stored on the SIM-card in their mobile phones. The SIM-card is assumed to be have an embedded WLAN radio so when they enter the train their eID is sent via a wireless network to the ticket server (TS) which determines the departure station and travel class (first or second) and uses the the SMS server to send a ticket via the mobile network to the passenger. At the destination station when the system detects that the passenger leaves the train the traveller's account is deducted accordingly. It is assumed that each passenger's identity can be determined from the eID. See Figure 1 for an illustration of the ticket system architecture[2].

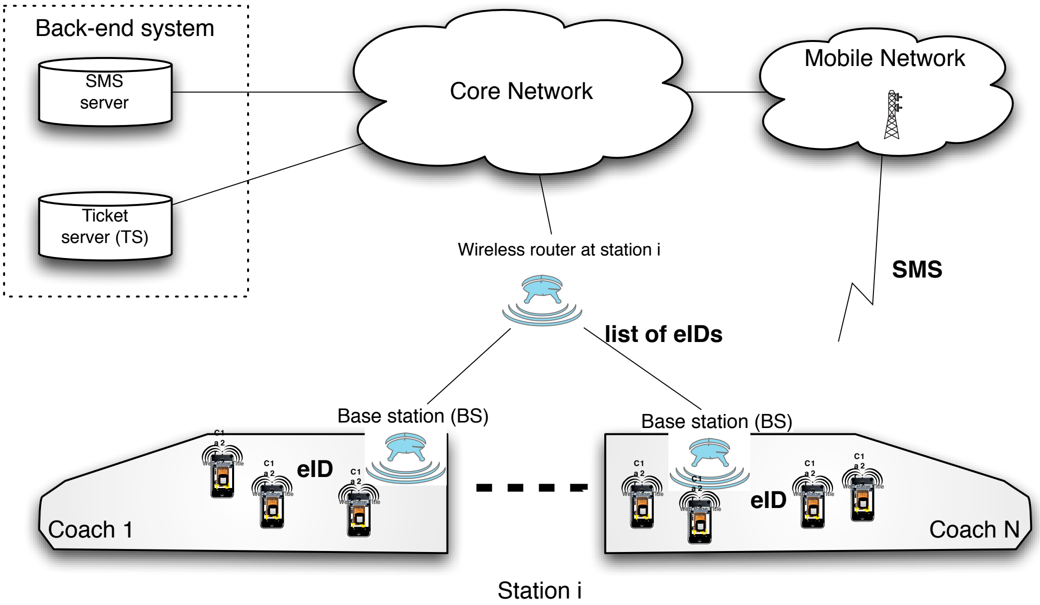


Figure 1: Illustration of the "be-in, be-out" ticket system.

The railway company depends on correct registration of eIDs from all stations a passenger passes along the route. The first eID registered determines the departure station, while the last in a sequence of eIDs is the arrival station. All eIDs in between is from transit stations along the route. If an eID at one of the transit stations, i.e. between the departure and arrival stations, is lost the ticket is free. Furthermore, if three or less eIDs are registered the ticket is free. The price of the ticket depends of the number of stations travelled, so for the railway company even loosing eIDs in the beginning and end of the trip implies lost revenue for the railway company.

The probability of loosing an eID at any station is $p_{\text{eID}} = 0.01$. The ticket costs 1 €/station. An n-station train ride consists of a departure station, $n-1$ transit stations, and an arrival station. The total cost of an n-station train ride is n €.

[1] In case of divergence between the English and the Norwegian version, the English version prevails.
[2] Details regarding confirmation, corrections, reservations, and ticket control are left out for simplicity.

a) Assume that a passenger X travels $n = 4$ stations train ride. What is the probability that the passenger is charged the full price 4 € for the trip? What is the probability that only 3 € can be charged due to eID loss? What is the probability that the ticket is free? What is the expected price of a ticket on a $n = 4$ stations train ride?

- Probability of 4€ ticket : $P(4\,€) = (1 - p_{\text{eID}})^5 = 0.99^5 = 0.9510$
- Probability of 3€ ticket : $P(3\,€) = 2p_{\text{eID}}(1 - p_{\text{eID}})^4 = 2 \cdot 0.01 \cdot 0.99^4 = 0.0192$
- Probability of free ticket: $P(0\,€) = 1 - P(5€) - P(4€) = 0.0298$
- Expected price of ticket: $C_5 = 4 \cdot 0.9510 + 3 \cdot 0.0192 = 3,8616€$

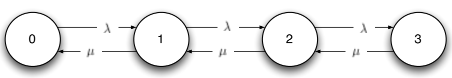
Let's now focus on the SMS server. It has one processor with negative exponentially distributed service times with expectation μ_{SMS}^{-1}=0.5. The messages to be processed arrive at the server according to a Poisson process with intensity $\lambda_{\text{SMS}} = 0.5$. First, we assume that the SMS server has an infinite queuing capacity.

b) Make a state diagram (Markov model) of the SMS server in order to study the message sojourn times in the server. What is used to identify the state of the server in this respect? Use Kendall notation to describe the queuing model this represent. Determine the expected sojourn time in the SMS server (queuing and service). What is the probability of non-zero waiting time, i.e., $P(W > 0)$?

- Markov model, see Figure 6.15, page 167
- System state is the number of messages in the system (in queue and service)
- Kendall: $M/M/1$
- Expected time in system: $E(S) = 1/(\mu_{\text{SMS}} - \lambda_{\text{SMS}}) = 1/(2 - 0.5) = 0.67$
- $P(W > 0) = 1 - p_0 = A = \lambda/\mu$

Now, assume that the SMS server has a finite queuing capacity of $q_{\max} = 2$ and that requests are served in the same order as they arrive.

c) How is this system described in Kendall's notation? Make a state diagram (Markov model) of the SMS server with finite queuing capacity of $q_{\max} = 2$. Determine the steady state probabilities in your SMS server model. Obtain numerical values of the expected number in the system, the variance of the number in the system, and the SMS message loss ratio.

- Kendall: $M/M/1/3/FIFO$
- Markov model 

- Symbolic expression: $p_i = p_0(\lambda_{\text{SMS}}/\mu_{\text{SMS}})^i$ where $p_0 = \sum_{i=0}^{3}(\lambda_{\text{SMS}}/\mu_{\text{SMS}})^i$.
- Numerically:

- $p_0 = 0.7529$
- $p_1 = 0.1882$
- $p_2 = 0.0471$
- $p_3 = 0.0118$

- Expected number in the systems; $E(S) = \sum_{i=0}^{3} i \cdot p_i = 0.1882 + 2 \cdot 0.0471 + 3 \cdot 0.0118 = 0.3176$
- Variance of number in the system: $\text{Var}(S) = \sum_{i=0}^{3} (i - E(S))^2 p_i = 0.3814$
- Message loss ratio:
  - average message arrivals; λ_{SMS}
  - average message losses; $\lambda_{\text{SMS}} \cdot p_3$
  - loss ratio: $\lambda_{\text{SMS}} \cdot p_3 / \lambda_{\text{SMS}} = p_3 = 0.0118$ (equivalent to the time congestion)
- The queue constraint $q_{\max} = 2$ means that the system capacity is 3. However, if you wrongly assume system capacity 2, you loose points on the model but get full score on the reminding if this is done correctly. The correct values for $q_{\max} = 1$ (system capacity 2) are as follows
  - Kendall: $M/M/1/2/FIFO$
  - probabilties $\mathbf{p} = \{0.7619, 0.1905, 0.0476\}$
  - $E(S) = 0.2857$
  - $\text{Var}(S) = 0.2993$
  - loss ratio $= 0.0476$

The passengers arrive at station i according to a Poisson process with intensity $\xi = 0.01$ [min^{-1}]. The trains arrive at station i with interarrival time following a Weibull$-k$ distribution with $\lambda = 0.1$ scale parameter and shape parameter $k = 2$. (See the table in the formula sheet. In addition you need to know $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$, $\Gamma(x+1) = x\Gamma(x)$ and it might be helpful to know that $\int_0^\infty t \cdot e^{-(a \cdot t)^2} dt = \frac{1}{2a^2}$).

d) What is the probability density distribution of the time to next train arrives observed from an arbitrary point in time (forward recurrence time)? What is the expected time until next train arrives observed from an arbitrary point in time? What is the train interarrival time distribution when $k = 1$? What is the expected time to next train in this case?

- Forward recurrence time from Eq (24): $f_T(t) = \frac{1 - F_X(t)}{E(X)}$ where
  - $F_X(t) = 1 - e^{-(\lambda t)^2}$
  - $E(X) = 1/\lambda \cdot \Gamma(1/2 + 1) = 1/2\sqrt{\pi} \cdot 1/\lambda = \frac{\sqrt{\pi}}{2\lambda} = 8.86$ minutes
- Then $f_T(t) = \frac{e^{-(\lambda t)^2}}{\frac{\sqrt{\pi}}{2\lambda}} = \frac{2\lambda}{\sqrt{\pi}} e^{-(\lambda t)^2}$

- Expected time to next train; $E(T) = \int_0^\infty t \cdot f_T(t)dt = \frac{2\lambda}{\sqrt{\pi}} \int_0^\infty t \cdot e^{-(\lambda t)^2} dt = \frac{2\lambda}{\sqrt{\pi}} \cdot \frac{1}{2\lambda^2} = \frac{1}{\sqrt{\pi}\lambda}$

- Numerically, $E(T) = 5.64$ minutes
- With $k = 1$, the Weibull is a negative exponential distribution, $F_X(t) = 1 - e^{-\lambda t}$, and we know that the forward recurrence time follows the same distribution with the same expected value due to the memory-less property of the n.e.d.

The railway company is interested in evaluation of the performance of the wireless router. The company needs to know if the eIDs of all passengers will be registered in the short time a train is at the station. It is found that the wireless router and its communication with the base stations on board the train is the bottleneck so the back-end system (ticket and SMS servers) and the core and mobile networks are ignored. The performance of a router at an arbitrary station i is to be simulated and the effect of the eID losses to be estimated.

e) Describe a random variate generator of the train interarrival times that is Weibull distributed with scale parameter $k = 2$. Justify the choice of method.

- Answer: An inverse transform of the Weibull distribution exits and since this is the most efficient method this is the obvious choice. The inverse transform ($U \in [0, 1]$):

$$\begin{aligned} F(x) &= U \\ 1 - e^{-(\lambda x)^\gamma} &= U \\ e^{-(\lambda x)^\gamma} &= 1 - U = U \\ -(\lambda x)^\gamma &= \log(U) \\ (\lambda x)^\gamma &= -\log(U) \\ \lambda x &= (-\log(U))^{1/\gamma} \\ x &= \frac{1}{\lambda}(-\log(U))^{1/\gamma} \end{aligned}$$

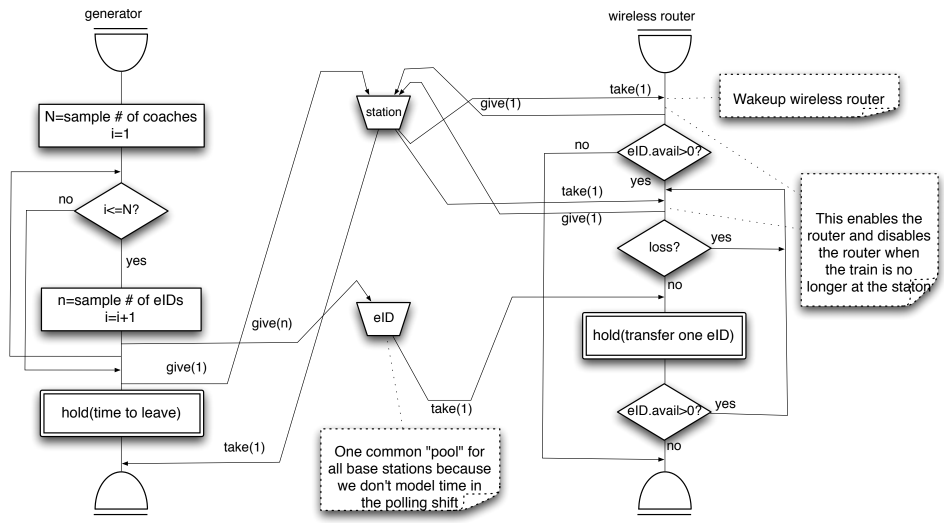
The wireless router at station i is polling each base station (one for each coach on the train) in sequence. It starts with the base station in coach 1, reads all eIDs collected by this base station, and continue to coach $2, 3, \cdots$ and so on until all base stations on the train are polled. The time it takes to transfer the eIDs from the base station to the wireless router depends on the number of eIDs which again depends on the number of passenger in each coach. The probability of loosing an eID is $p_{\text{eID}} = 0.01$. The polling of the base stations starts when the doors are closed. If the train get out of reach of the wireless router before all eIDs are registered the remaining eIDs are lost.

f) Regard one station and one train stop. Identify all random behaviours (times, numbers) that constitutes the dynamics of the simulation model of the base stations and the wireless router at the station. Specify the entities and resources in your simulation model and describe the model by use of activity diagrams. How is the system state defined? What is an *event* and give examples of events in your model? Show clearly in your model how you collect data required to evaluate the performance as requested.

- Randomness:
  - # passengers in each coach when it leaves the station (i.e. the number of eIDs collected)

- time window available from door closes to the train is out of reach by the wireless router.
- transmission processing time (will be multiplied by the # of eIDs)
- # of eID lost
- # coaches <<It is not obvious from the text that this is random >>

- entities
  - generator (of coaches and eIDs)
  - wireless router
- resources
  - station
  - eID



- System state given by number of transferred eIDs and number of remaining eIDs in each coach (if time at station is directly linked to the number of new and leaving passenger this must be included in the system state).
- Events; what changes the system state (ex. eID sent from BS, eID lost, eID received at wireless router)
- Statistics: The model described above simulates on train stop at one station. The eIDs are modelled as BIN. The final report contains information about the "max" value (which will be the number of eIDs generated at the station) and the "now" value is the eIDs that was not served by the wireless router due to either loss on transmission between base station and router, or because the train left the station before the polloing was completed. The loss ratio is therefor "now/max" from the BIN report. To generate a number of i.i.d samples for the loss ratio the replication method must be applied and the simulator rerun with different seed. DEMOS has a built-in method for this purpose.

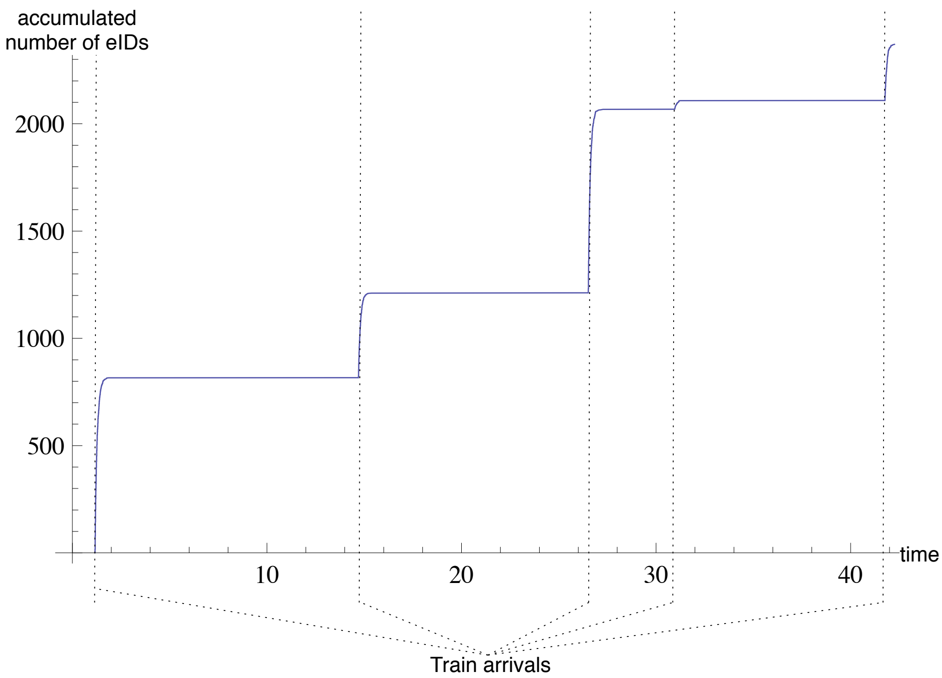


Figure 2: Plot of accumulated number of eIDs at a station as a function of time

Measurements on the wireless router records the time of eID arrivals. In Figure 2 a plot of the accumulated number of eIDs received up to time t is given. The number of eIDs observed is $n = 2980$ and the sum of all observed eIDs interarrival times is $\sum_{i=1}^{n} T_i = 21.9667$ while the sum of their squares is $\sum_{i=1}^{n} T_i^2 = 117.462$.

g) Estimate the average interarrival time $\bar{T}$ and its standard error $S_{\bar{T}}$. Is it reasonable to assume that the interarrival times of eIDs are independent, identically distributed? Justify your answer.

- Estimated average interarrival times; $\bar{T} = \frac{1}{n}\sum_{i=1}^{n} T_i = 0.00737$
- Estimated standard error; $S_{\bar{T}} = \sqrt{\left(\frac{\sum_{i=1}^{n} T_i^2 - n\bar{T}^2}{n-1}\right)/n} = 0.00364$
- The interarrival times for the eIDs when the train is at the station might be i.i.d. while the interarrival times between last eID of the current train and first eID of the next train will follow a different distribution.

The back-end system consists of two servers (ticket and SMS server) that are installed on one processing unit (PU) each. The railway company has three available PUs. Figure 3 shows an example with the ticket server running on processing unit 1 (PU_1) while the SMS server is running on PU_2. The third (PU_3) is a spare unit for both. If PU_1 fails, a reconfiguration of the ticket server occurs and the server is moved to PU_3. Similar for the SMS server if PU_2 fails. No processing takes place during reconfiguration. After a repair the system is not reconfigured (but a re-indexing of the PUs takes place, so the ticket server runs on PU_1, etc) A single PU cannot host more than one server process. The $\mathrm{PU_i}$, ($i = 1, 2, 3$), fails according to a Poisson process with failure rate λ_{PU}. They are repaired by one shared repairman with repair time that is negative exponentially distributed (n.e.d.) with expected time μ_{PU}^{-1}. The reconfiguration time is n.e.d. with rate γ.

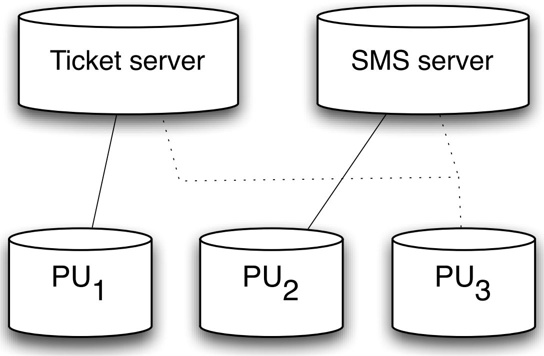
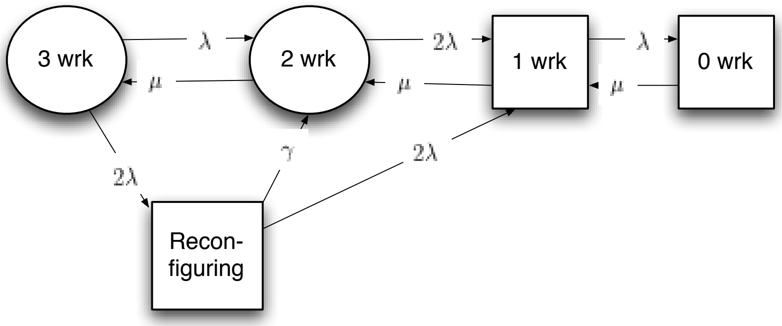


Figure 3: Subsystem of ticket and SMS servers provided on three processing units

h) Make a model of the subsystem that consists of the ticket and SMS servers, and the three PUs with the objective to determine the (simultaneous) availability of the two services. Identify the system failure states where either ticket or SMS servers are not working. Establish a set of equations to determine the steady state probabilities in your model (the equations should not be solved). Given that the state probabilities are known, what is the steady state availability for this subsystem? What is the system failure rate?

- Markov model in the following figure ("x wrk" means that x PUs are working);



- If the standby unit fails the number of working units are reduced by one, if one of the primary units fails we need to reconfigure the server process and during the reconfiguration the system is not available.
- When more than one until has failed, the system has failed.
- The failure states are "Reconfiguration", "1 wrk" and "0 wrk". (below, states are indexed according to the number of working servers and "r" means reconfiguring.)
- Equations
  - $3\lambda p_3 = \mu p_2$
  - $(\mu + 2\lambda)p_2 = \gamma p_r + \mu p_1 + \lambda p_3$

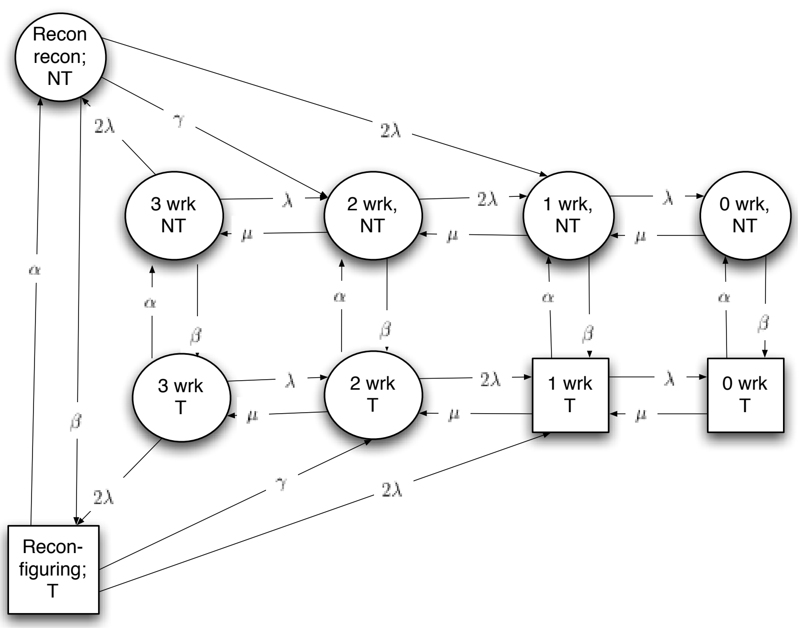
  - $(\gamma + 2\lambda)p_r = 2\lambda p_3$
  - $-(\lambda + \mu)p_1 = 2\lambda p_r + 2\lambda p_2 + \mu p_0$
  - $p_3 + p_2 + p_1 + p_0 + p_r = 1$

- Steady state availability: $A = p_3 + p_2$
- System failure: $\Lambda = 2\lambda p_3 + 2\lambda p_2$ or alternatively $\Lambda = \gamma p_r + \mu p_1$

The unavailability of the ticket and SMS servers is not critical in the periods where there are no trains at any stations, i.e. no eIDs are sent, i.e. if the back-end system is failed while no trains are on a station no eIDs are lost. It is of interest to find the fraction of the time when eIDs are lost due to the back-end system failure. Let's assume that the time between arrivals of any train at any station is n.e.d. with expectation β^{-1}. The time at the station when eIDs are transferred is assumed to be n.e.d. with expectation α^{-1}.

i) Extend your dependability model from the previous task to include the arrival and departure of trains at a station (i.e. start/stop transmission of eIDs). Define the system failure states to be the states where eIDs are lost.

- Markov model in the following figure;



- Since the train arrivals and failure and repair processes can be assumed to be independent, the system may also be analysied by product form solution. This means that $p_{x,y} = p_x \cdot p_y$ where $x = 3, 2, 1, 0, r$ and $y = \mathrm{NT}, \mathrm{T}$. Train arrival and departure from the station can be modelled by the figure below which gives: $1 - p_{\mathrm{T}} = p_{\mathrm{NT}} = \frac{\alpha}{\beta+\alpha}$.