HETEROGENEOUS FINITE-SOURCE RETRIAL QUEUES WITH SERVER SUBJECT TO BREAKDOWNS AND REPAIRS

B. Almási, J. Roszik, J. Sztrik

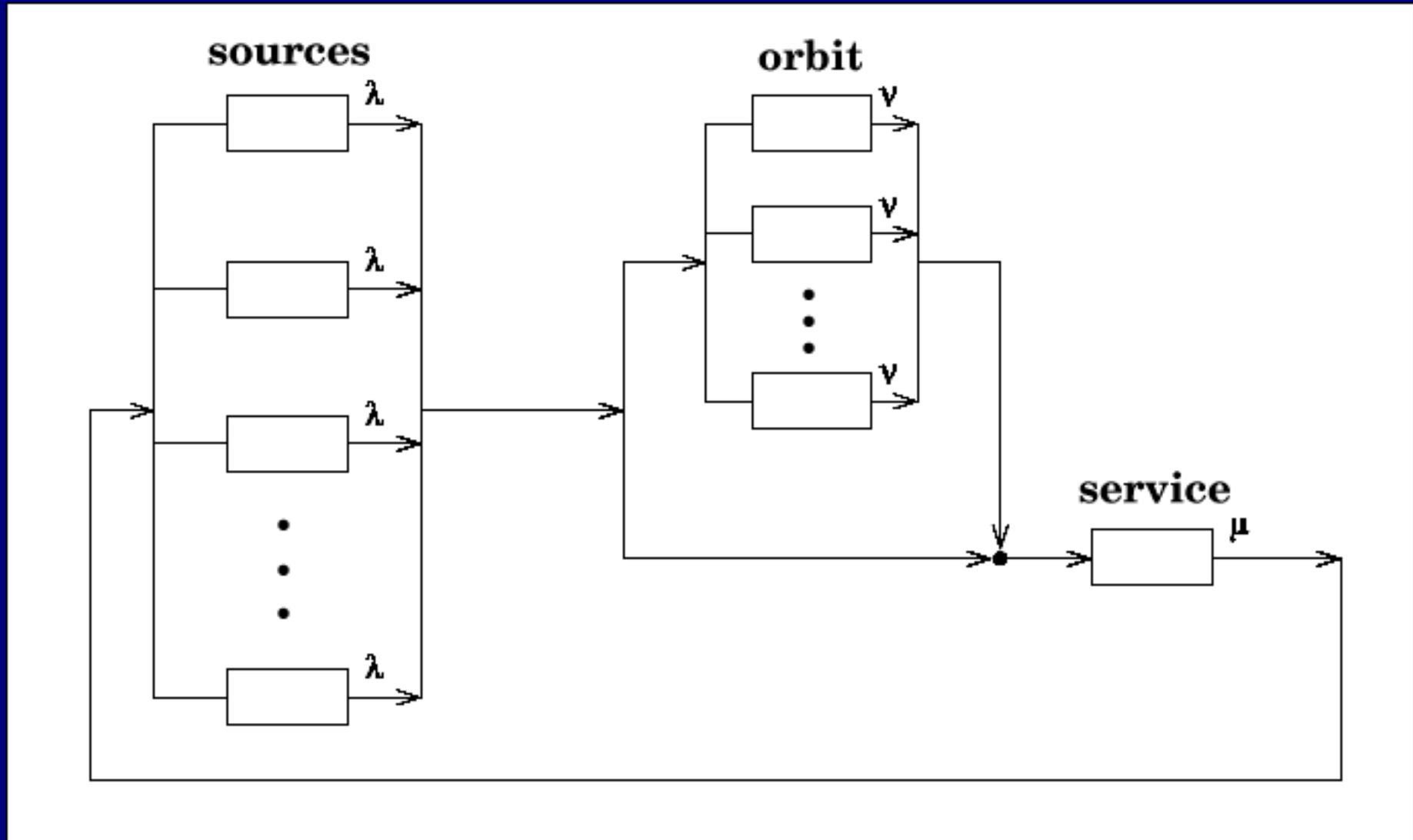Institute of Mathematics and Informatics, University of Debrecen
Debrecen, Hungary

e-mail: jsztrik@math.klte.hu
www: http://it.math.klte.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

$$X(t) = (Y(t); (\alpha_{C(t)}; \beta_1, ..., \beta_{N(t)}), t \geq 0)$$

where

$Y(t) = 0$ if the server is up,
$Y(t) = 1$ if the server is down,
$C(t) = 0$ if the server is idle,
$C(t) = 1$ if the server is busy,
$\alpha_{C(t)}$ is the index of the request under service at time t if the server is busy.
$N(t)$ is the number of sources of repeated calls at time t,
$\beta_j, j = 1, ..., N(t)$ is the indices of request staying in the orbit.

We define the stationary probabilities:

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

$$P(q; j; 0) = \lim_{t \to \infty} P(Y(t) = q; \alpha_1 = j; N(t) = 0),$$

$$P(q; 0; i_1, ..., i_k) = \lim_{t \to \infty} P(Y(t) = q; C(t) = 0; \beta_1 = i_1, ..., \beta_k = i_k),$$

$$P(q; j; i_1, ..., i_k) = \lim_{t \to \infty} P(Y(t) = q; \alpha_1 = j; \beta_1 = i_1, ..., \beta_k = i_k).$$

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

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

Hence the *server utilization*

$$U_S = \sum_{j=1}^{K} U_j.$$

2. Utilization of source i

$$U^{(i)} = \mathrm{P} \ (\text{ source } i \text{ generates a new primary call }).$$

3. Utilization of the repairman

$$U_R = E[Y(t)] = \sum_{j=0}^{K} \sum_{k=0}^{K-1} \sum_{i_1,...,i_k \neq j} P(1; j; i_1, ..., i_k).$$

4. Availability of the server

$$A_S = 1 - U_R.$$

5. Mean response time of source i

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

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

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

$$E[S_i] = E[D_i] + 1/\lambda_i \geq 1/\lambda_i,$$

where $E[D_i]$ denotes the mean delay time due to the failure of the server.

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

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

6. Mean waiting time of source i

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

7. Mean number of calls staying at the service facility

$$M = E[C(t) + N(t)] = \sum_{i=1}^{K} P^{(i)} = \sum_{i=1}^{K} (P_S^{(i)} + P_O^{(i)}) = \sum_{i=1}^{K} P_S^{(i)} + \sum_{i=1}^{K} P_O^{(i)}.$$

8. Mean rate of generation of primary calls

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

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

Validation of results

|	non–rel. retrial(cont.)	non–rel. retrial(orbit)	non–rel. FIFO
Number of sources:	3	3	3
Request's generation rate:	0.2, 0.3, 0.5	0.2, 0.3, 0.5	0.2, 0.3, 0.5
Service rate:	1, 1.2, 1.1	1, 1.2, 1.1	1, 1.2, 1.1
Retrial rate:	1e+20	1e+20	-
Server's failure rate:	0.002	0.002	0.002
Server's repair rate:	0.04	0.04	0.04
Utilization of the server:	0.578593008176	0.578593460071	0.578595143583
Mean response time			
Source 1:	1.61016598407	1.61027143737	1.6109393482
Source 2:	1.41365083148	1.41357129589	1.41287007613
Source 3:	1.35362123345	1.35362137877	1.35372999206
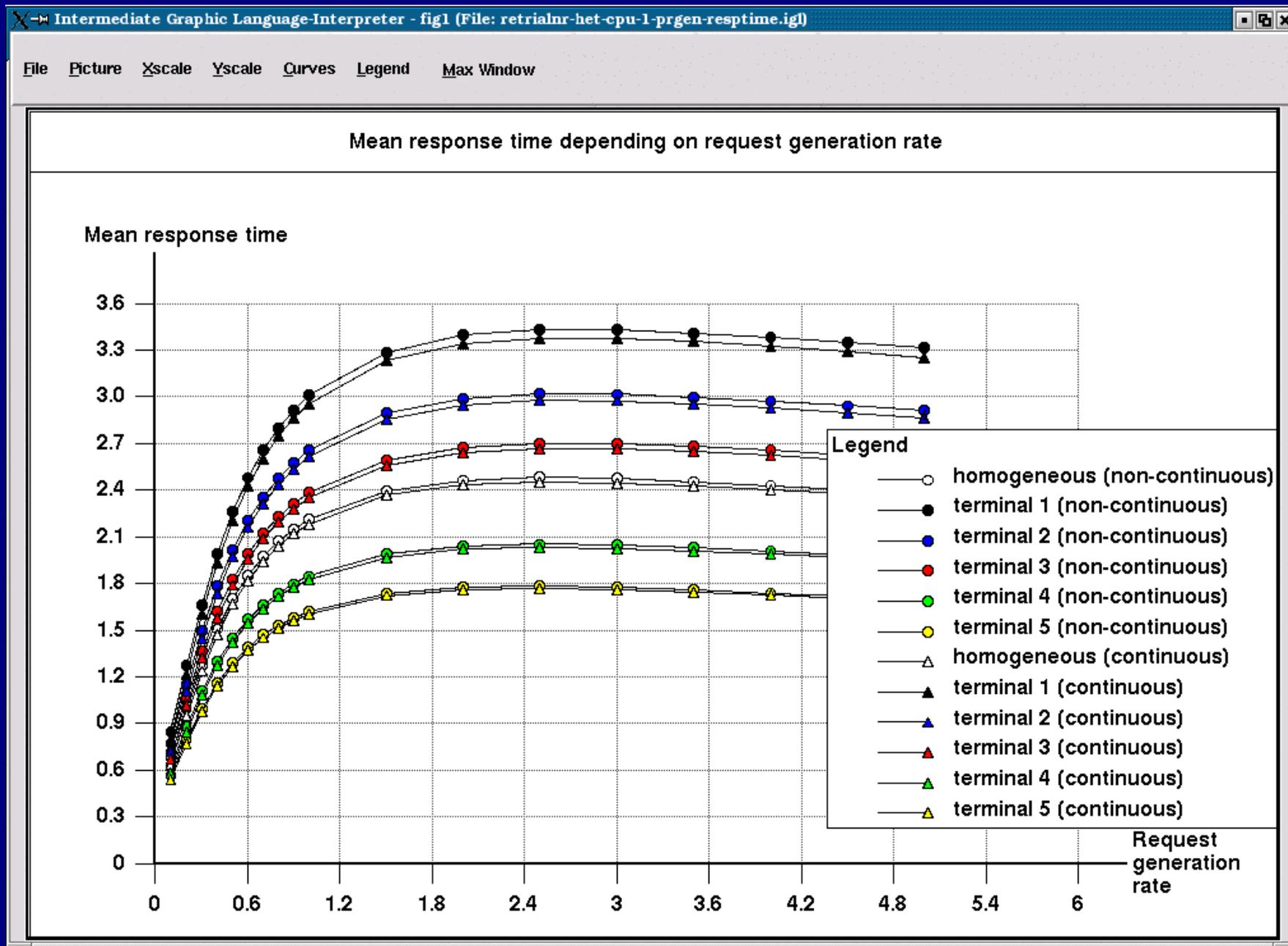
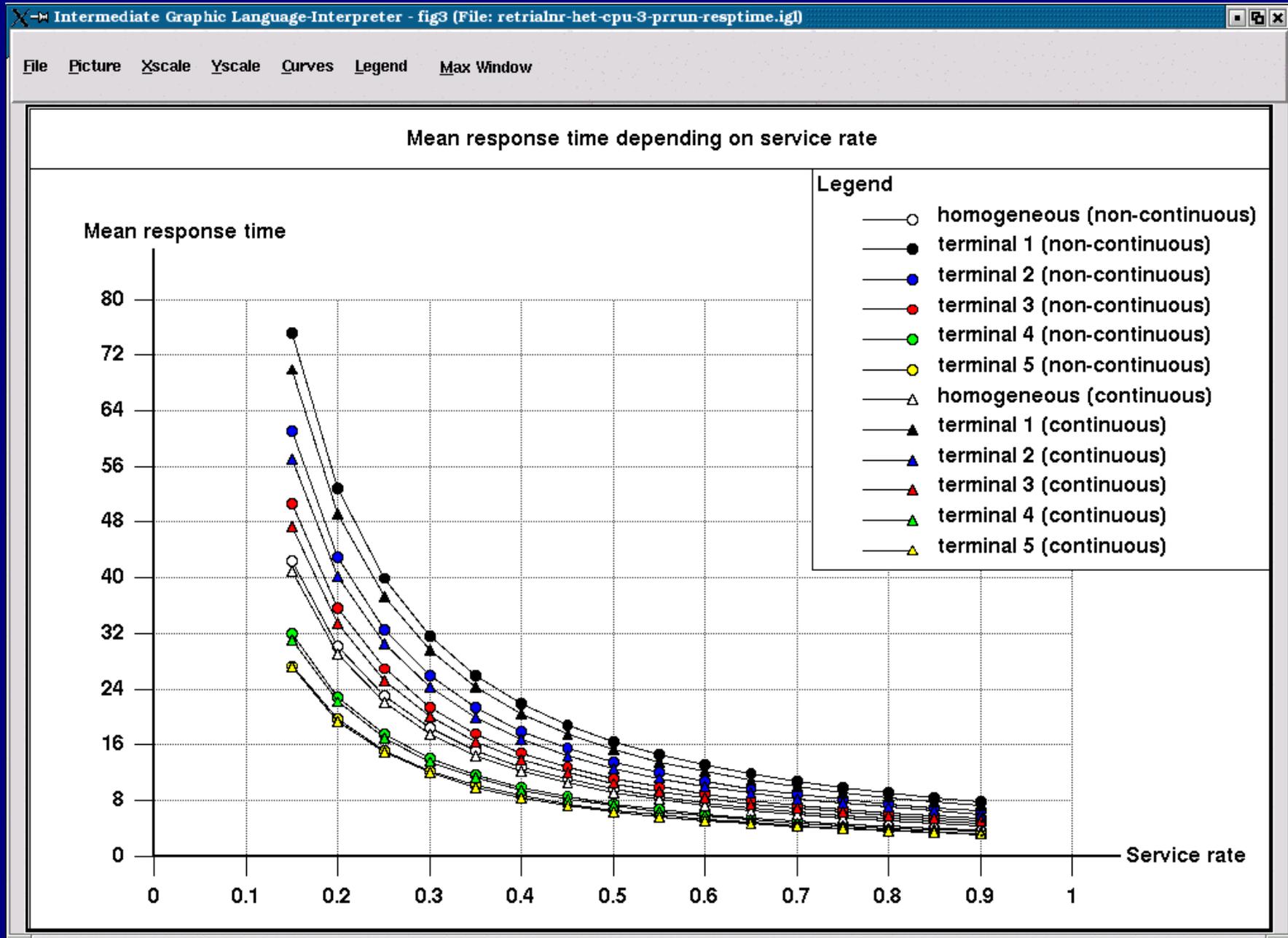*E*[*T*] versus primary request generation rate
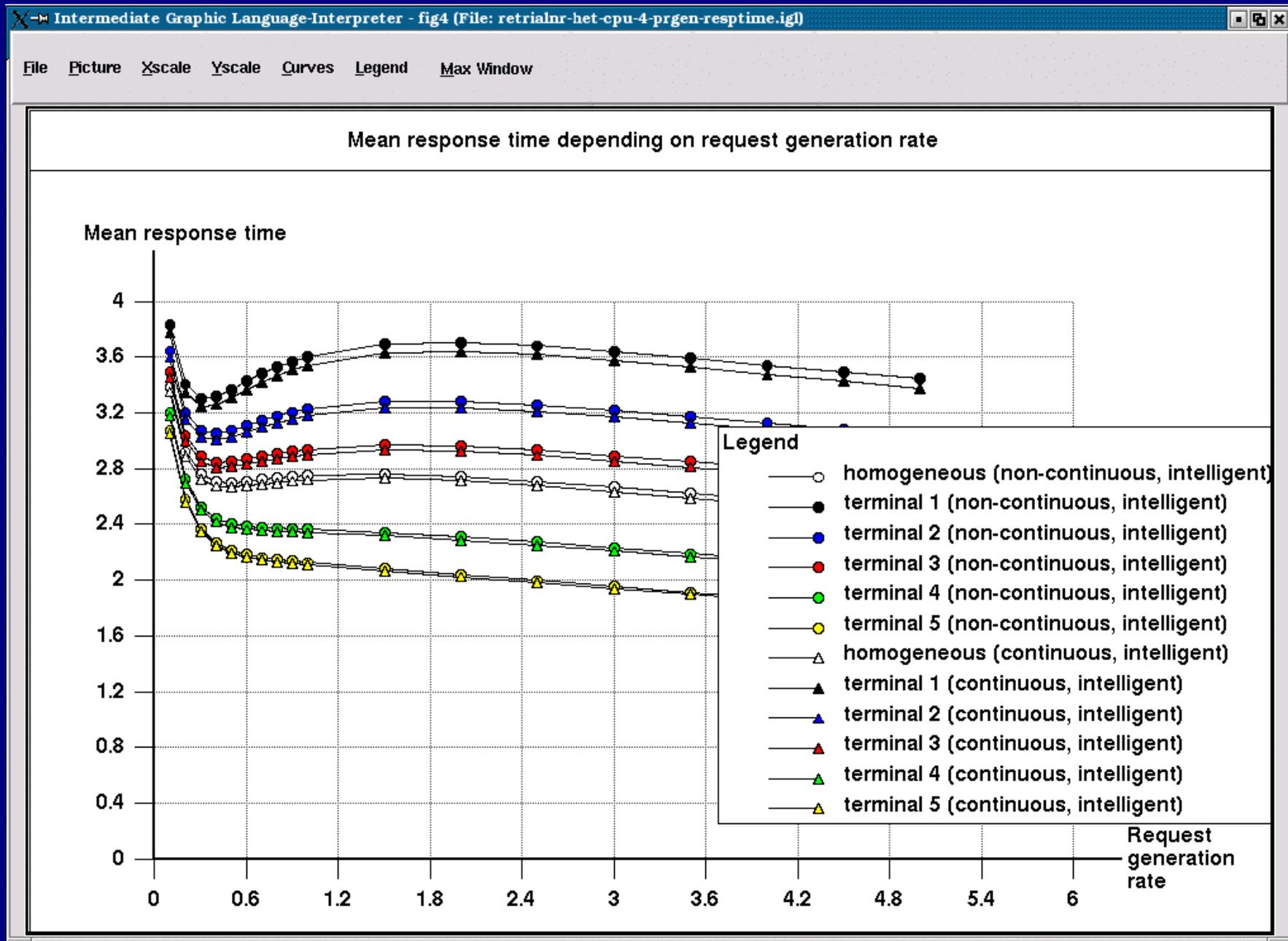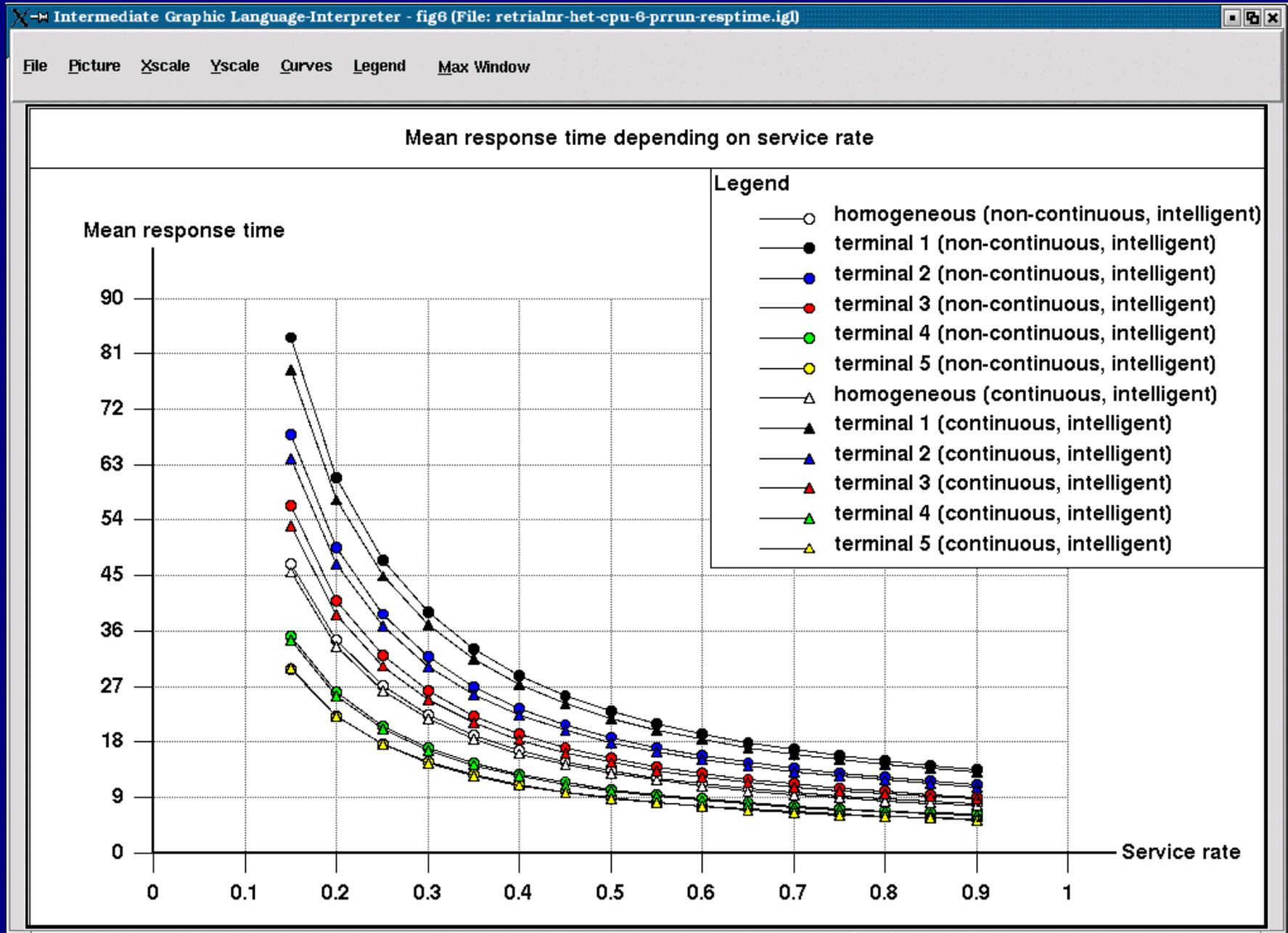
$E[T]$ versus primary request generation rate
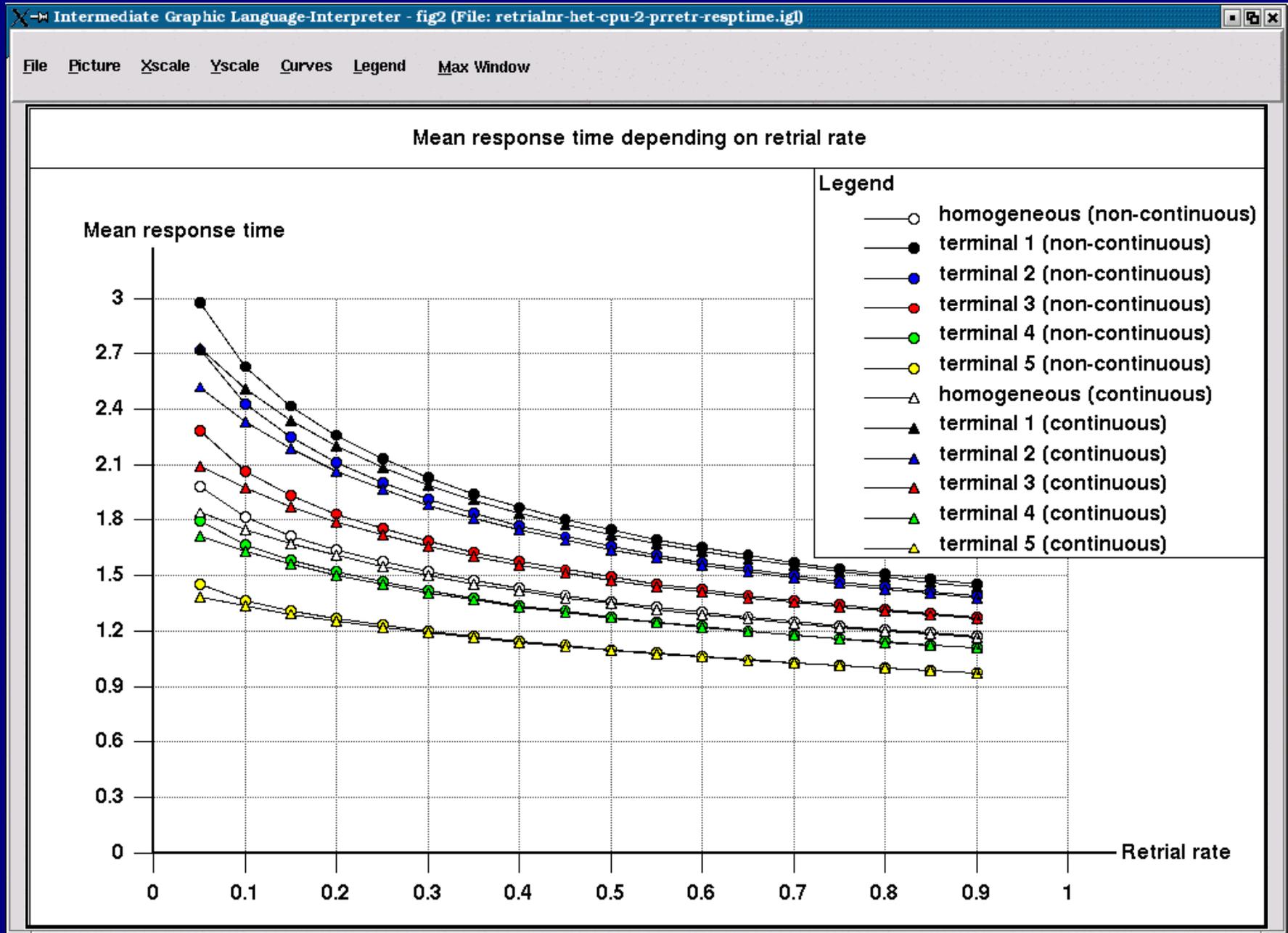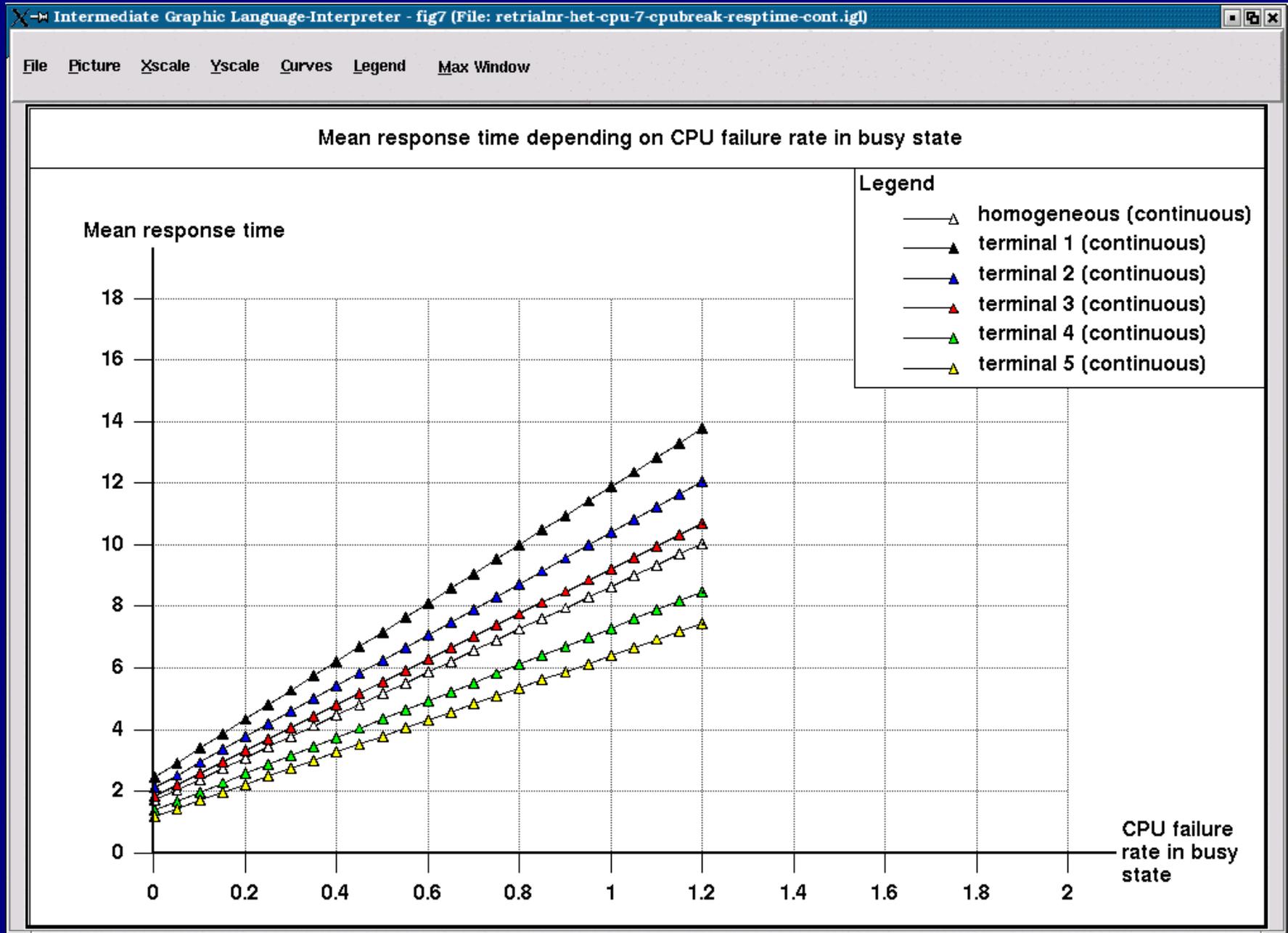
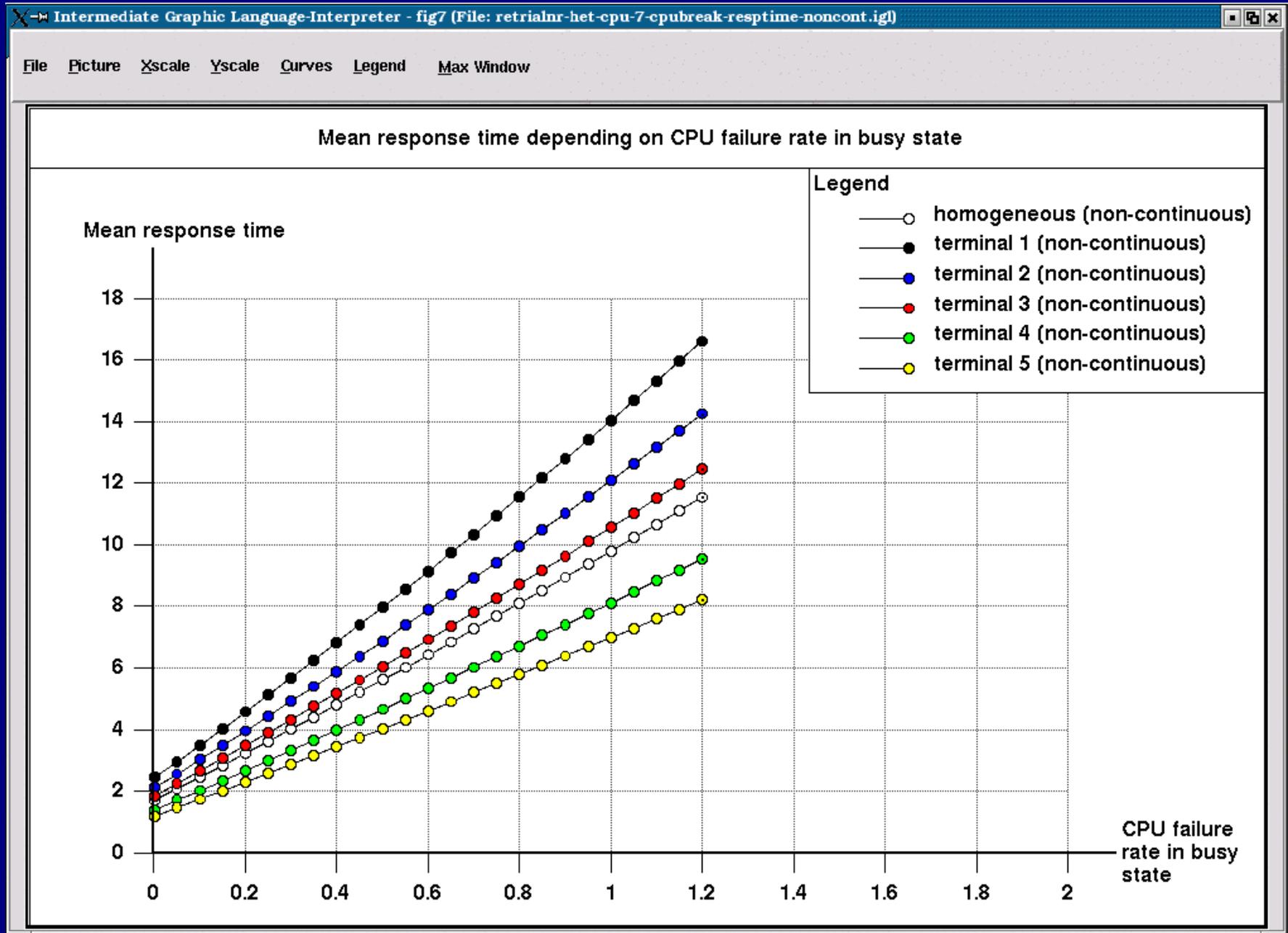$E[T]$ versus retrial rate
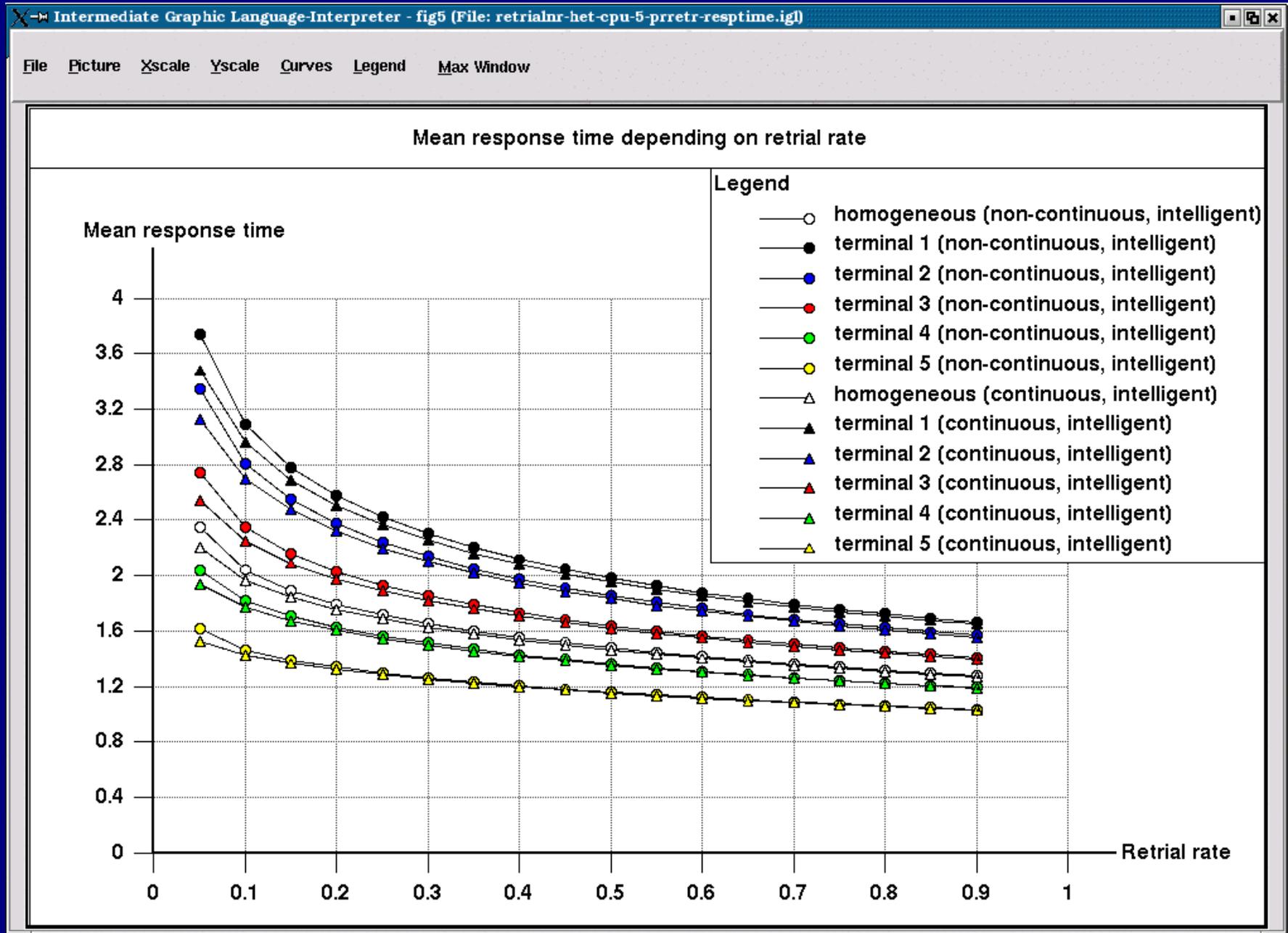
$E[T]$ versus retrial rate
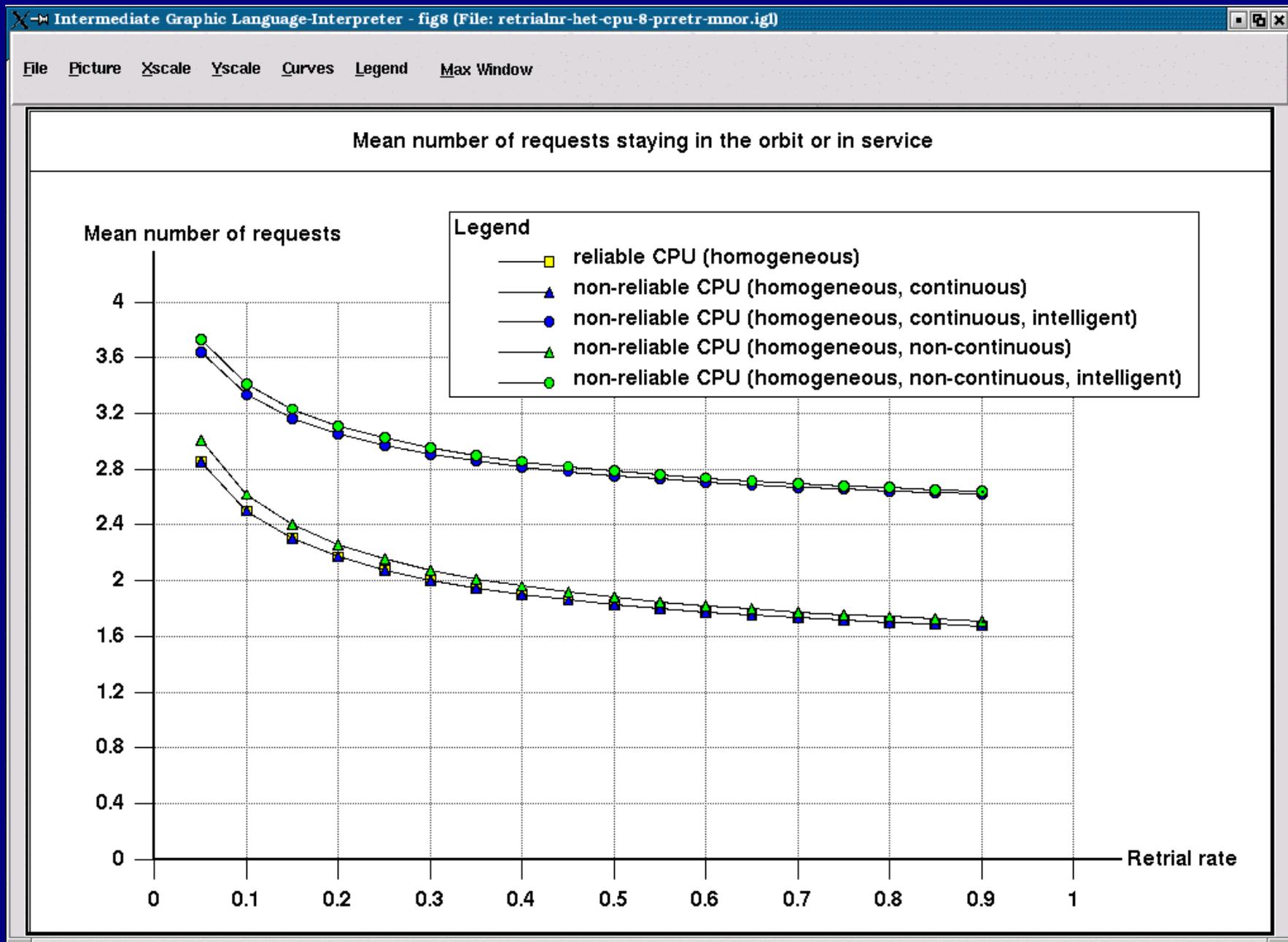
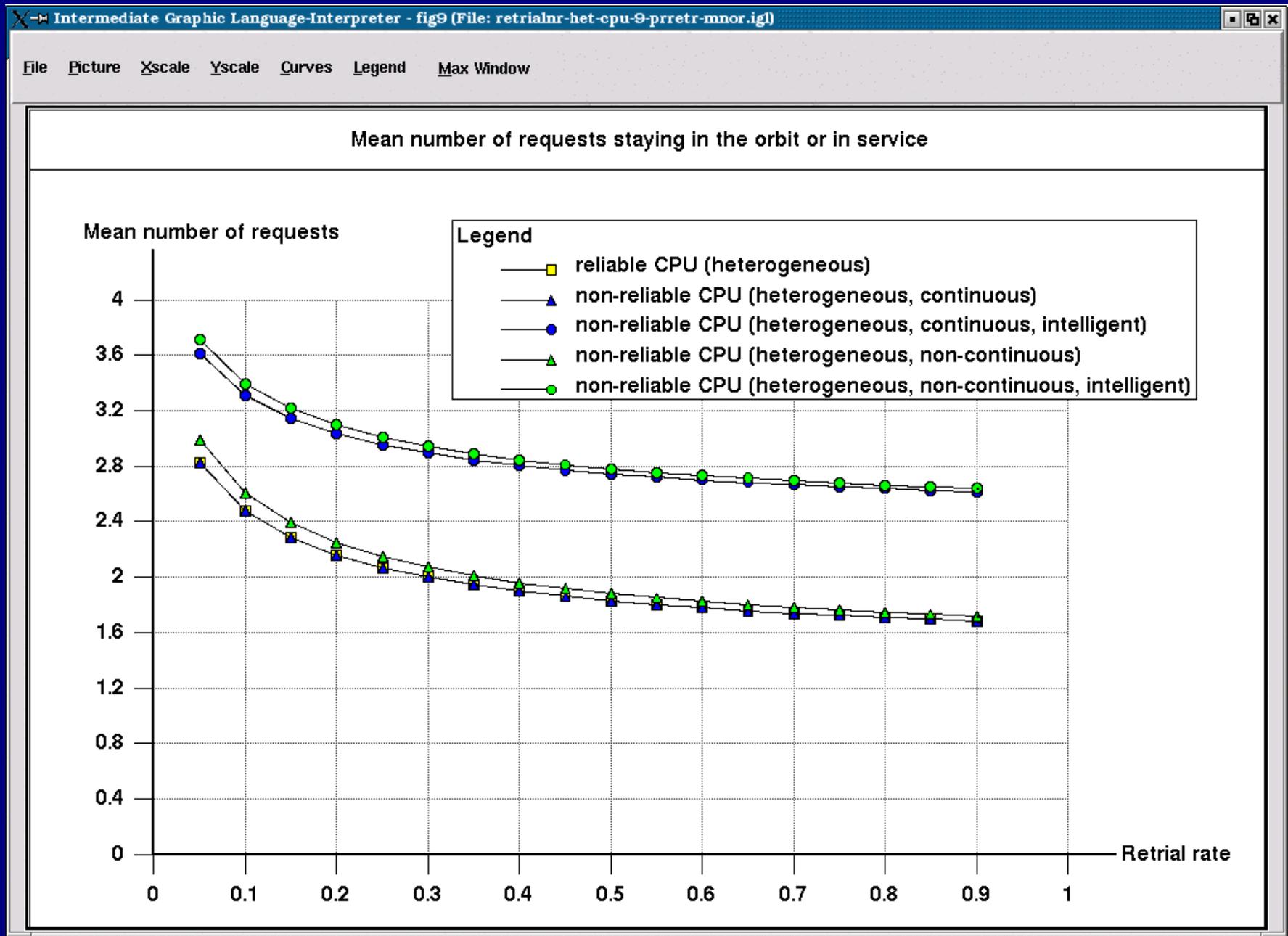$E[T]$ versus service rate

$E[T]$ versus service rate

$E[T]$ versus CPU failure rate in busy state

$E[T]$ versus CPU failure rate in busy state

M versus retrial rate

M versus retrial rate

References

[1] **Almási B., Bolch G. and Sztrik J.** Heterogeneous finite-source retrial queues, *Journal of Mathematical Sciences* (to appear).

[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.

[5] **Wang Jinting, Cao Jinhua and Li Quanlin** Reliability analysis of the retrial queue with server breakdowns and repairs, *Queueing Systems Theory and Applications* 38(2001), 363–380.