香港城市大學
City University of Hong Kong

396EM Airline Operations and Scheduling/ 6075MAA Airline Scheduling and Operations

Lecture 6
Input Preparation and Basics of SIMUL8

Developed & Revised :
Dr Yang Dai (Coventry)
Rossella Lau (Scope)
Presented: Kinki Leung

專業 創新 胸懷全球
Professional · Creative For The World

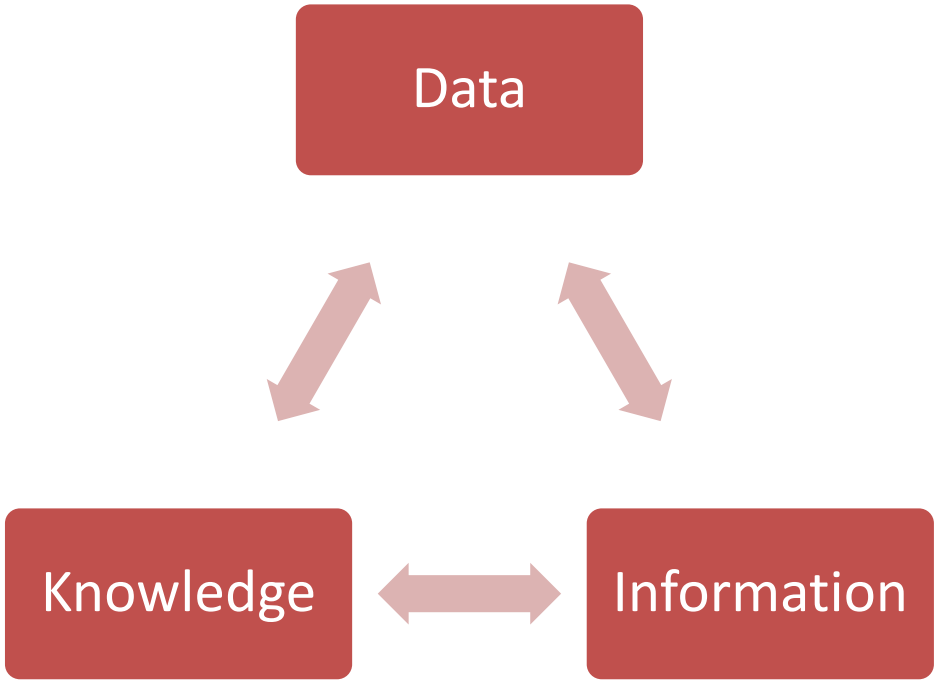
SCOPE School of Continuing and Professional Education 專業進修學院
Coventry University

Outline of the Lecture

- Data, Information and Knowledge
- Categories of Data
- Data collection and methods
- Simul8's basic building blocks
  - Properties
  - Arrival time
  - Graphics
  - Routing
  - Label

Difference between data and information

- Information is typically viewed as data with interpretation and that have been analysed for some purpose.
- How about knowledge?



Introduction

- Simulation is known as a "Data Hungry" technique therefore enough data should be collected and provided for modelling purposes.

- The validity of simulation models depend on the accuracy of the input data.

- Data should be precisely gathered by applying appropriate data collection tools and techniques.

Data Definition

- Data can be defined as a non identifying collection of facts from which conclusions may be drawn.
- Or Information about how our services are accessed by clients

	A	B	C	D	E	F	G
1	AHRQ Prevention Quality Indicators						
2	Dehydration Admision Rate (PQI 10)						
3							
4	Counties/Numbers highlighted in GREEN are significantly lower than the National Average.						
5	Counties/Numbers in RED are significantly higher than the National Average.						
6							
7	County Name	Cases	Population	Crude Rate	Risk Adj. Rate LCL	Risk Adjusted Rate	Risk Adj. Rate UCL
8	Adair	79	13,774	5.74	4.62	5.19	5.76
9	Allen	28	14,299	1.96	1.41	2.00	2.59
10	Anderson	12	15,453	0.78	0.25	0.84	1.42
11	Ballard	8	6,538	1.22	0.24	1.03	1.83
12	Barren	102	31,112	3.28	2.56	2.93	3.31
13	Bath	15	8,943	1.68	0.84	1.55	2.26
14	Bell	122	23,055	5.29	4.52	4.96	5.41
15	Boone	68	78,320	0.87	0.85	1.14	1.42
16	Bourbon	20	15,245	1.31	0.70	1.26	1.81
17	Boyd	32	39,393	0.81	0.39	0.72	1.06
18	Boyle	32	22,387	1.43	0.88	1.34	1.79
19	Bracken	18	6,700	2.69	1.78	2.63	3.47
20	Breathitt	40	12,381	3.23	2.84	3.50	4.15
21	Breckinridge	23	15,006	1.53	0.94	1.50	2.07
22	Bullitt	23	52,112	0.44	0.23	0.58	0.93
23	Butler	9	10,366	0.87	0.18	0.86	1.54
24	Caldwell	13	10,281	1.26	0.39	1.00	1.61
25	Calloway	28	29,186	0.96	0.50	0.90	1.30
26	Campbell	54	66,477	0.81	0.53	0.80	1.07
27	Carlisle	5	4,215	1.19	0.00	0.93	1.89
28	Carroll	20	7,950	2.52	1.77	2.56	3.35
29	Carter	18	21,160	0.85	0.37	0.85	1.34
30	Casey	47	12,646	3.72	2.72	3.30	3.89

PQI 1 / PQI 2 / PQI 3 / PQI 5 / PQI 7 / PQI 8 / PQI 9 / **PQI 10** / PQI 11 / PQI 12 / PQI 13 / PQI 14 / PQI 15

Data Requirements

- Split into three types (Pidd 2003)
  - Contextual data
    - background information that provides a broader understanding of an event, person, or item.
  - Data for model realization
  - Data for model validation

Contextual data

- In order for the modeller and clients to develop a thorough understanding of the problem situation some data needs to be available, for instance, a layout diagram, basic data on process capability and beliefs about the cause of problems that are being experienced.

- At this stage large data collection exercises should be avoided, if possible, since the data are only required for developing an understanding and are generally not needed for detailed analysis.

- These data are very much part of the conceptual modelling process because they are necessary for the development of the conceptual model.

Data for model realization

- In moving from the conceptual model to a computer model many data are required, for example, detailed data on cycle times and breakdowns, customer arrival patterns and descriptions of customer types, and scheduling and processing rules.

- It may be necessary to carry out a detailed collection exercise to obtain these data.

- These data are directly identified from the conceptual model, since this describes all the components of the model and the detail associated with them.

- As such, identification of the data required for model realization is an output of the conceptual modelling process

Data for model validation

- Determining that the contextual data and the data required for model realization and validation are sufficiently accurate for the purpose at hand.

  - White-Box Validation
    - check the internal functioning of the system (requires programming knowledge)
  - Black-Box Validation
    - high level of testing that focuses on the behavior of the software (doesn't require programming knowledge)
  - Experimentation Validation
  - Solution Validation

Obtaining Data

- Some data are immediately available, others need to be collected

Categories of Data Availability and Collectability.

Category A	Available
Category B	Not available but collectable
Category C	Not available and not collectable

Category A

- Category A data are available either because they are known or because they have been collected previously.

- Example, service times and arrival rates in an airline

Category B

- Data that fall into this category often include service times, arrival patterns, machine failure rates and repair times, and the nature of human decision-making.

- It might require interviews with subject matter experts such as staff, equipment suppliers.

Category C

- Category C data are not available and cannot be collected. These often occur because the real-world system does not yet exist, making it impossible to observe it in operation.

- Unfortunately, category C data are not uncommon.

- Many simulation studies involved category C data.

Data accuracy

- Data collection exercises should be devised to ensure that the sample size is adequate and as far as possible recording errors are avoided

- What mechanisms can be put in place to monitor and avoid inaccuracies creeping into the data collection?

Data Collection

- Data collection is simply a term used to describe a process of preparing and collecting data. There are various methods of data collection such as questionnaire, personal interviewing, observation, and stop watch.

- The purpose of data collection is to:
  - Obtain information to keep on record.
  - Make decisions about important issues.
  - Pass information on to others.

Bootstrapping

- Bootstrapping is particularly useful when there is only a small sample of data available.
- In effect, it is a way of extending the use of traces.
- Bootstrapping has a wider application in simulation than just representing unpredictable variation in the inputs to a model.
- It is also useful in the analysis of simulation output and in model validation.

Bootstrap Samples from Trace Data

Original Data			
Call arrival time (min)	Call inter-arrival time (min)	Bootstrap 1: inter-arrival time	Bootstrap 2: inter-arrival time
0.09	0.09	0.67	0.02
0.54	0.45	1.80	0.09
0.99	0.45	0.02	0.45
1.01	0.02	0.09	0.24
1.25	0.24	0.45	0.78
1.92	0.67	0.67	1.80
2.14	0.22	0.74	0.45
2.92	0.78	0.74	0.22
3.66	0.74	0.78	0.67
5.46	1.80	0.24	0.45

Preparing for Data Collection

- Key activities for data collection:
  - Schedule data collection visits.
  - Prepare data collection forms (set of questions) needed for field visits.
  - Prepare information materials and tools for data collectors.
  - Arrange for regular communications.

Data Collection: Quantitative Tools

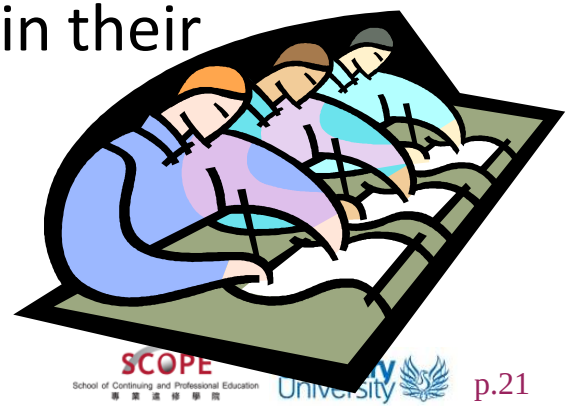
- A number of tools and techniques can be used to collect data.
- Most popular tools and techniques used in data collection are:
  - Structured interview
  - Questionnaire
  - Observation
  - Stopwatch

Advantages of Quantitative Tools

- Lower administrative costs.
- More straight forward analysis of data.
- Simple and short questions.
- Non-threatening.
- Wide geographic distribution possible.

Questionnaire

- It relies on obtaining information directly from individuals by raising a number of questions.
- This technique is widely used for collecting data in survey method.
- These questionnaires may be written in a manner that allows the researcher to know about the behaviour of the addressee, their attitudes to their work environment and how they perceive things in their organisation.



Questionnaire Mechanics

- Pay attention to format, cover letter.
- List questions in descending order of usefulness.
- Put related questions together; use transitions.
- Use branching questions.
- Pay attention to questionnaire appearance.
- Carefully plan mailing.
- Special care for Web questionnaires.

Interview

- An interview can be defined as a focused conversation in which one person asks questions (as interviewer) and other answers them (as respondent).

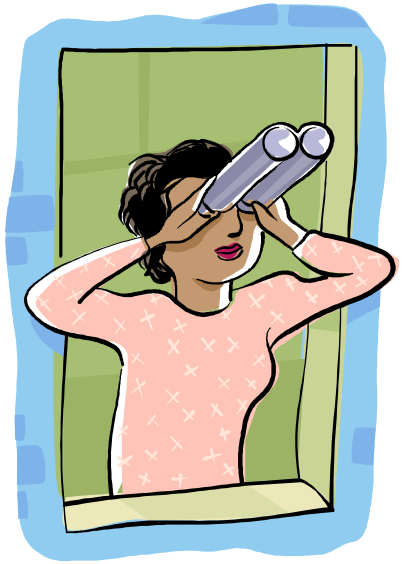
- Interviews can be categorised in a structured or unstructured way:
  - A structured interview is a questionnaire administrated in a face-to-face setting.
  - An unstructured interview is where the interviewer raises a certain topic for the interviewee to comment on.

Interviewing Mechanics

- Simple language / avoid jargon.
- Non-leading questions with only one interpretation.
- No double-barreled questions.
- Prepare carefully ahead of time.
- Use probes for clarification.
- Avoid suggesting answers, agreeing, and disagreeing.
- Use silence.

Observation

- Observation relies on the researchers' ability to gather data though their senses - and allows researchers to document actual behaviour rather than responses related to behaviour



- Or gathering information by watching people, events, or geographic areas.

Observation Mechanics

- Use observation recording sheet; take field notes.
- Carefully choose site, dates, and vantage point.
- Convert field notes to expanded accounts soon.

Stop Watch

- A watch that can be instantly started and stopped by pushing a button and used to measure an exact duration of time.
- It helps if the stopwatch is held by a colleague rather than an outsider.

- Direct stopwatch measurements are appropriate for operations lasting a few minutes with extensive manual intervention.
- It is used to record process times for one or more than one resource (operator or machine).

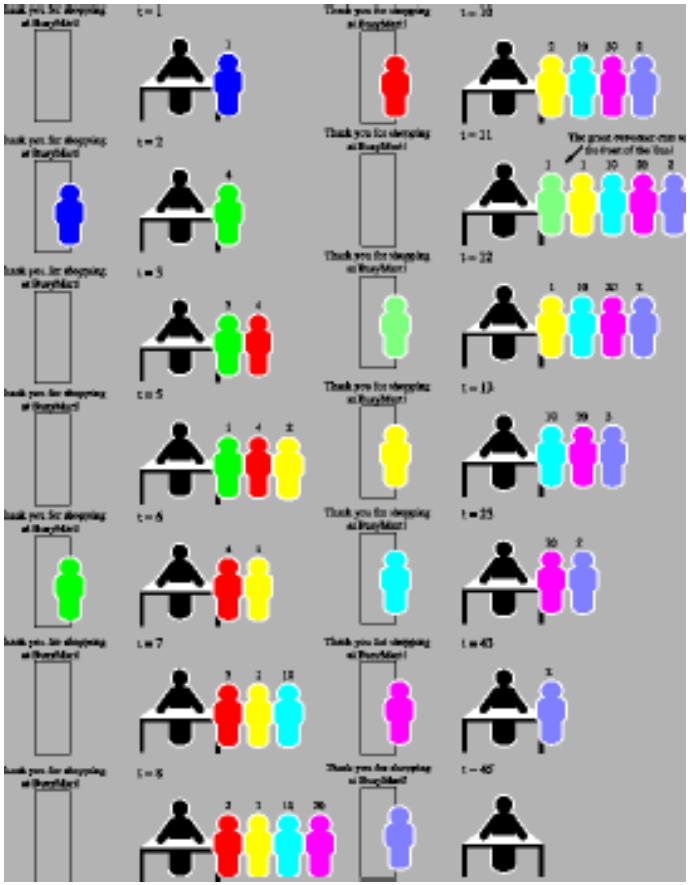
Data collection forms

Work type	Work Item #	Arrival time	Server#	Service Time	……

Inter-arrival Time Modelling

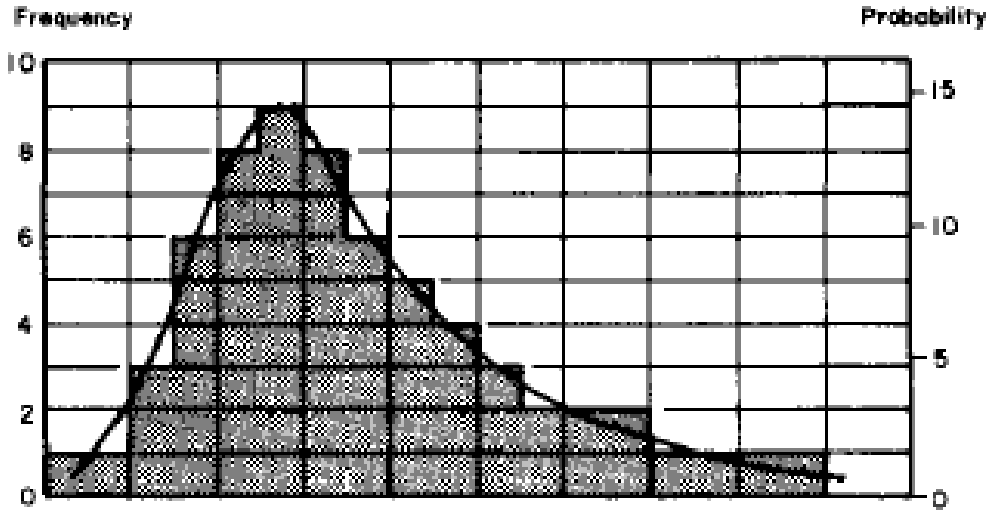
- The simplest arrival process is one where we have completely regular arrivals (i.e., the same constant time interval between successive arrivals) – impossible!

- In a random stream which is more applicable in real life phenomena's, successive customers arrive after intervals which independently are randomly distributed.



Distribution Definition

- An arrangement of values of a variable showing their observed or theoretical frequency of occurrence.



Score	Frequency	Cum Frequency
40	1	20
35	1	19
34	3	18
31	1	15
29	2	14
28	1	12
27	2	11
26	2	9
24	3	7
23	1	4
22	2	3
17	1	1

Families of Probability Distributions

- The number of different probability distributions is unlimited. Statistical distributions are defined by some mathematical functions or probability density function (PDF).

- Certain families of distributions give good approximations of the distributions of many random variables.

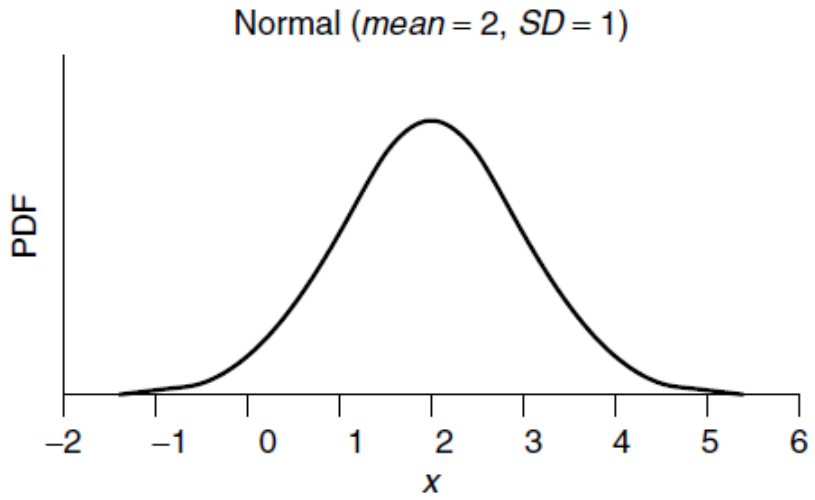
- Important families of continuous distributions include:
  - Normal(Gaussian), Negative Exponential, Gamma (Erlang), Lognormal, Weibull, Extreme Value…

- Important families of discrete distributions include:
  - Binomial, Multinomial, Poisson, Hyper-Geometric, Negative Binomial…

Continuous distributions

- The normal distribution is an example of a continuous distribution. It is specified by two parameters: mean (its location) and standard deviation (its spread)

- Two other continuous distributions, which are commonly used in simulation models, are the negative exponential and Erlang distributions

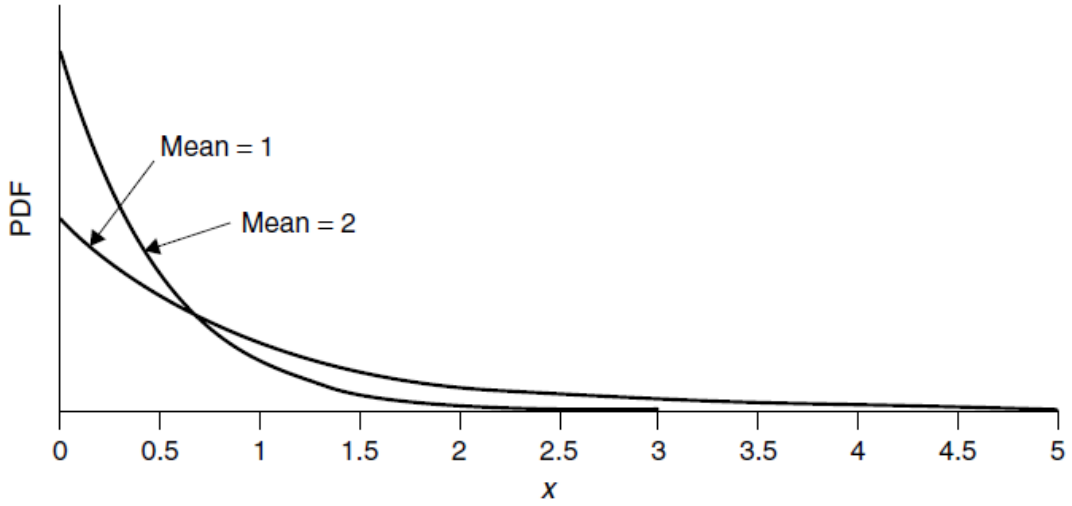


Negative Exponential Distribution

- It is used to sample time between events

  (e.g., inter-arrival time)

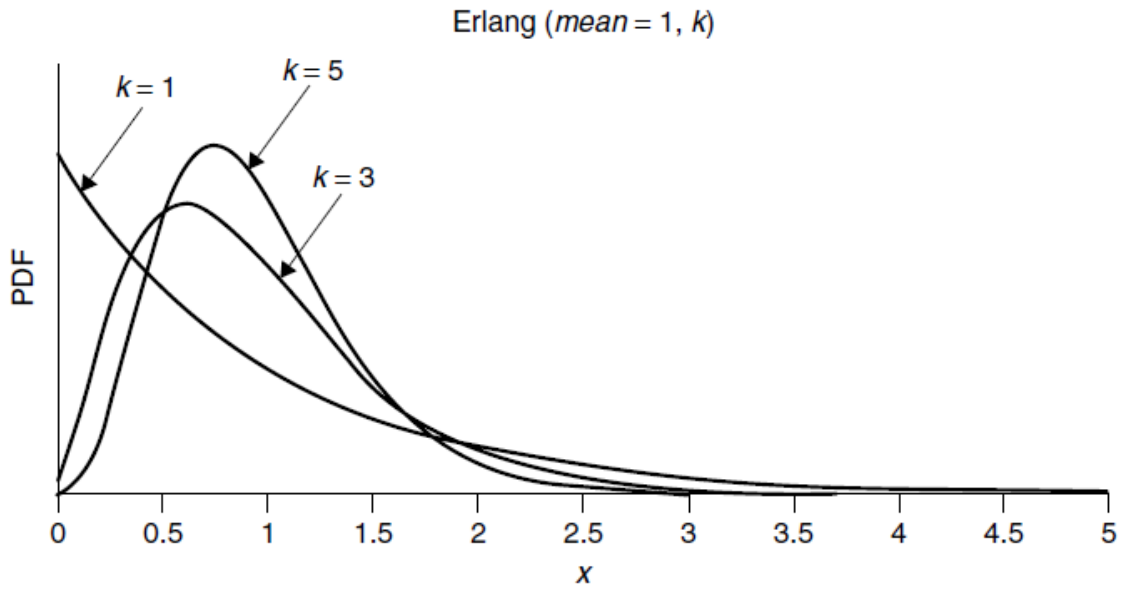
  - It gives a high probability of sampling values of x close to zero and a low probability of sampling higher values
  - It implies that the majority of arrivals occur quite close together with occasional long gaps
  - Close relationship with Poisson Distribution



Erlang distributions

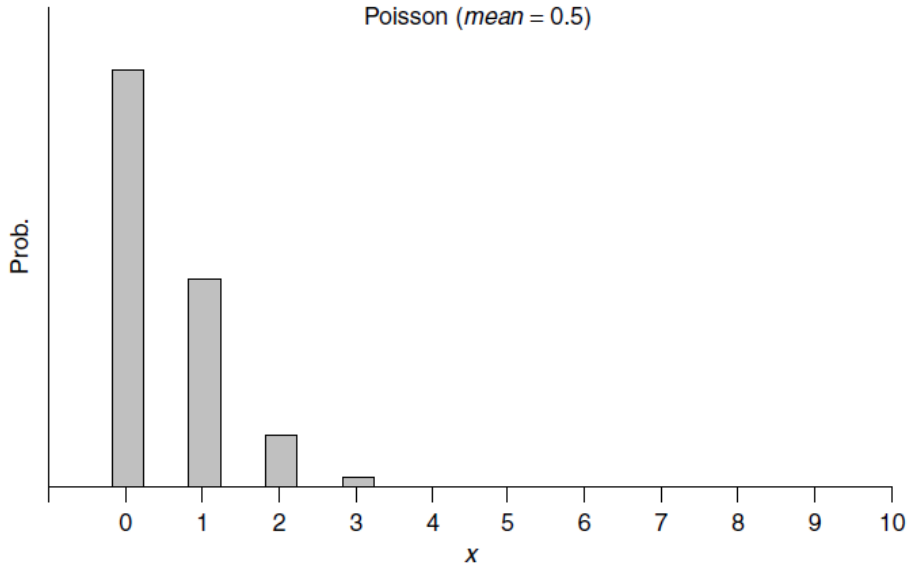
- The Erlang distribution is used to represent the time to complete a task and also the inter-arrival time, particularly if the arrivals are not closed (e.g., priority banking in public estate area)

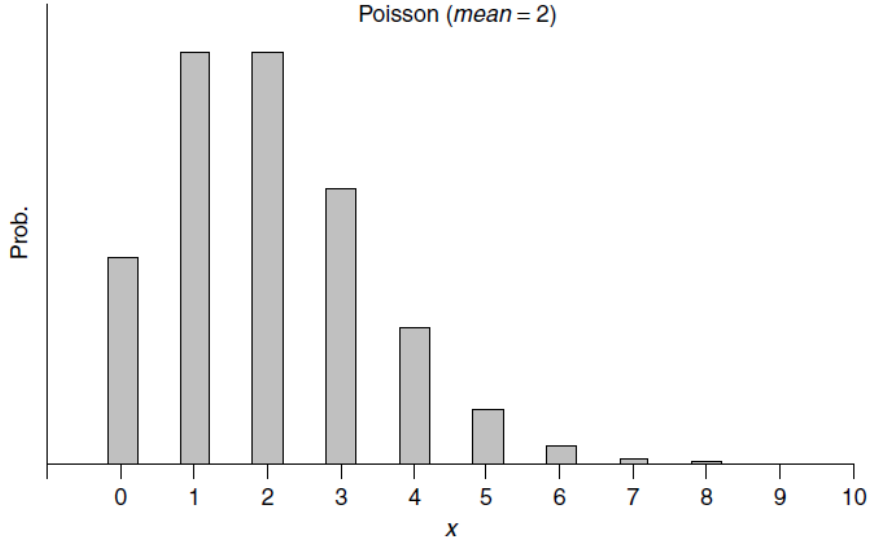
- For values of k greater than one it overcomes the limitation of the negative exponential distribution by giving a very low probability of near z



Poisson Distribution

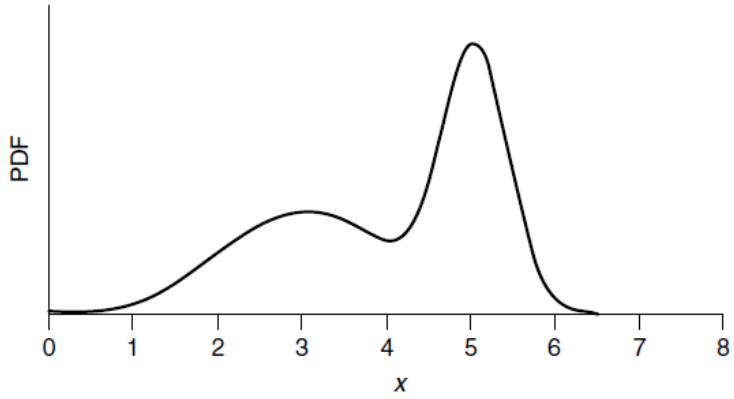
- A discrete distribution, used to sample number of events in an interval of time
  - The arrival is random with a mean (per time unit; p.m., p.h.)
  - A close relationship with Exponential:
  - Number of arrival is λ in Poisson,
    inter-arrival is 1/ λ in Exponential

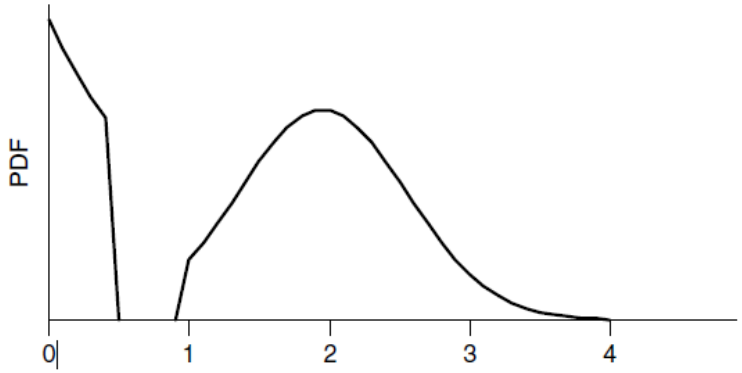




Approximate Distribution

- How about when distribution does not fall into the given distribution methods / models?





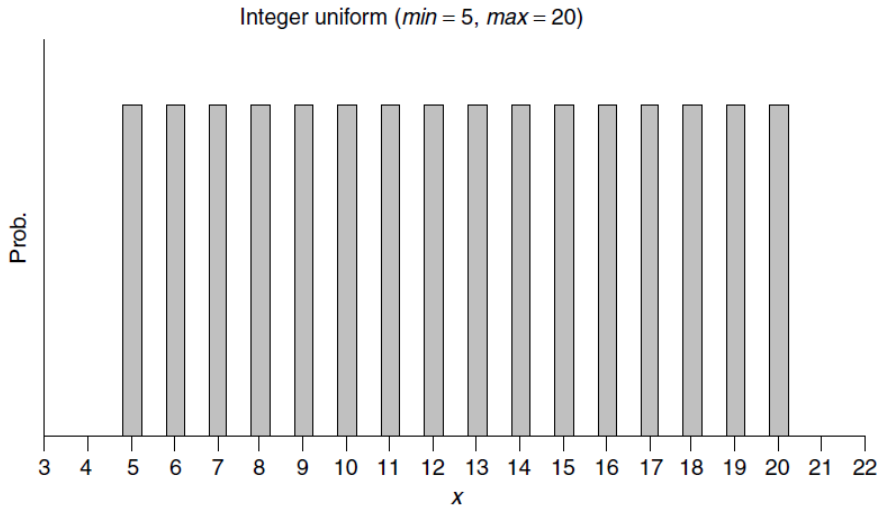
- Use approximate distributions
  - Uniform
  - Triangular
  - Empirical
  - ……

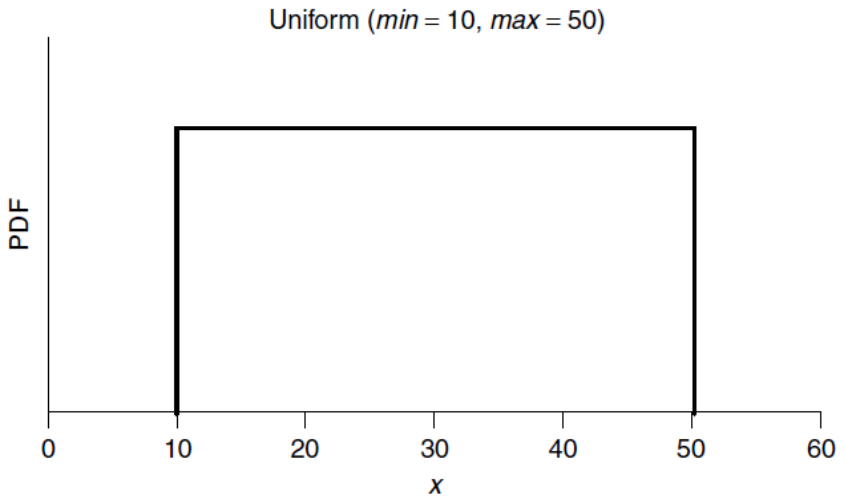
Uniform distribution

- A continuous distribution bounded at the upper and lower limits. It is useful for situations where there is a random occurrence between the upper and lower values, but where little else is known.

- No sample values occur outside this range.

- Tends to be used when it is not possible to collect data about the actual situation but there is knowledge about the broad range of the data.

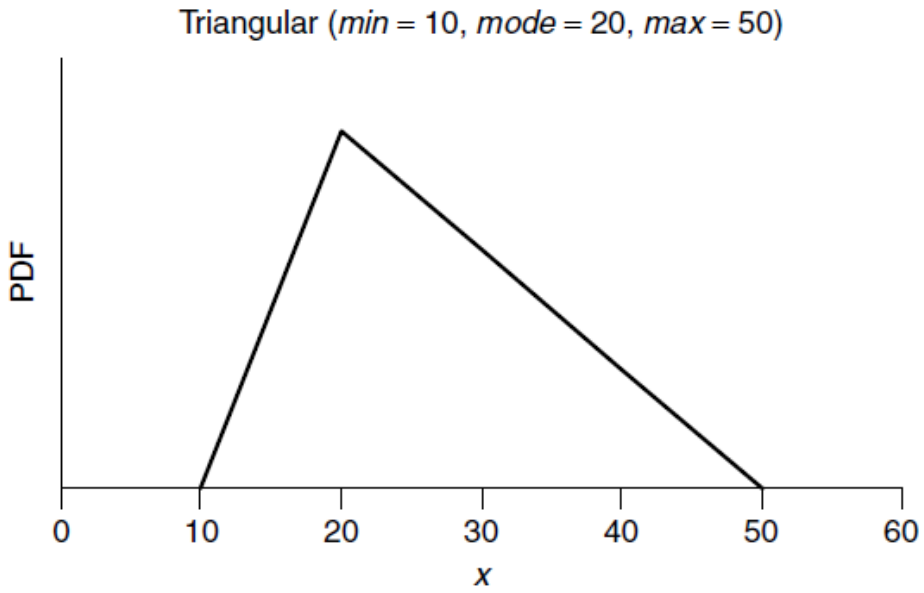




Triangular distribution

- The triangular distribution provides a slightly more sophisticated approximation than the uniform distribution by including a third parameter, the mode, or most likely value.

- The triangular shape can be quite similar to the shape of an Erlang distribution



Traces

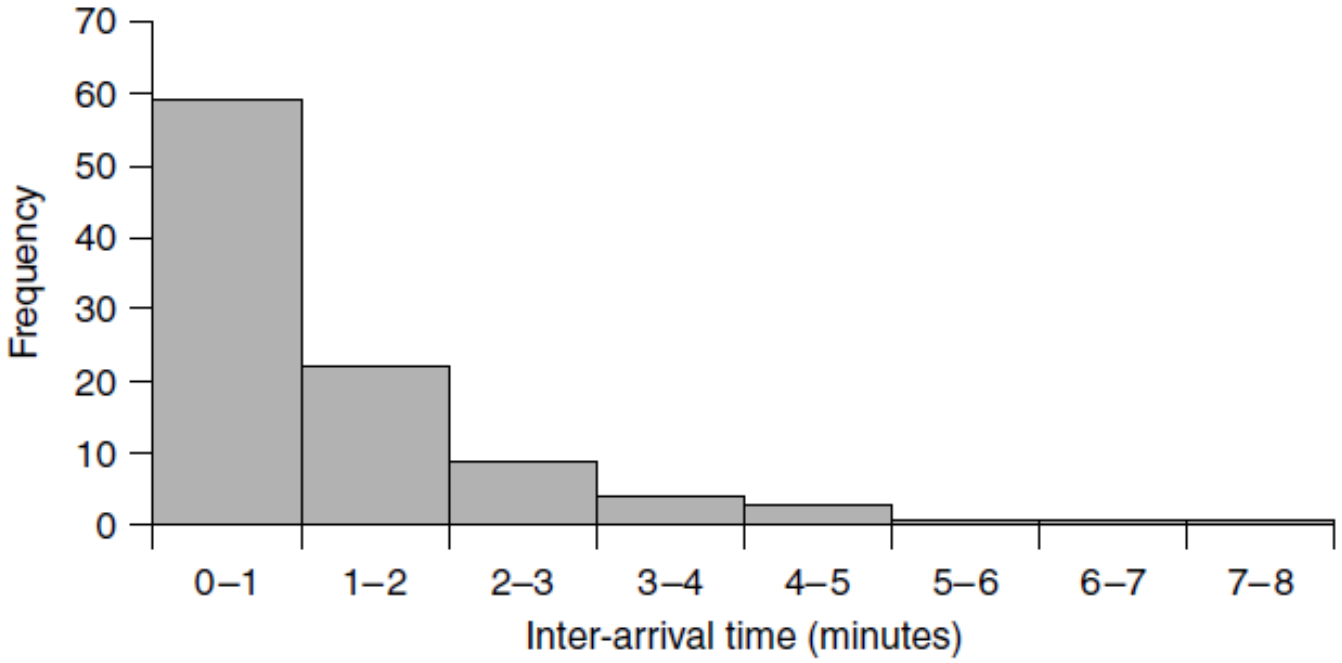
- A trace is a stream of data that describes a sequence of events
- Traces are normally obtained by collecting data from the real system.
- Automatic monitoring systems often collect detailed data about system events and so they are a common source of trace data.
  - In a call centre, the call handling system collects data on call arrivals, call routes and call durations.
  - Factory monitoring systems collect data on machine failures, down times and fault types

Example of a Trace: Call Arrivals at a Call Centre.

Call arrival time (minutes)	Call type
0.09	1
0.54	1
0.99	3
1.01	2
1.25	1
1.92	2
2.14	2
2.92	3
3.66	3
5.46	2
.	.
.	.
.	.

Empirical distribution

- An empirical distribution shows the frequency with which data values, or ranges of data
  - Values, occur and are represented by histograms or frequency charts.
  - They are normally based on historic data.

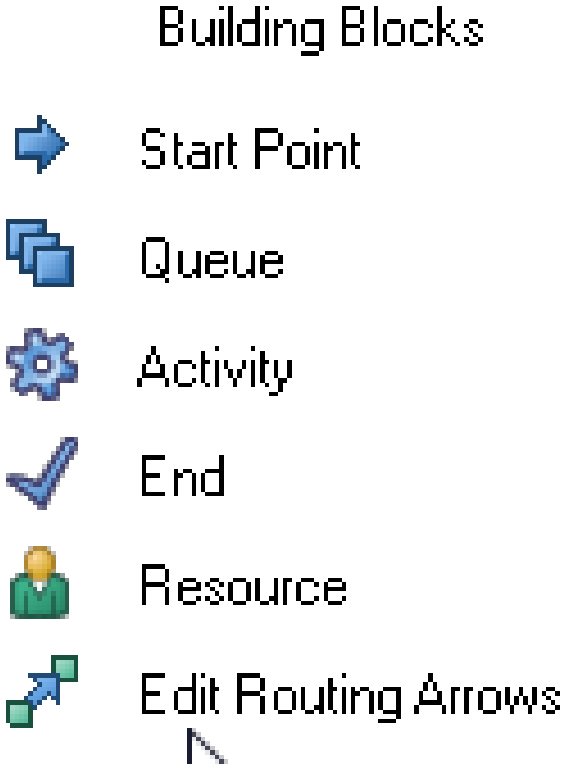


Basic components in SIMUL8

Sources: https://www.simul8.com/

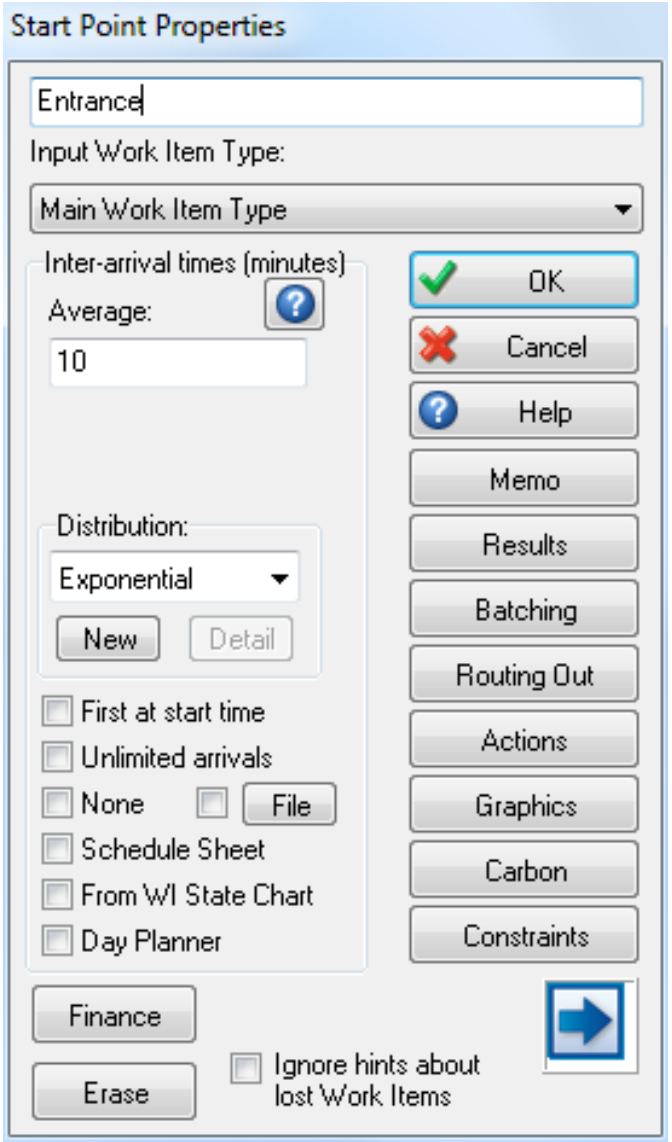
Building Blocks

- SIMUL8 provides the following basic building blocks to construct a simulation model:
  - Start Point
  - Queue
  - Activity
  - End
  - Resource
  - Routing Arrows

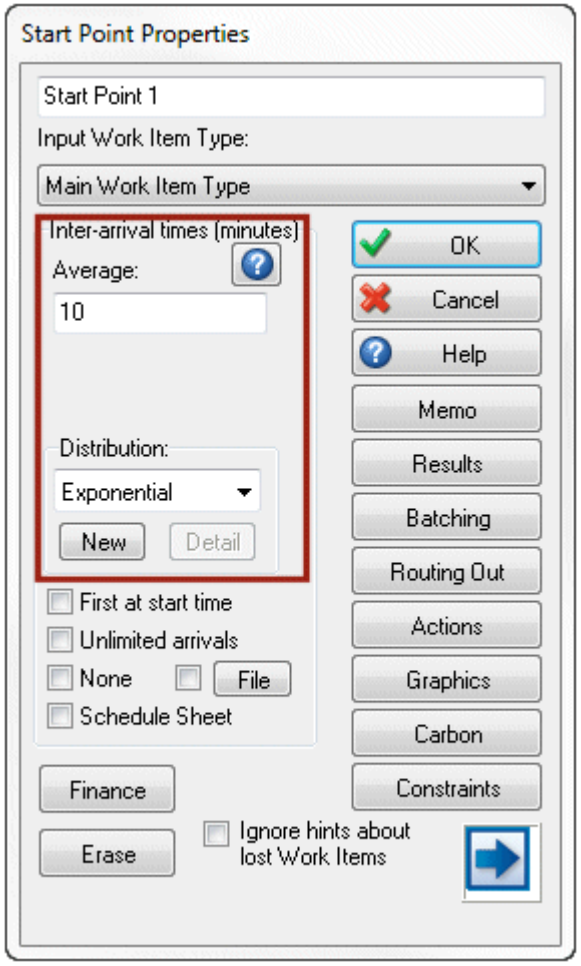


Start Point

- A Start Point is a work entry point where work to be done appears in your model for the first time.
- It uses to create entities to be fed into the simulated system. (i.e. this module imitates entities, their types and inter-arrival to the simulated system).
- You can have as many “Start Points” as you like. Each can feed in Work Item Types of different types (if required).



Arrival time distribution

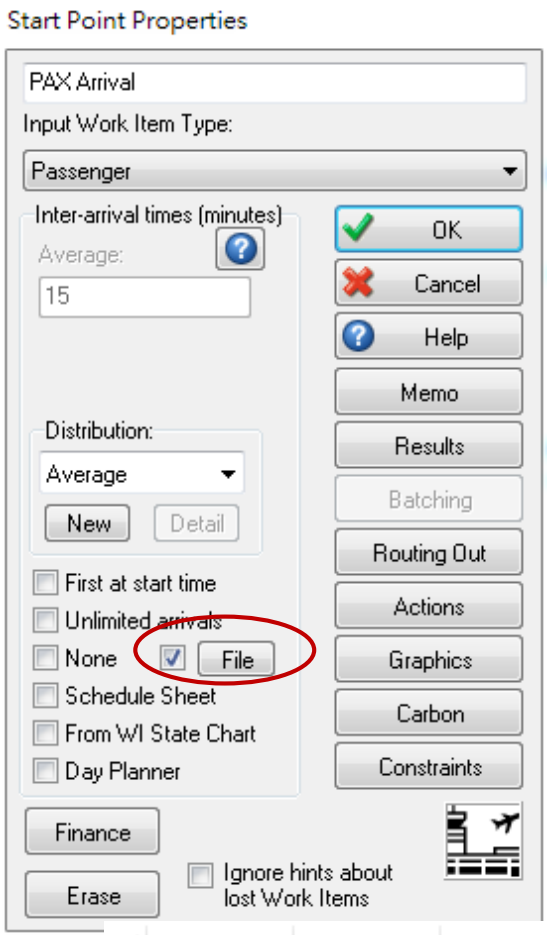


- Distribution Type
- Use the drop-down box to select the distribution you want to use to generate the time for the particular activity that the timing panel relates.
- Click the down-arrow button to see a list of available distributions (select the one you want) or click the NEW button to create a new distribution of your own.
- A number of standard distributions are available in SIMUL8:
  - Exponential
  - Fixed
  - Normal
  - Uniform
  - Average
  - Erlang
  - Triangular
  - .....

https://www.simul8.com/support/help/doku.php?id=features:timing_we

https://www.simul8.com/support/help/doku.php?id=features:distributions

Arrival time from csv file



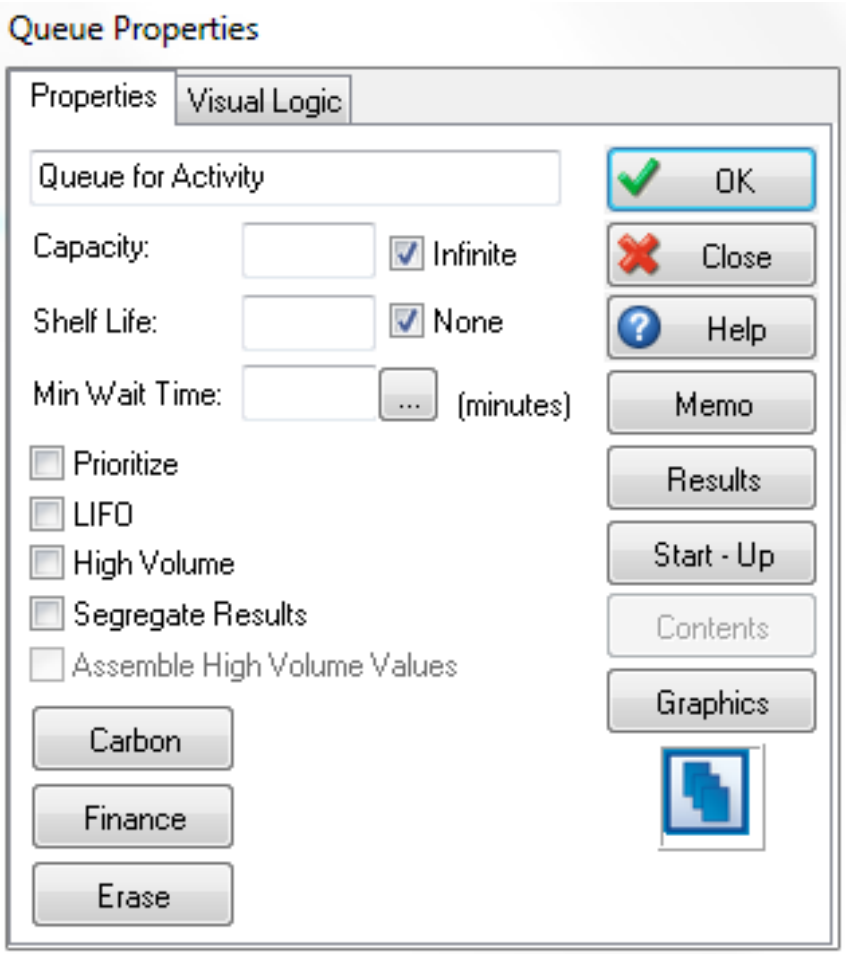
	A	B	C
1	Lab9-2_AirportA_FromCSV		
2			Airline Name
3		5	1
4		1	2
5		11	1
6		12	1

- Arrival Times from CSV File
- This feature reads from a csv file the exact times Work Items should be generated at a Start Point. It can also assign label values to the Work Items.
- Check the 'File' option on the Start Point dialog and then select the file to use.
- This feature requires the csv file to be formatted in the following way:
- Each row corresponds to a Work Item
- data must start on the 3rd row. The first cell in row 1 and 2 must be populated - row 1 could contain the simulation name and row 2 the column headings
- the time the Work Item enters the simulation must be in the 2nd column
- label values (if required) are contained in columns 3 and onwards. Row 2 must contain the label title exactly as defined in SIMUL8.
- The csv file must be closed before running the simulation.

https://www.simul8.com/support/help/doku.php?id=features:arrivalscsv&s[]=csv&s[]=file

Queue

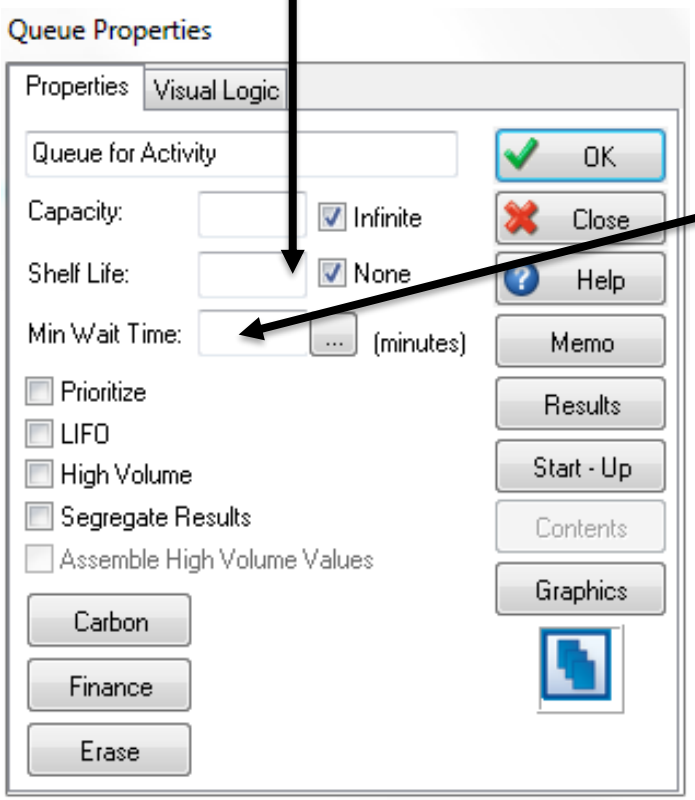
- A queue is a place where work to be done can wait until appropriate resources or work centres (Activities) are available.
  - Entities can not be processed by a work centre without passing a queue. A Simul8 warning pop up box appears in case of violating this condition.



Queue Components

Shelf Life Specify a time work items should not stay longer than in the queue. Number of time units before items in this queue become "out of date". Out of date items can be removed by Activities that have "expired only".

Capacity
Limit the capacity of the queue.



Min Wait Time (Delay)
Force work to stay in the queue for a specified amount of time. This value can be a fixed number, variable or label.

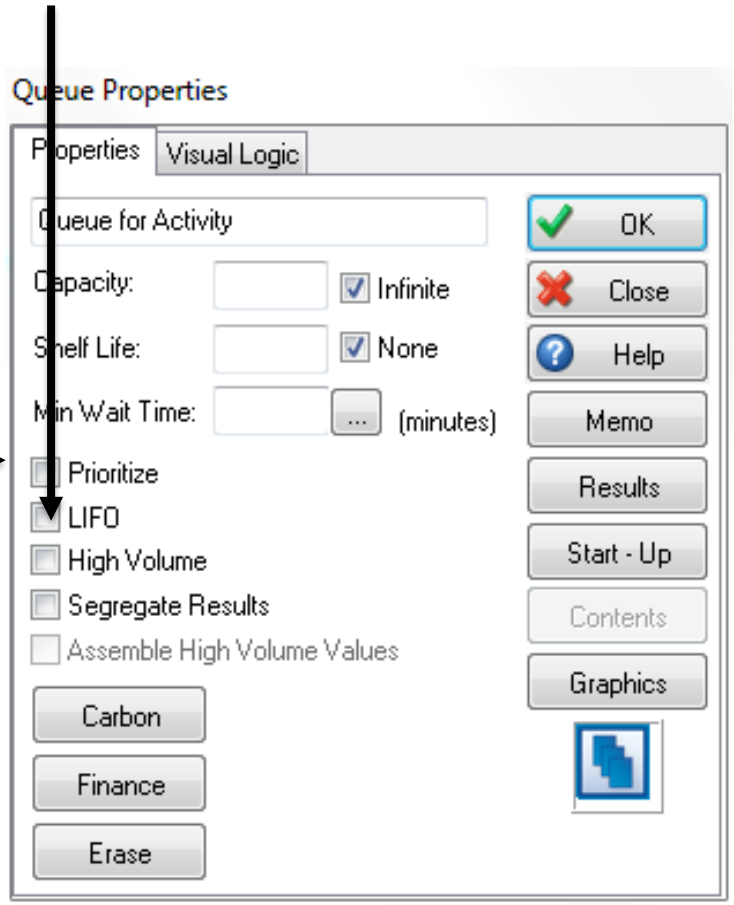
Further Queue Components

LIFO

Normally, work items join the back of a queue and don't leave until they reach the front, FIFO. With LIFO checked work that enters the queue goes to the front of the queue. Last In First Out.

Prioritise

Work can be prioritised in a queue based on a label value. The higher the value the closer to the front of the queue the work will be.



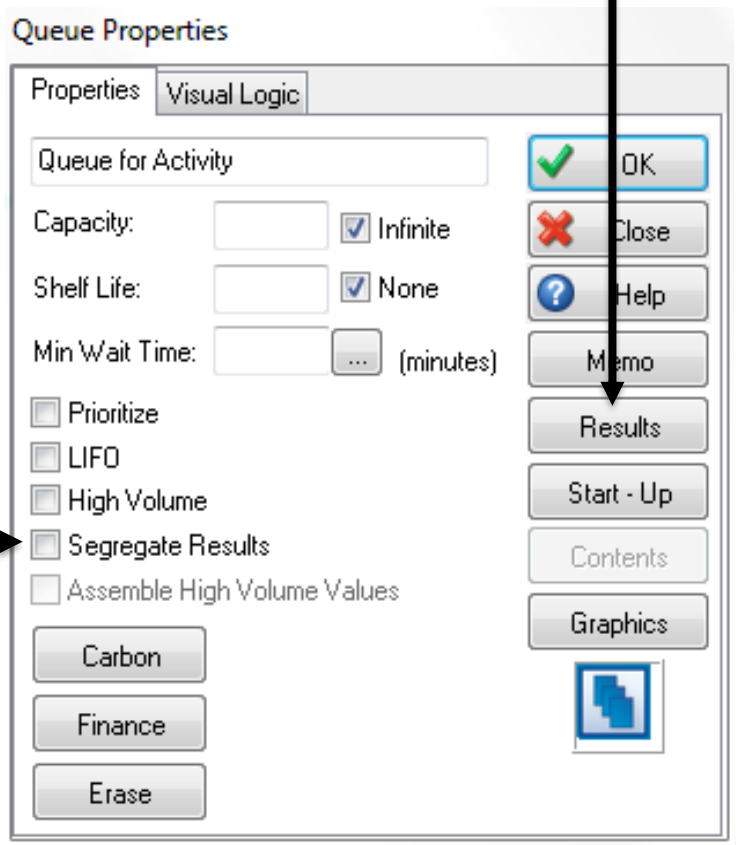
Results of Queue Simulation

Results

Collect various results about the queue in numeric and graphical format.

Segregate Results

Some of the results information for queue can be segregated by label to let you see, for example, how long different types of work spend in the queue.



Queue Graphics / Animation

Queue

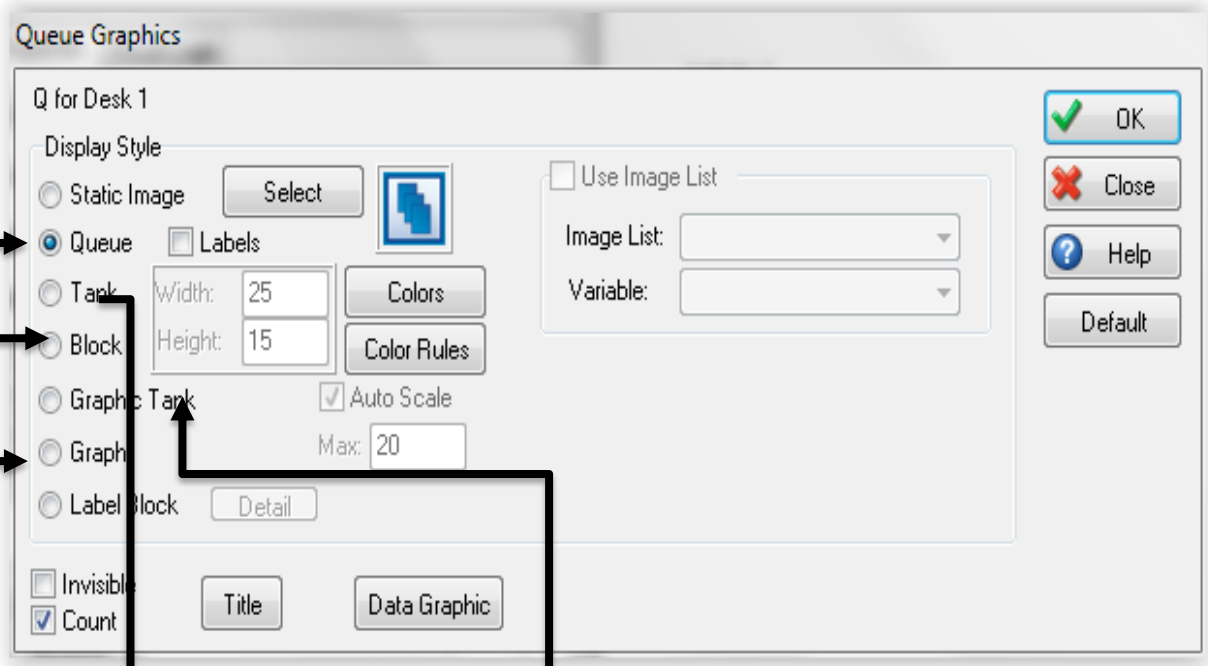
A queue displays all work items in the queue on screen.

Block

Each work item is represented by one block.

Graph

Display the contents of a queue as a graph.



Tank

A tank display shows a small box which fills as the queue fills. If the queue has a maximum size set then the tank will appear full at this size and will be filled proportionately as the queue is partly full. If the queue is of infinite size then one screen dot will be used to represent each work item in the queue.

Graphic Tank

Similar to the Tank Display option Graphic Tank shows an icon that fills as the queue fills but is more graphical.

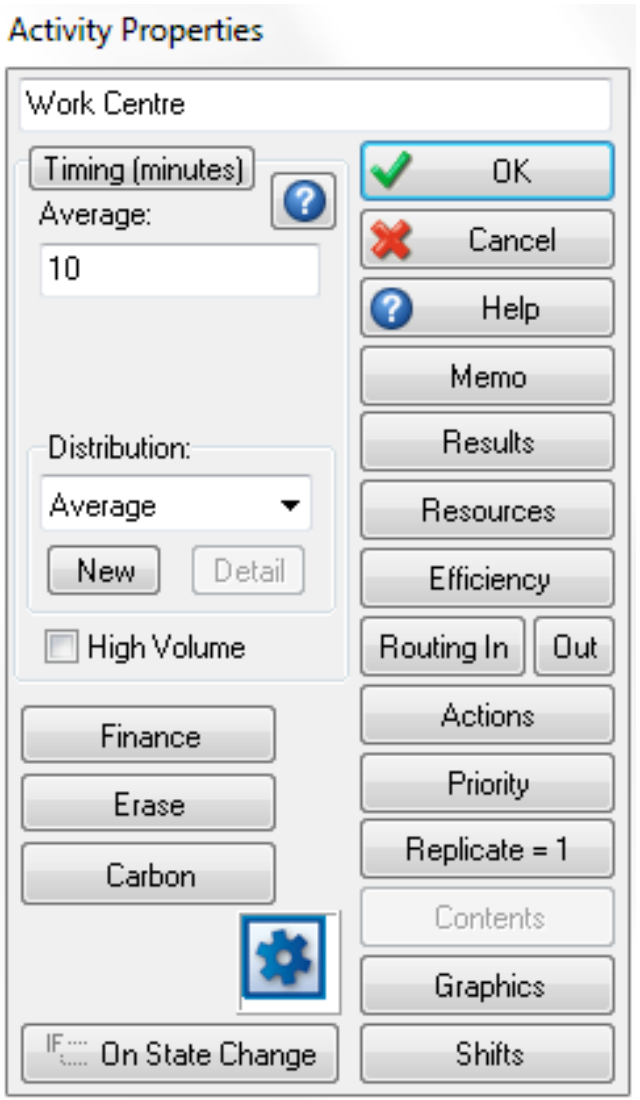
Animation of Queue

- Animates the queue so Work Item images are seen in a line as the simulation runs.
- Queues can be displayed at any angle with any distance desired between items in the queue.
- Change the direction a animated queue by dragging the handles that are displayed on the screen.
- Queues will normally be invisible unless you select the Show Route Arrows button. When the Route Arrows are displayed the Queue icon is displayed and the handles can be dragged around to change the direction that work items will queue in.

https://www.simul8.com/support/help/doku.php?id=features:graphics_sb

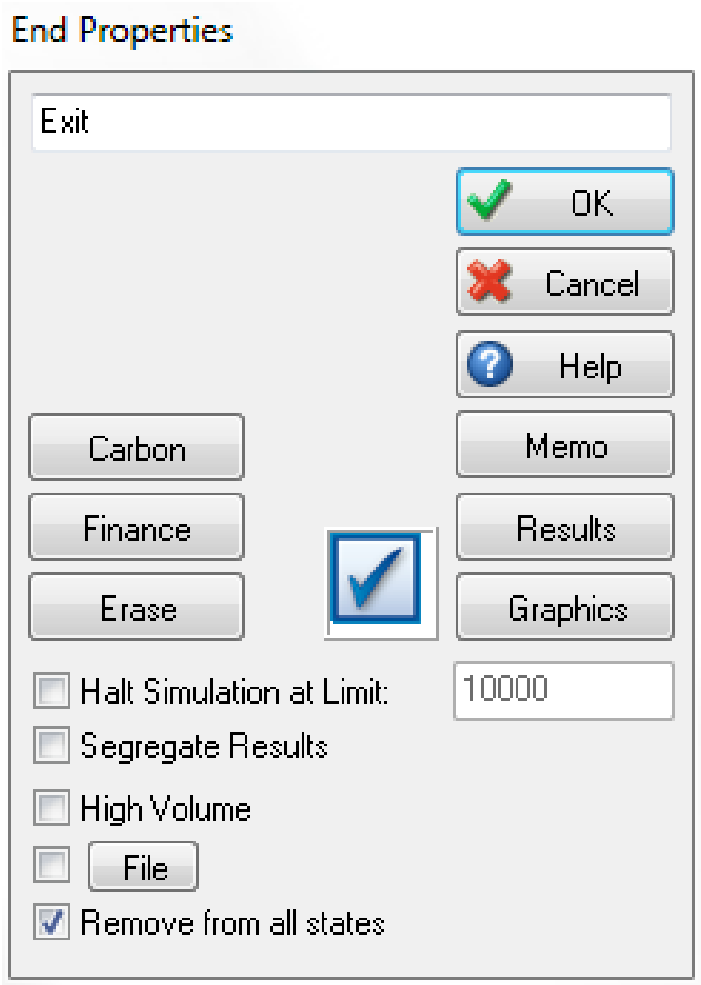
Activity

- An Activity is a place where work takes place on entities/ Work Items.
- Work done at Activity usually takes up time and sometimes requires the availability of resources. (set the timing panel)
- At an Activity (work centre), a work may be transformed in some way (perhaps by changing one or more of its labels).



End

- An End Point is a place where work that is complete or otherwise "finished" leaves the simulation.
  - At the point in time when each work item leaves, data is recorded about how long it has spent in the model (from the time when it entered through a “Start Point”).



Labels or Attributes

- When a new work item is generated into a Simulation, it is exactly the same as every other item. Labels let you tell items apart by attaching a Label, or property, to the item, and then giving the Label a numeric or Text value.

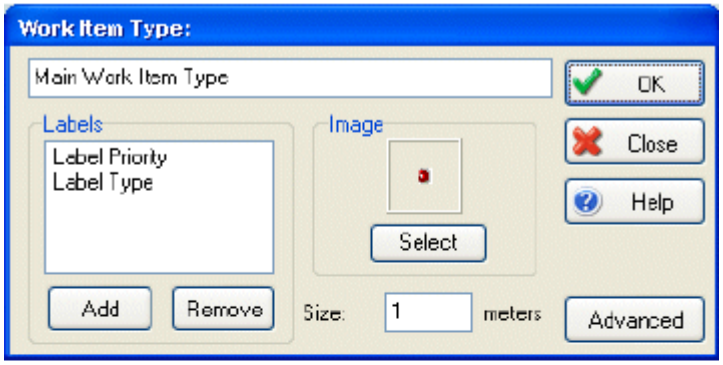
- An example of a Label would be Age. Everyone has an age, but ages differ form person to person. Part Number is another example of a Label. Each part has a part number that may be different.

Work Items

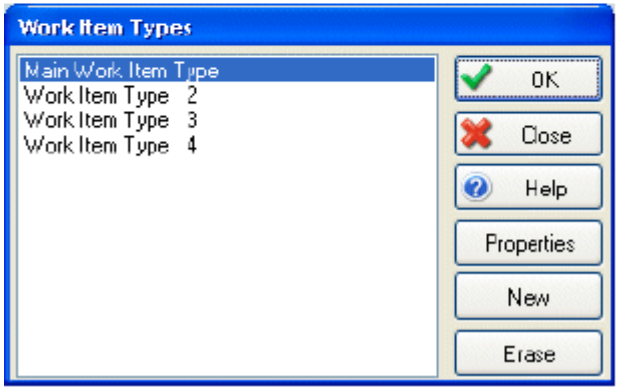
Work Items

A Work Item is the work which is done in the organisation being simulated. Examples are:

- Patients in a hospital
- Invoices in an Accounts department
- Products in a factory.



Individual work items flow through the objects in the simulation.



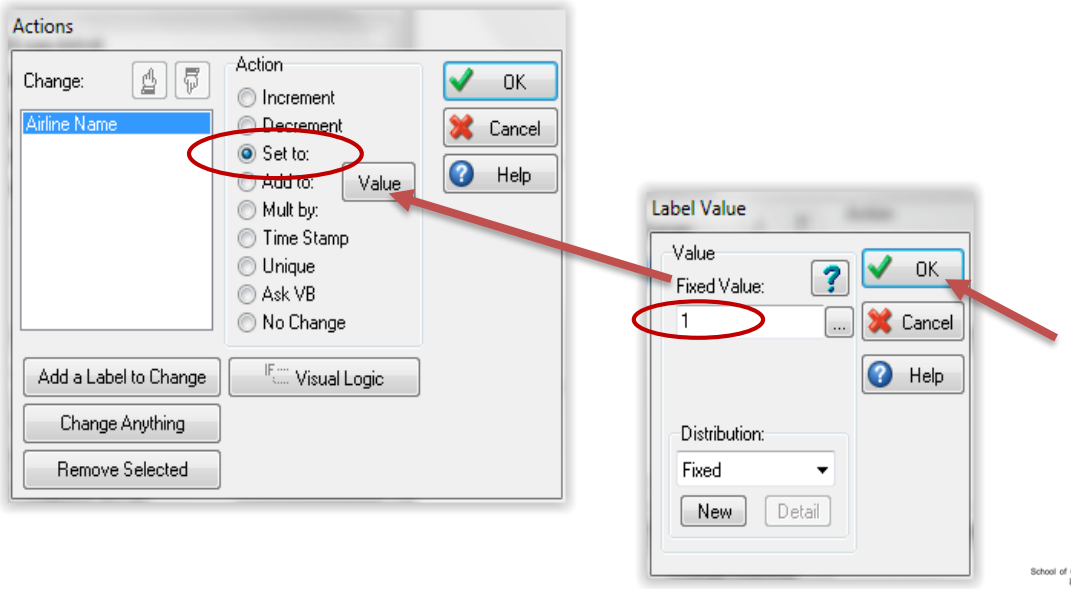
Labels can be attached to help differentiate between different work items or whole new work item types can be created.

Labels /Attributes

- Labels, or attributes as they're often called in other packages, can be attached to any work item in the simulation and used to store data about the work item.

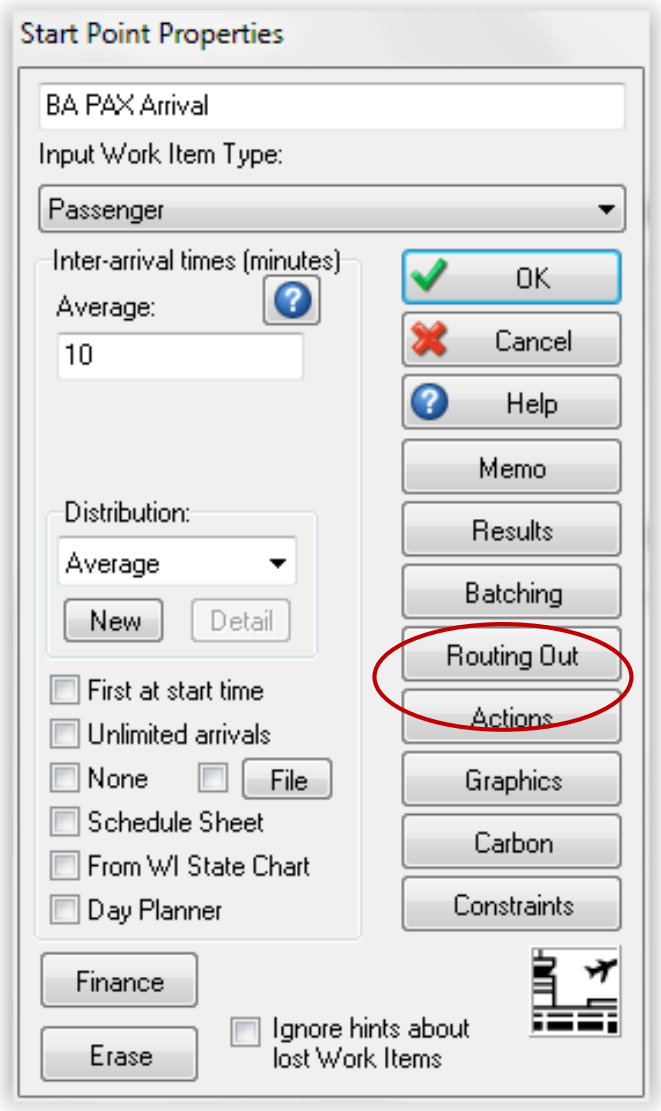
- For example you might attach a label “Airline Name“ and set this label to a code number as they enter your simulation of a terminal. Each code number would represent the Airline Name of a passenger.

- Labels can contain either text or numbers and can be changed throughout the simulation by the simulation objects that have Label Actions functionality.



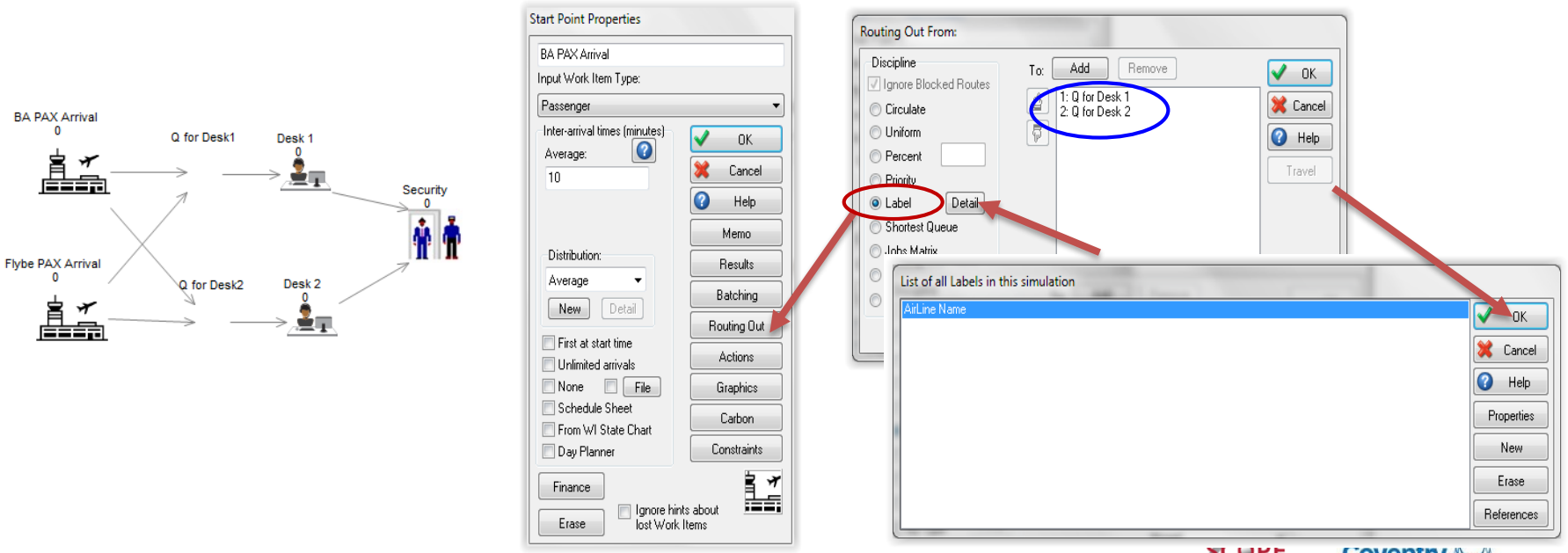
Start Point: Routing Out and Actions

- Routing Out Rules
- The following rules are available when controlling where work is sent to from a Start Point or an Activity:
  - Shortest Queue
  - Jobs Matrix
  - Circulate
  - Uniform
  - Percent
  - Priority
  - Label or Attribute
  - Passive
- Example: Lab8-2

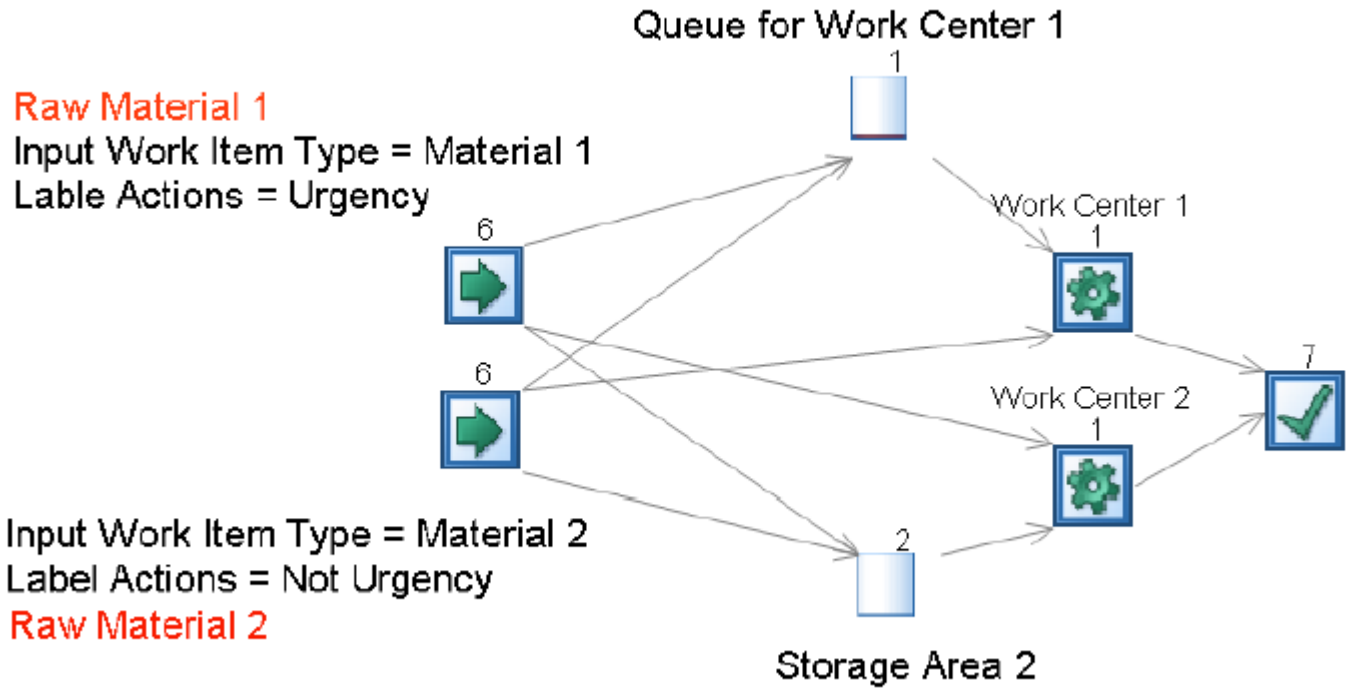


Label Routing Out: Example 1

- Label based **routing** will send work items along a route that corresponds to the value of a Label contained on the work item. For example if a work centre has 2 possible routes out a work item with label value 1 will be allocated to the first listed destination, and to the second destination if it has label value 2.



Label Routing Out: Example 2



- Two Products with total different attributes can be modelled as well.

Routing Out Options (I)

- Circulate
  - Work is circulated through the destinations in the list. That is the first work item will go to the first destination in the list, the second work item to the second,...., the Nth work item to the Nth destination on the list. The Nth+1 work item with go to the first destination etc.

- Uniform
  - Work is distributed randomly with equal chance of being sent to each of the routes.

- Shortest Queue
  - This method sends Work Items to the queue which is shortest.

Routing Out Options (II)

- Priority
  - The priority routing rule sends all work items to the first destination in the list, unless this is blocked, when the 2nd is tried… and so on.

- Percent
  - The destination work centre is decided randomly (like “Uniform” except that the exact percentage going to each destination can be specified. If the percentages do not add to 100 they will be pro-rata adjusted when you click OK to exit from the dialog box.

References

- Robinson, S. (2014) Simulation: The Practice of Model Development and Use. 2nd edn. Palgrave, Macmillan

- Shalliker, J, (2014) An Introduction to Simulation in the Service Industry using SIMUL8 2014 (release 21). SIMUL8