RETRIAL QUEUES FOR PERFORMANCE MODELLING AND EVALUATION OF HETEROGENEOUS NETWORKS

J. Roszik, J. Sztrik

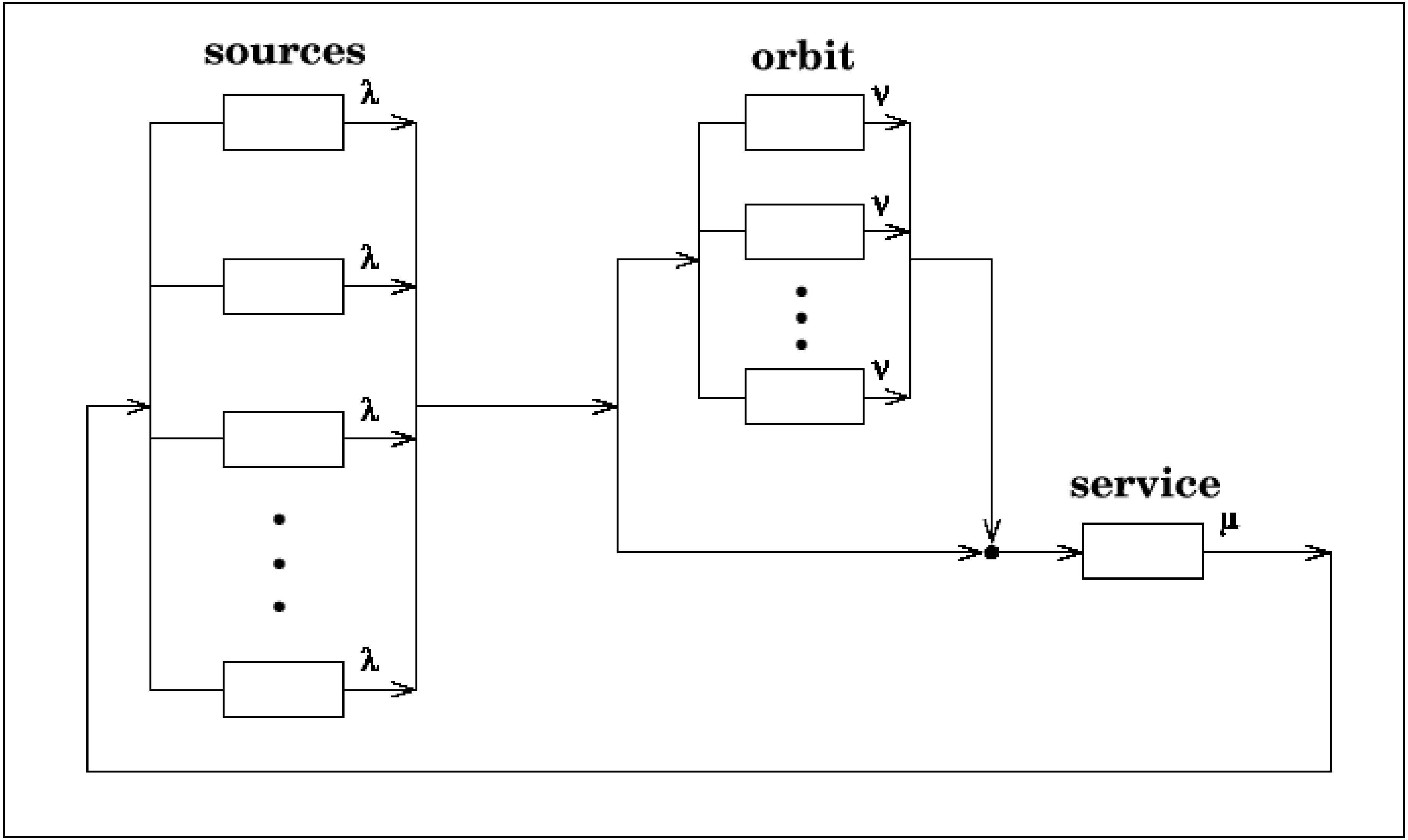
Institute of Informatics, University of Debrecen
Debrecen, Hungary

e-mail: jsztrik@inf.unideb.hu
www: http://irh.inf.unideb.hu/user/jsztrik/

OUTLOOK

- **The queueing model**
- **Applications**
- **Mathematical model**
- **Evaluation Tool MOSEL**
- **Case studies**
- **References**

The queueing model



Applications

- magnetic disk memory systems
- local area networks with CSMA/CD protocols
- collision avoidance local area network modeling

Mathematical model

$$P(0;0) = \lim_{t \to \infty} P(C(t) = 0, N(t) = 0)$$

$$P(j;0) = \lim_{t \to \infty} P(\alpha_1 = j, N(t) = 0), \quad j = 1, ..., K$$

$$P(0; i_1, ..., i_k) = \lim_{t \to \infty} P(C(t) = 0, \beta_1 = i_1, ..., \beta_k = i_k), \quad k = 1, ..., K-1$$

$$P(j; i_1, ..., i_k) = \lim_{t \to \infty} P(\alpha_1 = j, \beta_1 = i_1, ..., \beta_k = i_k), \quad k = 1, ..., K-1.$$

Once we have obtained these limiting probabilities the **main system performance measures** can be derived in the following way.

1. The server utilization with respect to source j

$$U_j = \text{P (the server is busy with source } j \text{)}$$

that is, we have to summarize all the probabilities where the first component is j. Formaly

$$U_j = \sum_{k=0}^{K-1} \sum_{i_1,\ldots,i_k \neq j} P(j; i_1, \ldots, i_k)$$

Hence the **server utilization**

$$U = E[C(t) = 1] = \sum_{j=1}^{K} U_j.$$

Let us denote by $P_W^{(i)}$ the steady state probability that request i is waiting (staying in the orbit). It is easy to see that

$$P_W^{(i)} = \sum_{j=0, j \neq i}^{K} \sum_{k=1}^{K-1} \sum_{i \epsilon (i_1, ..., i_k)} P(j; i_1, ..., i_k).$$

Similarly, it can easily be seen, that the steady state probability $P^{(i)}$ that request i is in the service facility (it is under service or waiting in the orbit) is given by

$$P^{(i)} = P_W^{(i)} + U_i.$$

2. Mean response time of source i

Let us denote by $E[T_i]$ the mean response time of customer i, and by γ_i the **throughput** of request i, that is, the mean number of times that request i is served per unit time. These are related by

$$\gamma_i = \frac{1}{E[T_i] + 1/\lambda_i} = \lambda_i(1 - P^{(i)}) = \mu_i U_i, \qquad i = 1, ..., K. \qquad (1)$$

For $P^{(i)}$ we have

$$P^{(i)} = \frac{E[T_i]}{E[T_i] + 1/\lambda_i} = \gamma_i E[T_i] = 1 - \frac{\gamma_i}{\lambda_i} \qquad i = 1, ..., K. \qquad (2)$$

which represents **Little's theorem** for request i in the service facility. It is easy to see that as a consequence of (**??**) we have

$$P^{(i)} = 1 - \frac{\mu_i}{\lambda_i} U_i$$

and

$$P_W^{(i)} = P^{(i)} - U_i = 1 - \frac{\mu_i + \lambda_i}{\lambda_i} U_i.$$

Alternatively, by the help of (**??**) we can express the mean response time $E[T_i]$ for request i in terms of U_i as

$$E[T_i] = \frac{P^{(i)}}{\lambda_i(1 - P^{(i)})} = \frac{1 - \frac{\mu_i}{\lambda_i} U_i}{\mu_i U_i} = \frac{\lambda_i - \mu_i U_i}{\lambda_i \mu_i U_i}. \quad (3)$$

3. Mean waiting time of source i

The mean waiting time of request i is given by

$$E[W_i] = E[T_i] - \frac{1}{\mu_i} = \frac{1}{\gamma_i} - \frac{1}{\lambda_i} - \frac{1}{\mu_i} = \frac{\lambda_i - (\mu_i + \lambda_i) U_i}{\lambda_i \mu_i U_i}. \quad (4)$$

At the same time we have another **Little's theorem** for request i waiting for service. Namely

$$P_W^{(i)} = \frac{E[W_i]}{E[T_i] + 1/\lambda_i} = \gamma_i E[W_i] \qquad i = 1, ..., K.$$

4. Mean number of calls staying in the orbit or in service

$$M = E[C(t) + N(t)] = \sum_{i=1}^{K} P^{(i)} = \sum_{i=1}^{K} (1 - \frac{\mu_i}{\lambda_i} U_i) = K - \sum_{i=1}^{K} \frac{\mu_i}{\lambda_i} U_i.$$

5. Mean number of sources of repeated calls

$$N = E[N(t)] = \sum_{i=1}^{K} P_W^{(i)} = \sum_{i=1}^{K} (1 - \frac{\mu_i + \lambda_i}{\lambda_i} U_i) = K - \sum_{i=1}^{K} \frac{\mu_i + \lambda_i}{\lambda_i} U_i.$$

6. Mean rate of generation of primary calls

$$\overline{\lambda} = \sum_{i=1}^{K} \gamma_i = \sum_{i=1}^{K} \lambda_i (1 - P^{(i)}) = \sum_{i=1}^{K} \mu_i U_i.$$

7. Blocking probability of primary call i

$$B_i = \frac{\lambda_i \sum_{j=1, j \neq i}^{K} \sum_{k=0}^{K-1} \sum_{i \neq i_1, \ldots, i_k} P(j; i_1, \ldots, i_k)}{\overline{\lambda}}.$$

Hence **blocking probability of primary calls**

$$B = \sum_{i=1}^{K} B_i$$

In particular, in the case of **homogeneous calls**

$$U_i = E[C(t)]/K, \quad i = 1, ..., K, \quad N = K - \frac{(\lambda + \mu)U}{\lambda},$$

$$\overline{\lambda} = \lambda E[K - C(t) - N(t)] = \mu U,$$

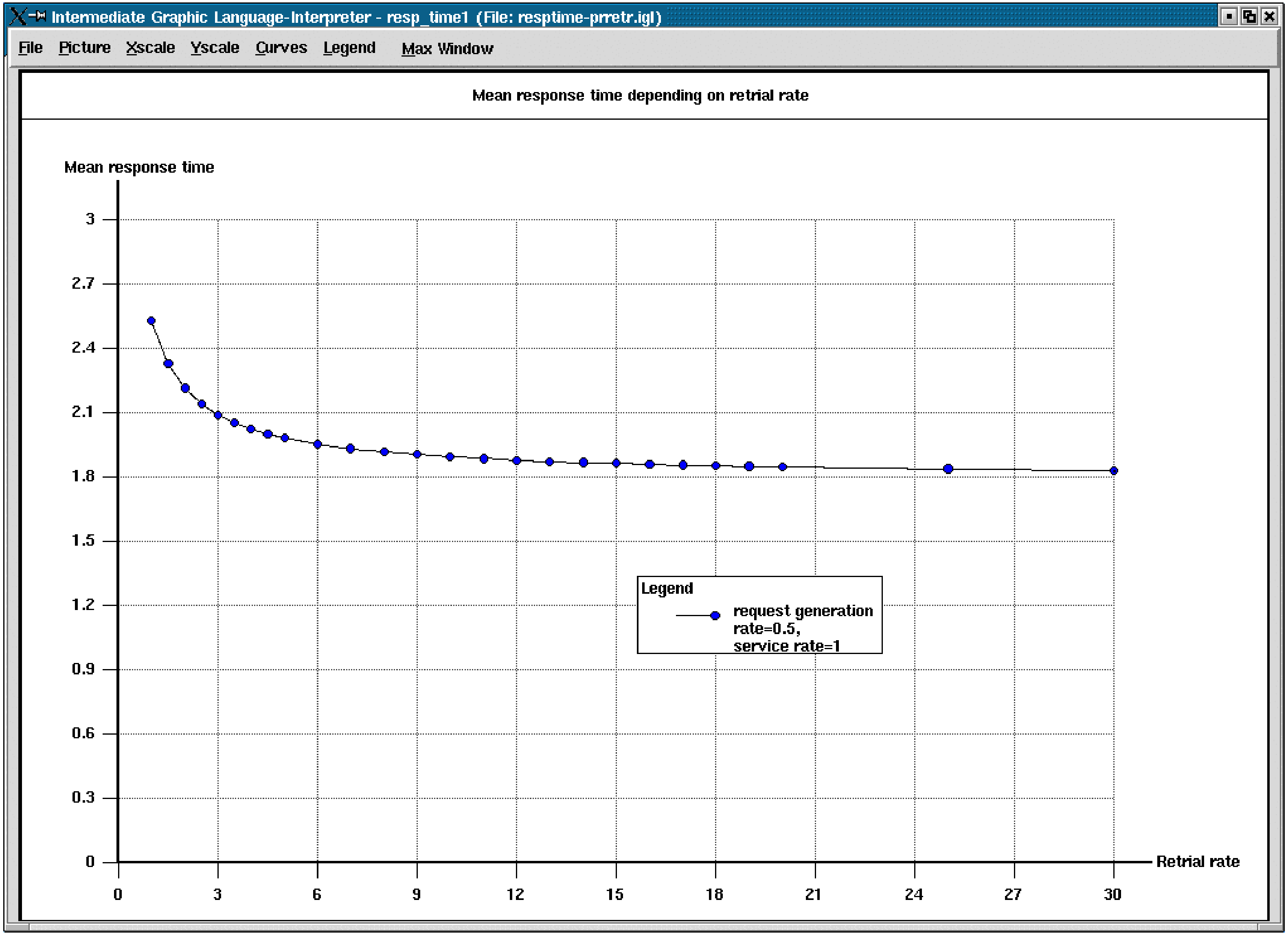
$$E[W] = \frac{N}{\overline{\lambda}} = K - \frac{1}{\mu U} - \frac{1}{\lambda} - \frac{1}{\mu},$$

$$B = \frac{\lambda E[K - C(t) - N(t); C(t) = 1]}{\lambda E[K - C(t) - N(t)]}.$$

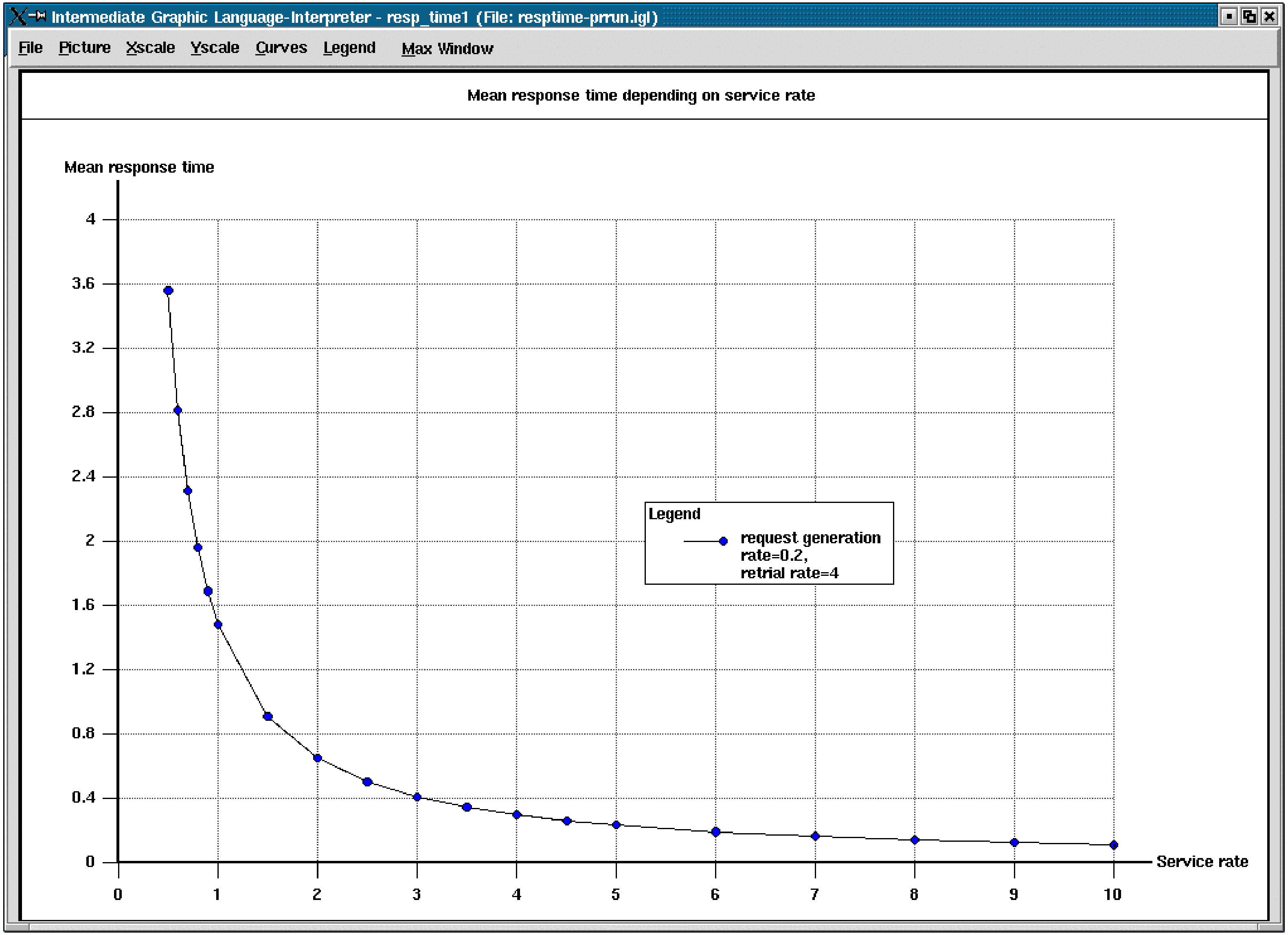
Evaluation Tool MOSEL

MOSEL (Modeling, Specification and Evaluation Language) developed at the University of Erlangen, Germany, is used to formulate and solved the problem.

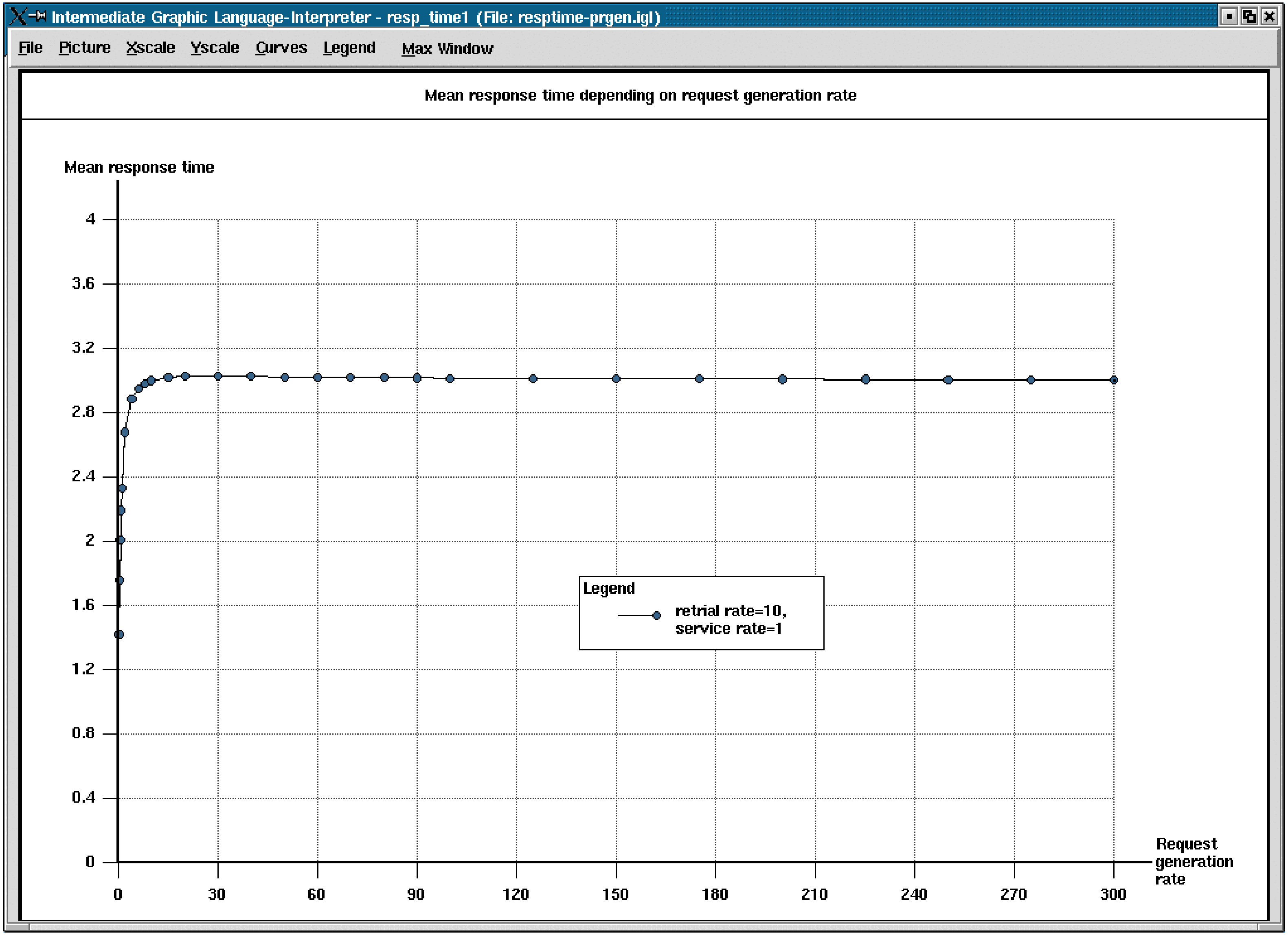
Case studies



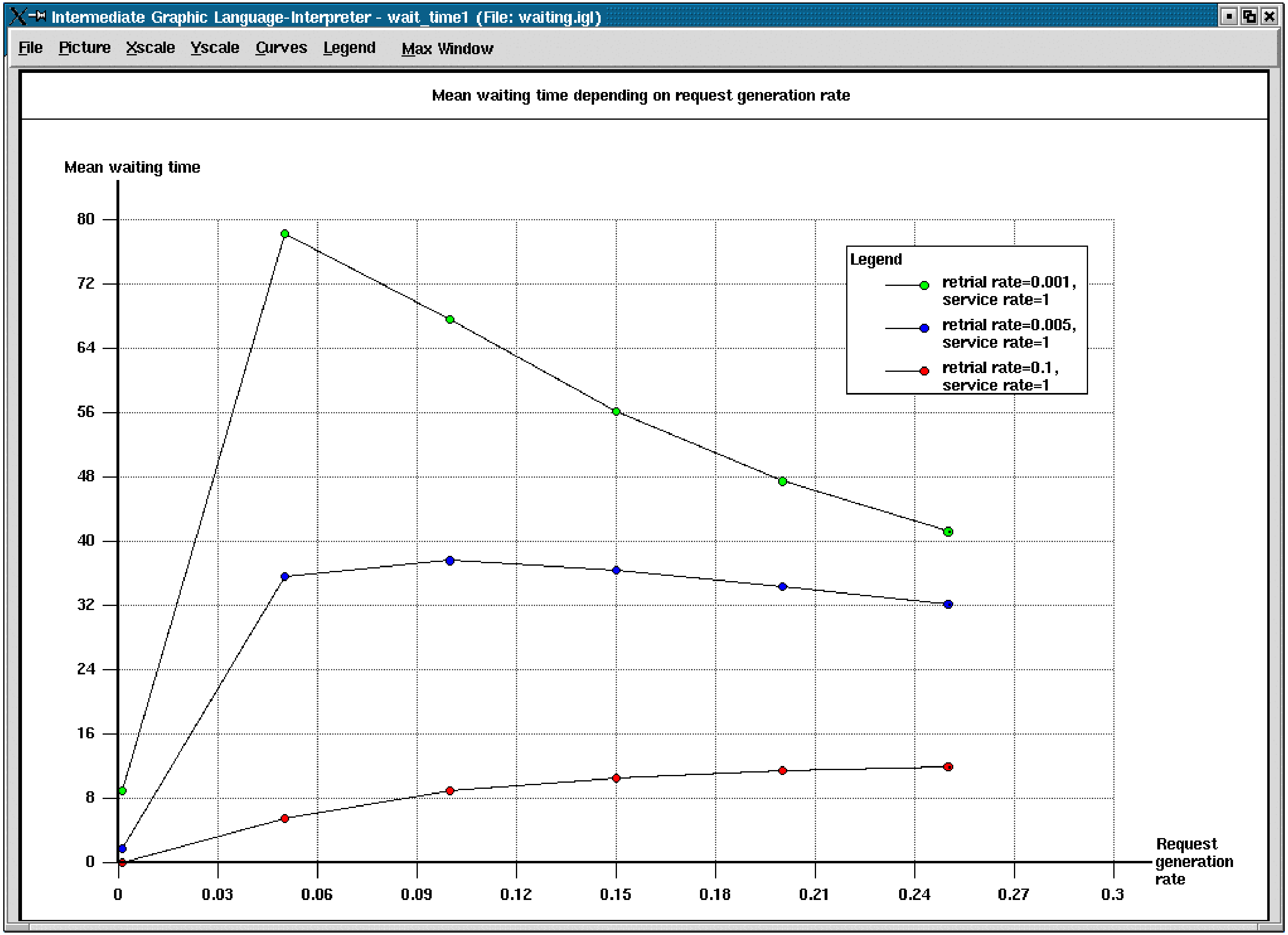
$E[T]$ versus retrial rate



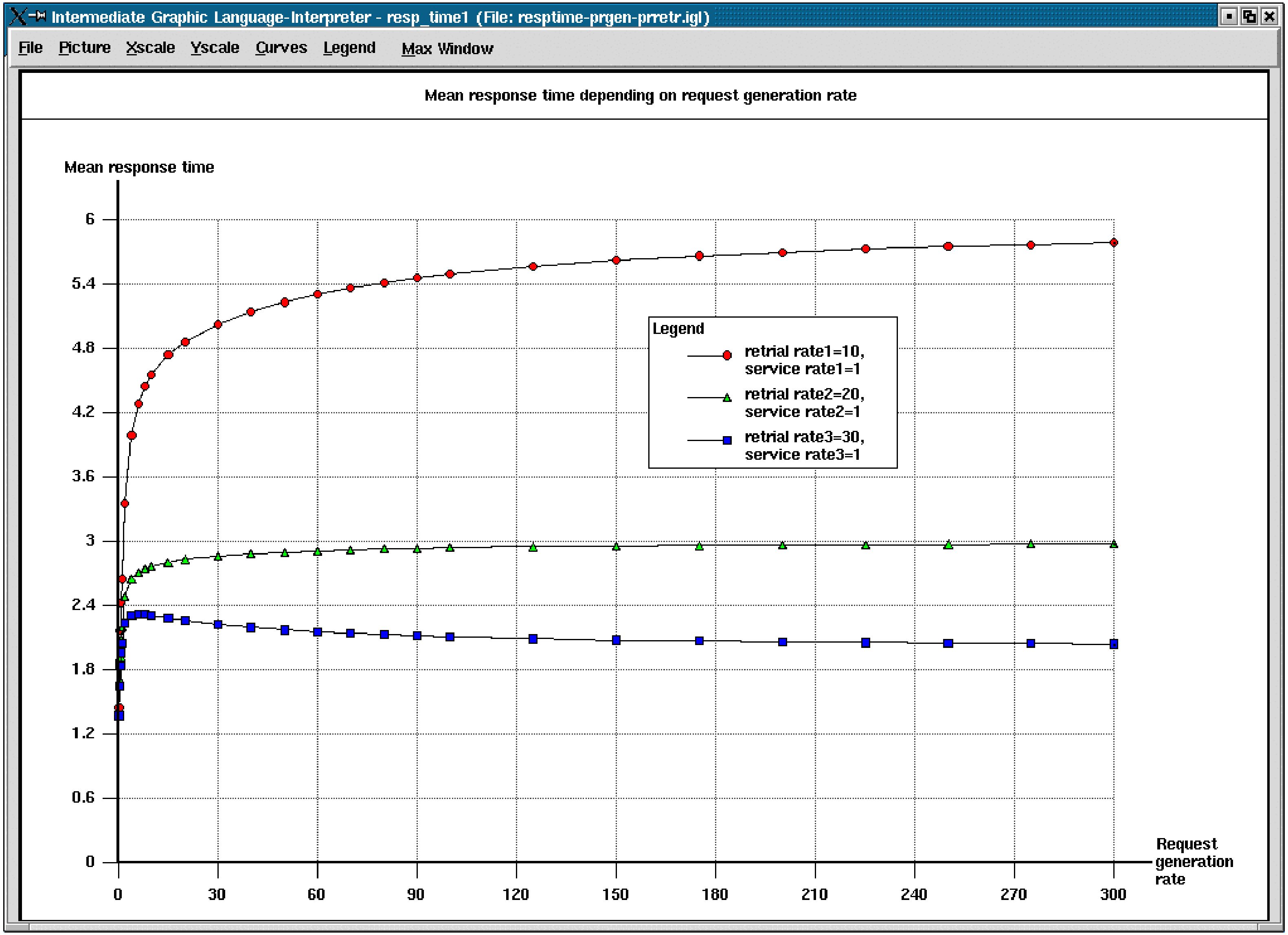
$E[T]$ versus service rate



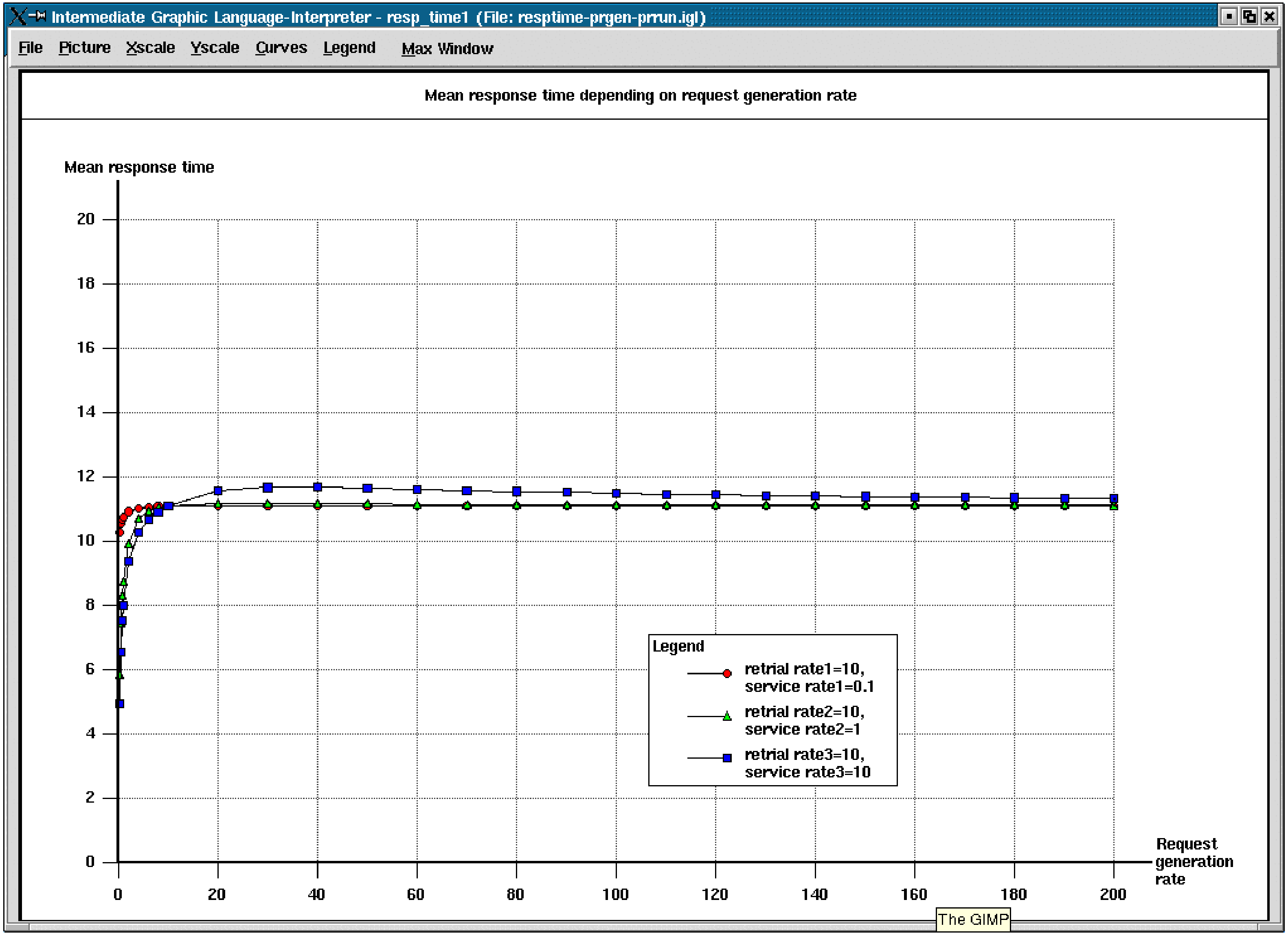
$E[T]$ versus primary request generation rate



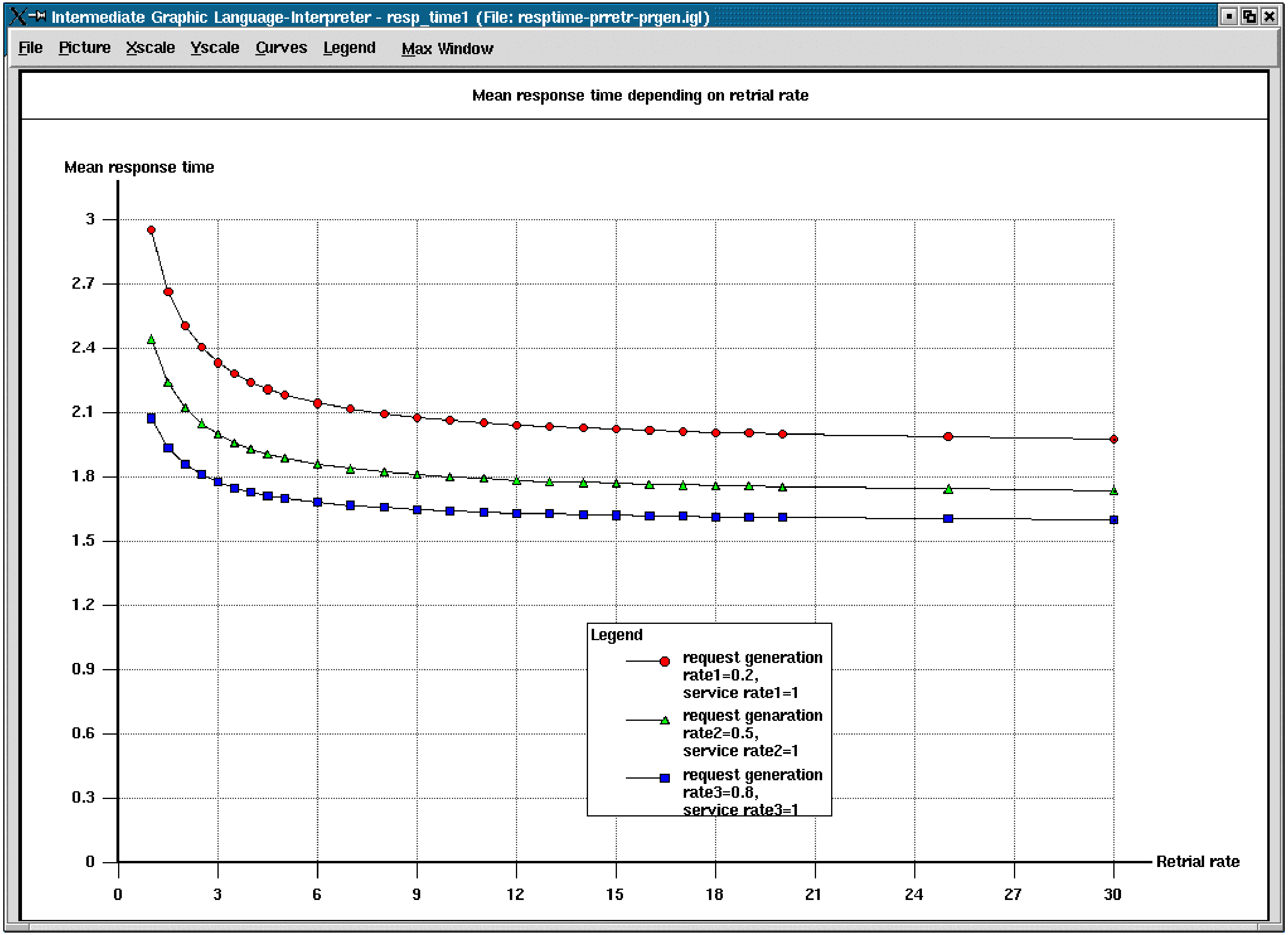
$E[T]$ versus primary request generation rate



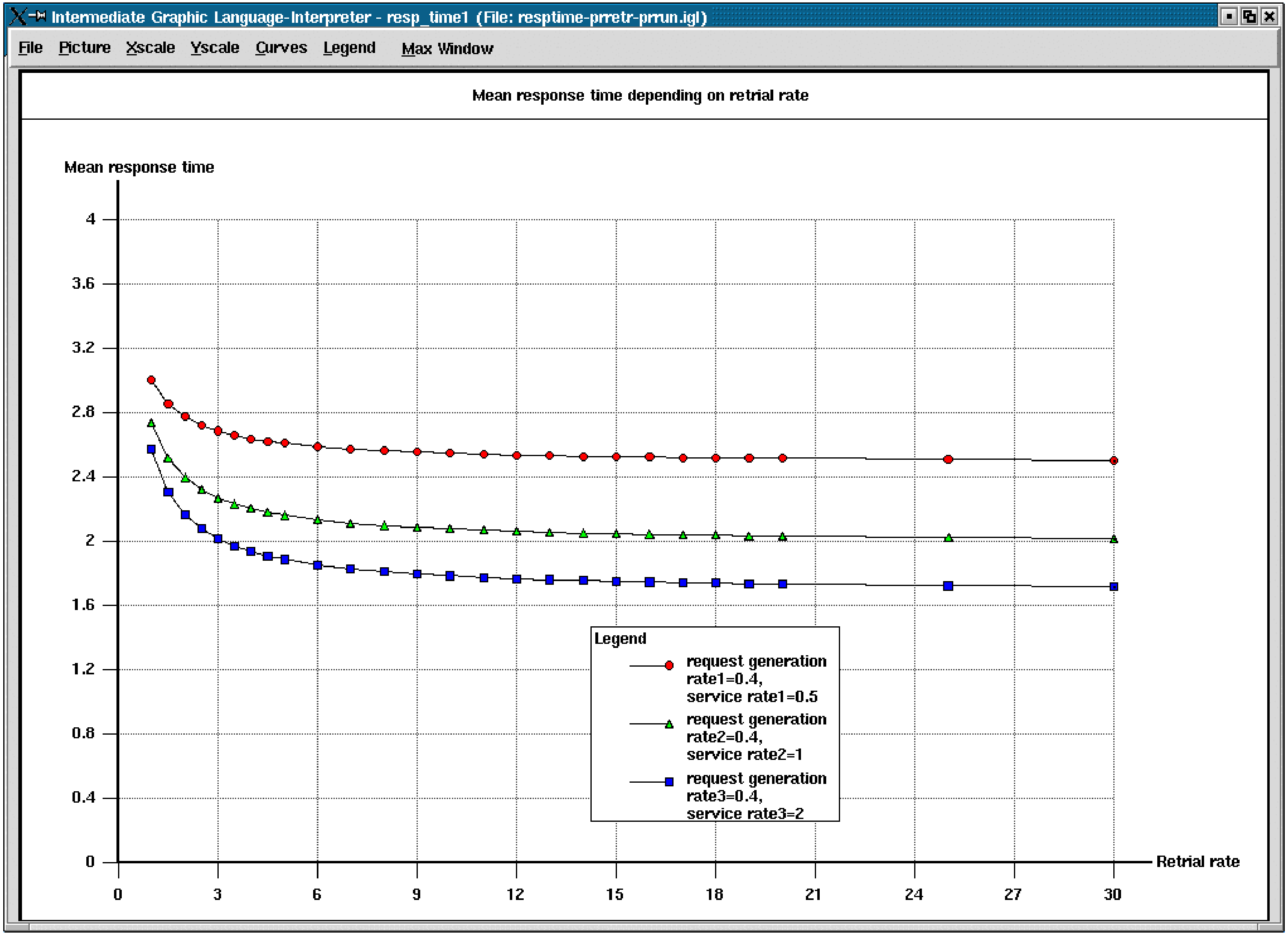
$E[T]$ versus primary request generation rate with homogeneous service and heterogeneous retrial



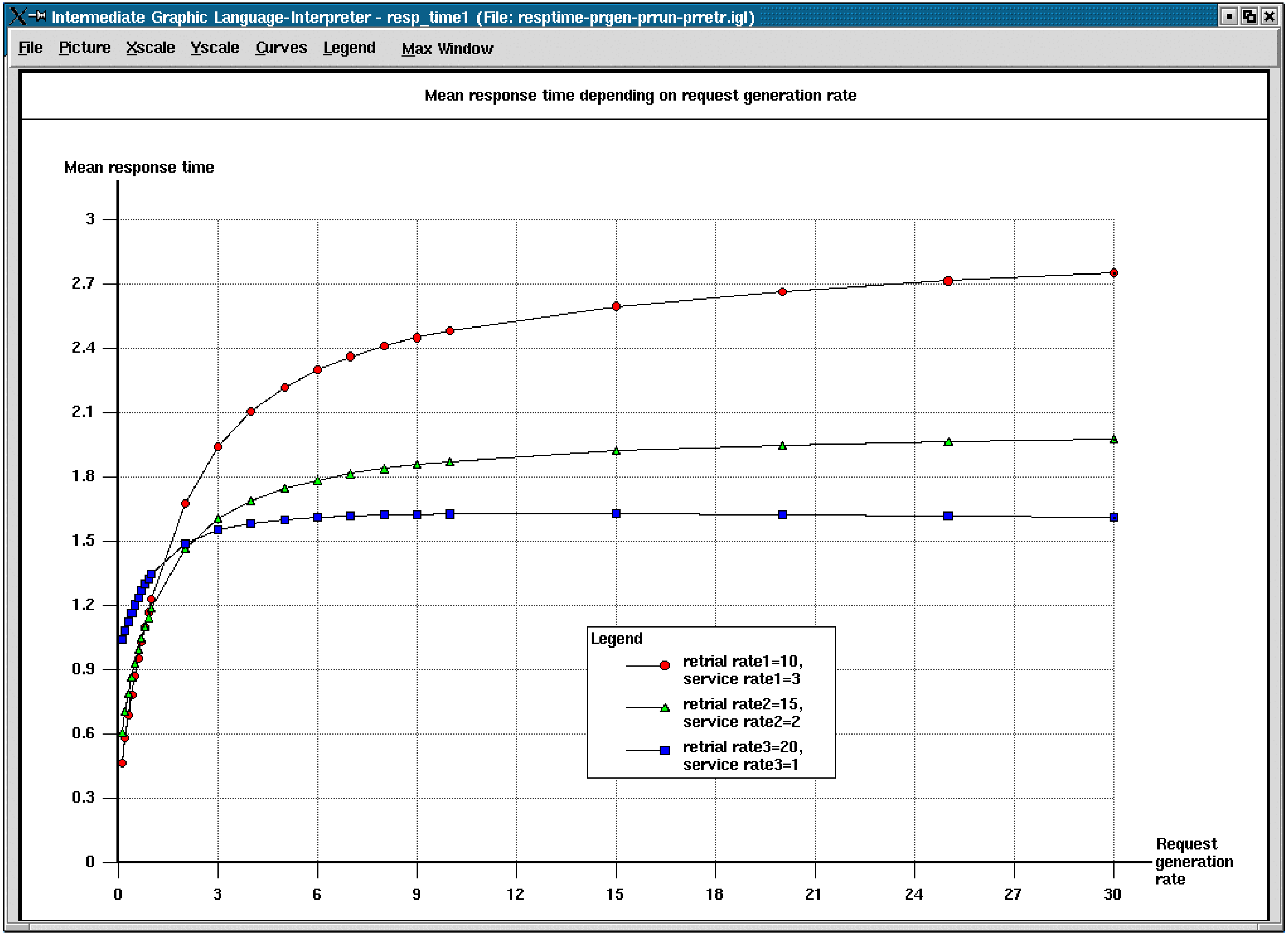
$E[T]$ versus primary request generation rate with homogeneous retrial and heterogeneous service



$E[T]$ versus retrial rate with homogeneous service and heterogeneous primary request generation



$E[T]$ versus retrial rate with homogeneous primary request generation and heterogeneous service



$E[T]$ versus primary request generation rate with heterogeneous service and heterogeneous retrial

References

[1] **Artalejo J.R.** Retrial queues with a finite number of sources, *J. Korean Math. Soc*. 35(1998) 503-525.

[2] **Begain K., Bolch G., Herold H.** *Practical Performance Modeling, Application of the MOSEL Language*, Kluwer Academic Publisher, Boston, 2001.

[3] **Falin G.I. and Templeton J.G.C.** *Retrial queues*, Chapman and Hall, London, 1997.

[4] **Falin G.I. and Artalejo J.R.** A finite source retrial queue, *European Journal of Operational Research* 108(1998) 409-424.