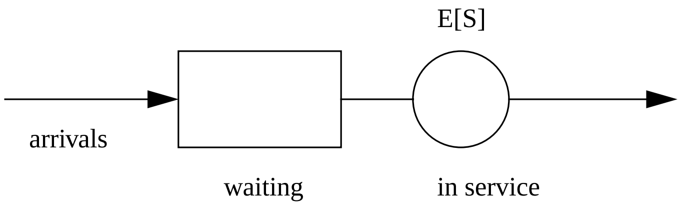
Simulation simple models and comparison with queueing theory.

Andrei Dorokhov

Three simple queueing models M/M/1, M/G/1 and series of M/M/1 have been simulated, and compared with calculations based on queueing theory for validation purposes.

1.0 M/M/1 model.

1.1 Simplest M/M/1 model description and formula.



This model consists of queueing station where jobs arrive with a negative exponential interarrival time distribution with rate λ. Furthermore, the job time service requirements are also negative exponentially distributed with mean $E[S] = \frac{1}{\mu}$

Simplest queueing model M/M/1 theory gives the following formula for mean number of jobs in the system and mean response time, provided that the system is in stable state, i.e. $\rho = \frac{\lambda}{\mu} < 1$

(EQ 1)

$$E[N] = \frac{\rho}{1-\rho}$$

(EQ 2)

$$E[R] = \frac{E[S]}{(1-\rho)}$$

where E[N] is the mean number of jobs in the system, E[R] - mean response time of the system, E[S] - mean serve time of the system, utilisation $\rho = \frac{\lambda}{\mu}$, and λ is mean job arrival rate, μ is mean job service rate, $E[S] = \frac{1}{\mu}$ is mean service time

For the case shown in the figure 1, $\lambda = 500$, $\mu = 1000$, so $\rho = 0.5$, and mean number of jobs in the system is 1.0, which is equal to the value obtained from the simulation.

1.2 Simulation model description.

In order to simulate simple queueing model M/M/1, Monarc simulation tool have been used. There is one regional center "cern", used and one type of Activity "Job.Analysis". Analyze factors set to (0.0,0.0,0.0), this mean that job read only "TAG" data(one event) from one database server and does not read others types of data and afterwards process this "TAG" data The bottleneck bandwidth is database read speed, which is set to 1MB/s. Mean size of "TAG" data is 1KB. Page size set to 1Byte. The other parameters - link speed for node and database and CPU power set to large values (but not infeasible!) in order to neglect these values in calculation using queueing theory. The size of "TAG" data is random value with mean 1KB, and negative exponential distribution (Marcovian process). So, database serves jobs with mean rate 1MB/s/1KB=1000 jobs/sec. Jobs arrive with a negative exponential inter-arrival time distribution with mean rate 500. Simulation performed for different job arrival rates. Mean number of jobs in the system and mean response time have been extracted from the simulation.

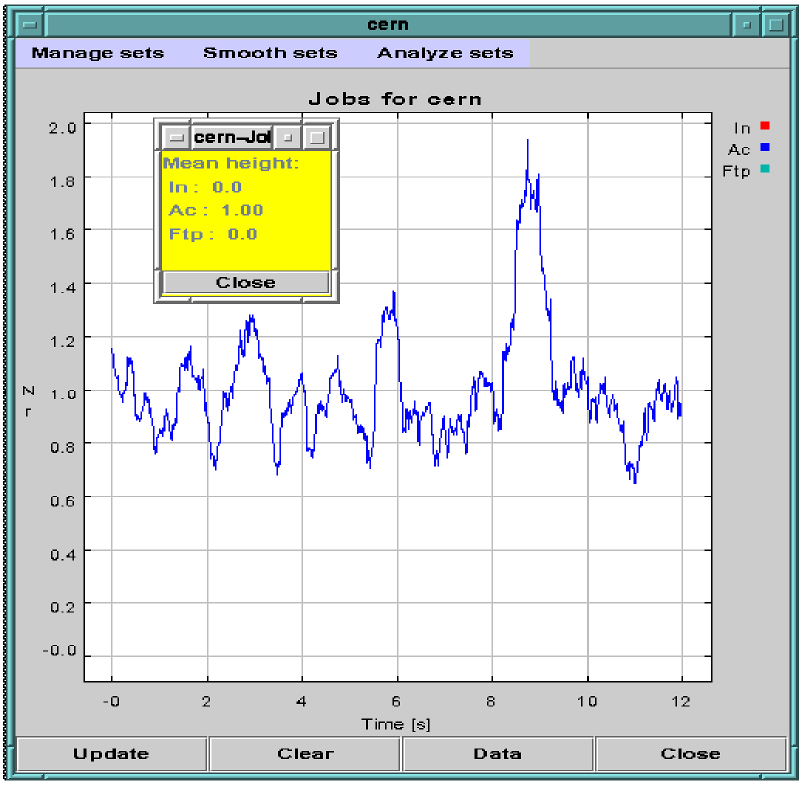


FIGURE 1.

Typical behavior for the number of jobs in the system is shown in the figure 1. In this case job arrival rate is taken 500 jobs/sec. One can see, that mean height (mean number of jobs in the system) is 1.0.

1.3 Comparison calculated M/M/1 model with simulation.

In the table 1 some values shown for different arrival rates Also on the figures 2 and 3 these dependencies are shown. E[R] is mean response time.E[N] - mean number of jobs in the system (including waiting and service)

TABLE 1.

Arrival rate	utilisation	E[N], from simulation	E[N] calculated	E[R],from simulation seconds	E[R] calculated seconds
1.0	0.001	0.001018	0.001001	0.001007	0.001001
10.0	0.01	0.001018	0.001010	0.010171	0.010101
50.0	0.05	0.001034	0.001053	0.052077	0.052632
100.0	0.1	0.001137	0.001111	0.112971	0.111111
200.0	0.2	0.001232	0.00125	0.246945	0.25
300.0	0.3	0.001538	0.001429	0.461039	0.428571
500.0	0.5	0.00199	0.0020	1.00087	1.0
700.0	0.7	0.003580	0.003333	2.497969	2.333333

FIGURE 2. Mean number of jobs for M/M/1 model.

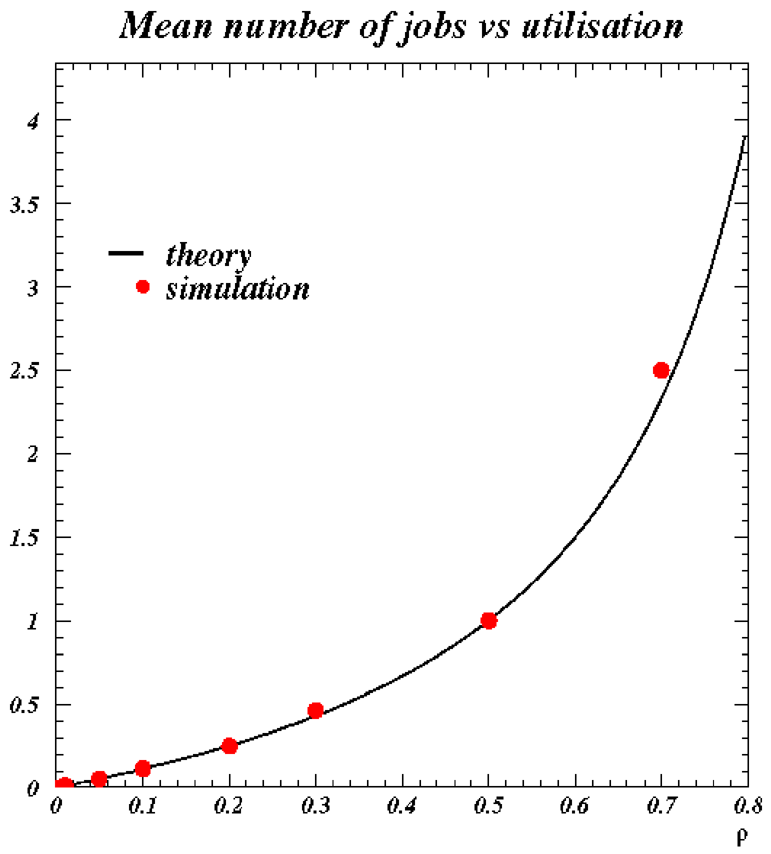
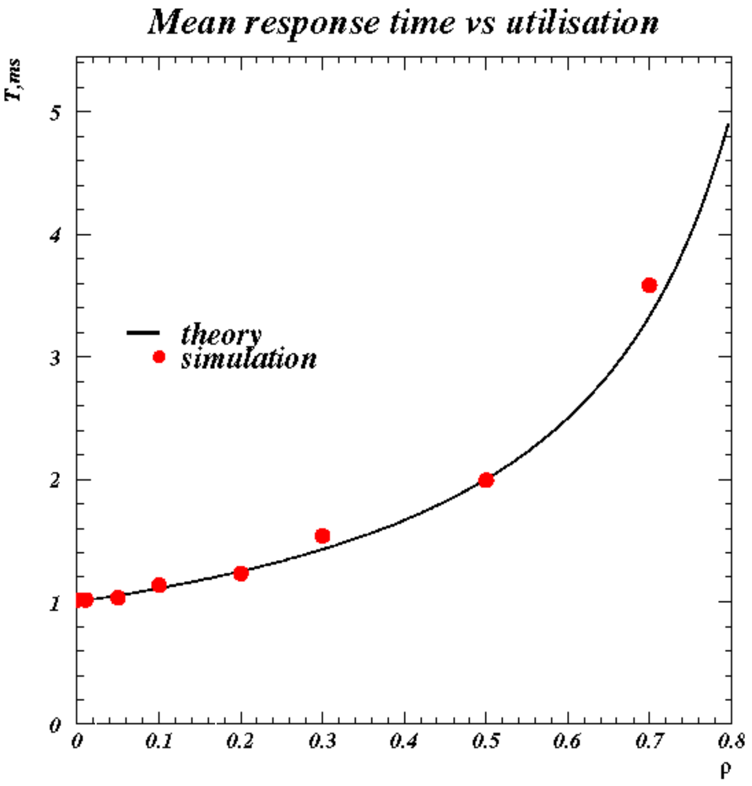
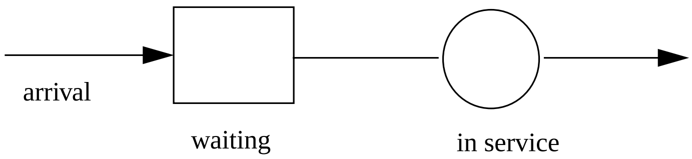


FIGURE 3. Mean response time for M/M/1 model



2.0 M/G/1 model

2.1 Formula and description for M/G/1 model.



This model consists of queueing station where jobs arrive with a negative exponential interarrival time distribution with rate λ. The job service time has general distribution with mean $E[S] = \frac{1}{\mu}$

Mean number of jobs in the M/G/1 system is given by Pollaczek-Khintchine(PK) formula (EQ 3):

$$E[N] = \lambda E[S] + \frac{\lambda^2 E[S^2]}{2(1-\rho)} \quad \text{(EQ 3)}$$

$$E[R] = E[S] + \frac{\lambda E[S^2]}{2(1-\rho)} \quad \text{(EQ 4)}$$

In the simulation one can use hyperexponential distribution of service time as general type of distribution.

Hyperexponential distribution is defined as: $f(x) = \sum_{i=1}^{r} \alpha_i \mu_i e^{-\mu_i x}$, mean service time is $E[S] = \sum_{i=1}^{r} \frac{\alpha_i}{\mu_i}$,

second moment is $E[S^2] = 2\sum_{i=1}^{r} \frac{\alpha_i}{\mu_i^2}$, utilisation is $\rho = \lambda \sum_{i=1}^{r} \frac{\alpha_i}{\mu_i}$

2.2 Simulation model description

Model used to simulate general distribution of service time is very close to the model used to simulate M/M/1 queueing system, but 50% job requires for two events, and 50% one event data type “TAG“, in other words, when job is submitted, with probability 50% it will demand for one event, and 50% for two events. So when job requires two event it will take twice more time to be served. The distribution of job arrival time is again negative exponential. Now we have hyperexponential distribution for serving time - general distribution (G), and Marcovian (M) for arrival time.

2.3 Comparison simulation and calculation of model M/G/1.

Simulation and calculation have been performed for different arrival rates. In the table 2 some results are summarized. See also figures 3 and 4, where mean values are plotted.

TABLE 2.

arrival rate	utilisation	E[N], simulated	E[N], calculated	E[R], simulated	E[R], calculated
1.0	0.0015	0.001503	0.001503	0.001526	0.001503
10.0	0.015	0.001511	0.001525	0.015002	0.015254
50.0	0.075	0.001720	0.001635	0.086201	0.081757
100.0	0.15	0.001862	0.0017941	0.189175	0.179412
200.0	0.3	0.002069	0.002214	0.417950	0.442857

TABLE 2.

arrival rate	utilisation	E[N], simulated	E[N], calculated	E[R], simulated	E[R], calculated
300.0	0.45	0.002818	0.002864	0.838326	0.859091
400.0	0.6	0.003612	0.0040	1.403495	1.6

FIGURE 4. Mean number of jobs for M/G/1 model

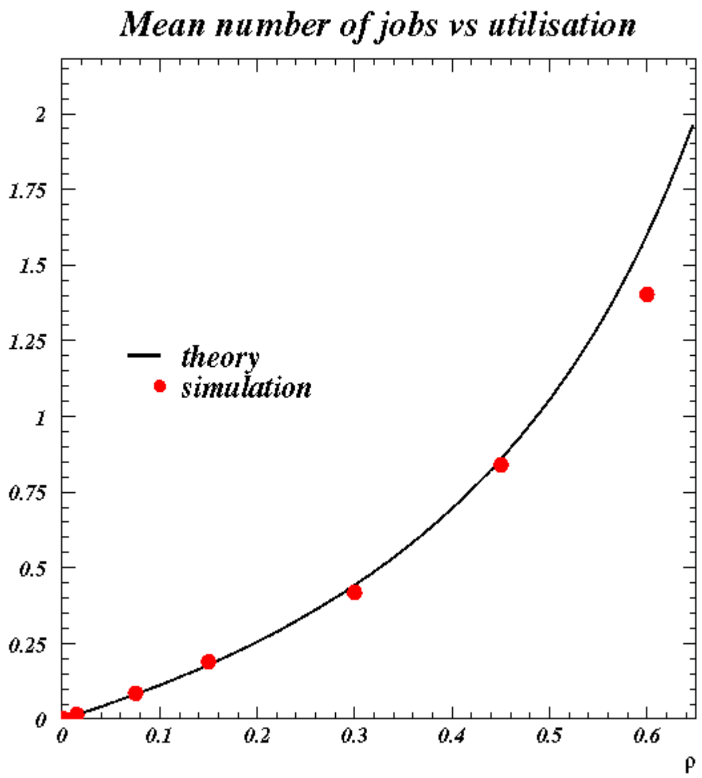
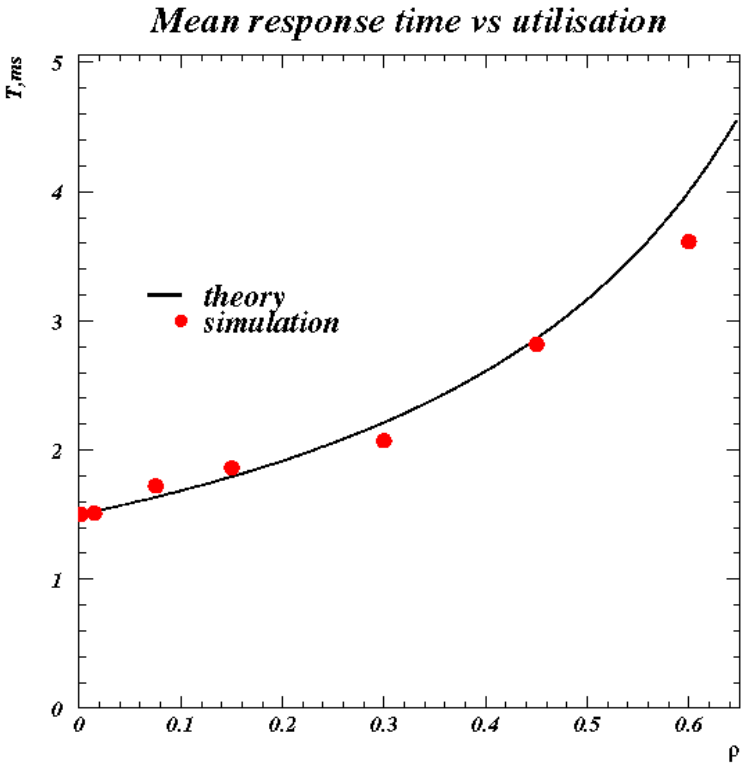


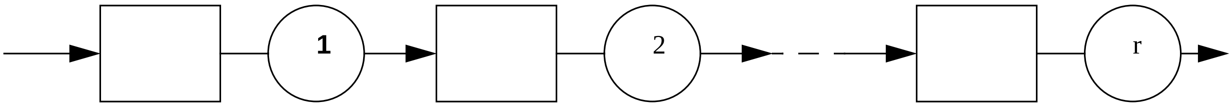
FIGURE 5. Mean response time for M/G/1 model



3.0 M/M/1 network queue model

3.1 Formula and description for M/M/1 network queue model.

This type of queueing model consists of series of M/M/1 queues.



Mean total number of jobs in the system is:

$$E[N] = \sum_{i=1}^{r} E[N_i] = \sum_{i=1}^{r} \frac{\rho_i}{1-\rho_i} \qquad \text{(EQ 5)}$$

Mean total response time of total network is:

$$E[R] = \sum_{i=1}^{r} E[R_i] = \sum_{i=1}^{r} \frac{E[Si]}{(1-\rho i)} \qquad \text{(EQ 6)}$$

where utilisation for each stage is $\rho_i = \frac{\lambda}{\mu_i}$.

3.2 SImulation model description.

This network model can be easily derived from simple M/M/1 model as described above, by setting Analyze Factors to (1,1,1). It means, that after reading and processing "TAG" data, the following data will be read and processed sequentially 100% of "AOD", 100% of "ESD" and 100% "RAW" data.(percentages from requested "TAG" data).As we have different size for different types of data, we obtain different service time (or service rate). But according to Burk's theorem the departure process from a stable single server M/M/1 queue with arrival and service rates and respectively, is a poisson process with arrival rate. So we can apply the same formula as for M/M/1 case for each stage of process and sum mean number of job and mean response time in each stage.

3.3 Comparison M/M/1 network queue mode simulated and calculated.

In the table 3 one can find the main results from simulation and calculation of M/M/1 network queue model. See also figures 5 and 6.

TABLE 3.

arrival rate	utilisation	E[N], simulated	E[N], calculated	E[R], simulated	E[R], calculated
0.0010	0.0010	1.150163	1.112011	0.001185	0.001112
0.01	0.01	1.104074	1.121202	0.011395	0.011212
0.05	0.05	1.149576	1.164139	0.056958	0.058207
0.1	0.1	1.210121	1.223131	0.120573	0.122313
0.2	0.2	1.383068	1.363061	0.280643	0.272612
0.3	0.3	1.525811	1.542695	0.443952	0.462808
0.5	0.5	2.176270	2.116314	1.121783	1.058157
0.7	0.7	3.057828	3.4519315	2.133265	2.416352

FIGURE 7. Mean number of jobs for network M/M/1 queue

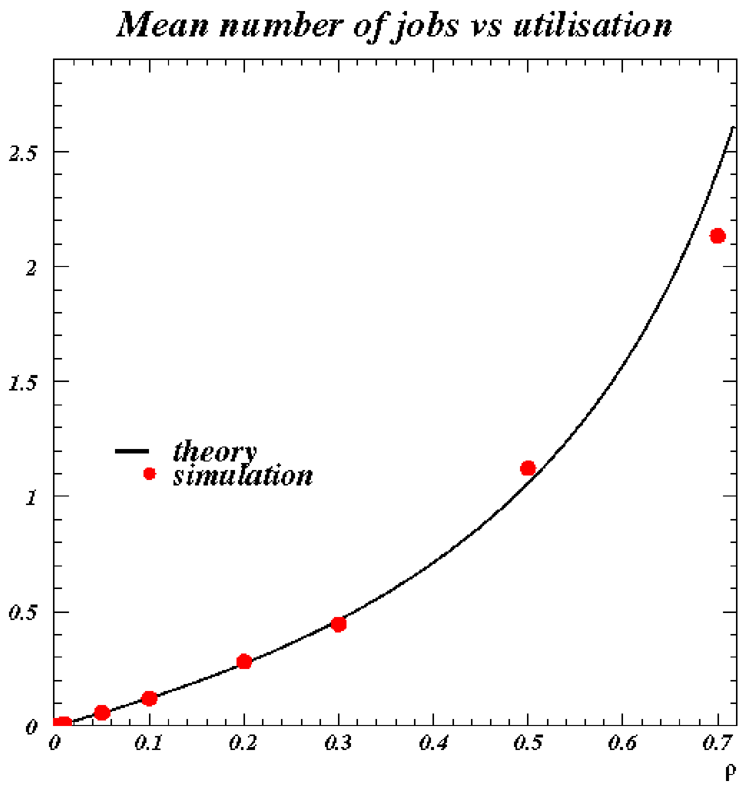
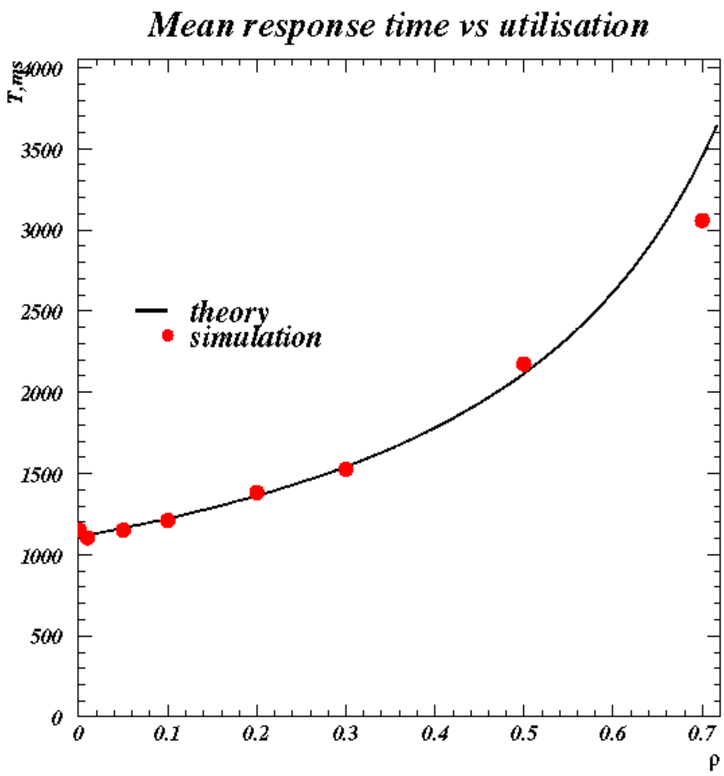


FIGURE 8. Mean response time for network M/M/1 queue



4.0 Conclusions

Three simple model have been simulated and compared with prediction of queueing theory. Good agreement between queueing theory and simulation results have been shown.