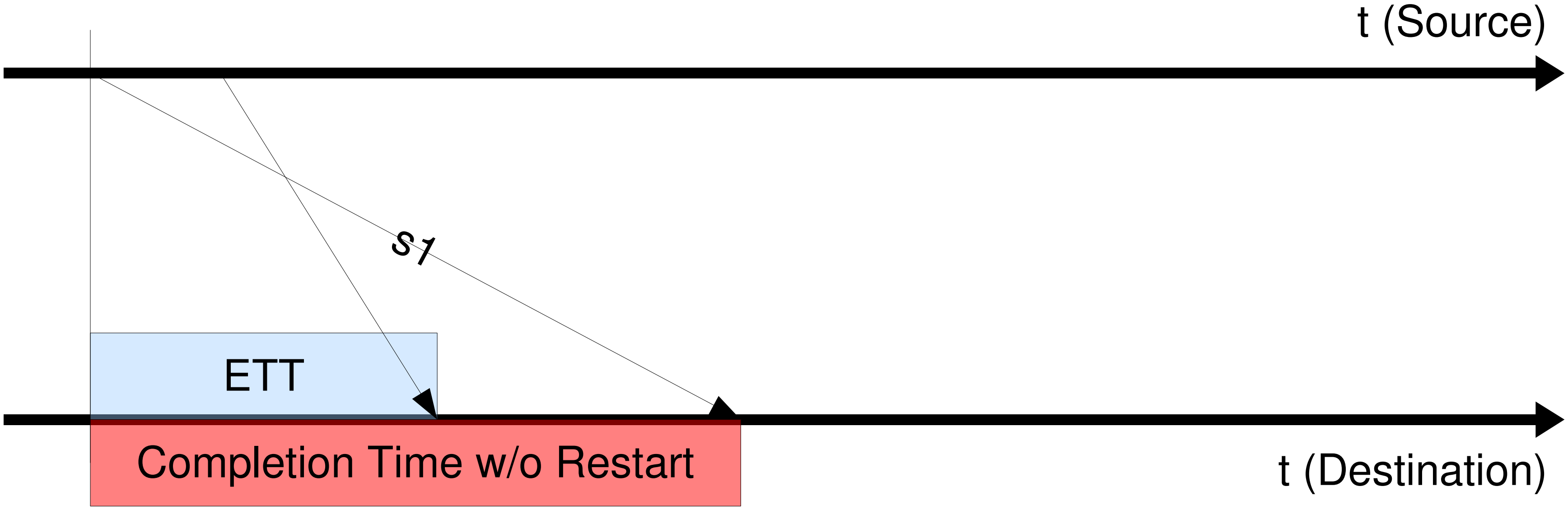
Phase-type Approximations for Message Transmission Times in WSRM

Philipp Reinecke and Katinka Wolter
{preineck, wolter}@informatik.hu-berlin.de

Short introduction to WSRM

- WSRM provides reliable SOAP message transmissions to Web Services
- WSRM destination sends acknowledgement messages for messages it has received
- WSRM source restarts unacknowledged message transmissions upon timeout expiration
- Metric considered here: Effective Transmission Time (ETT) – Time after which the message is available at the destination.



Experiment Setup

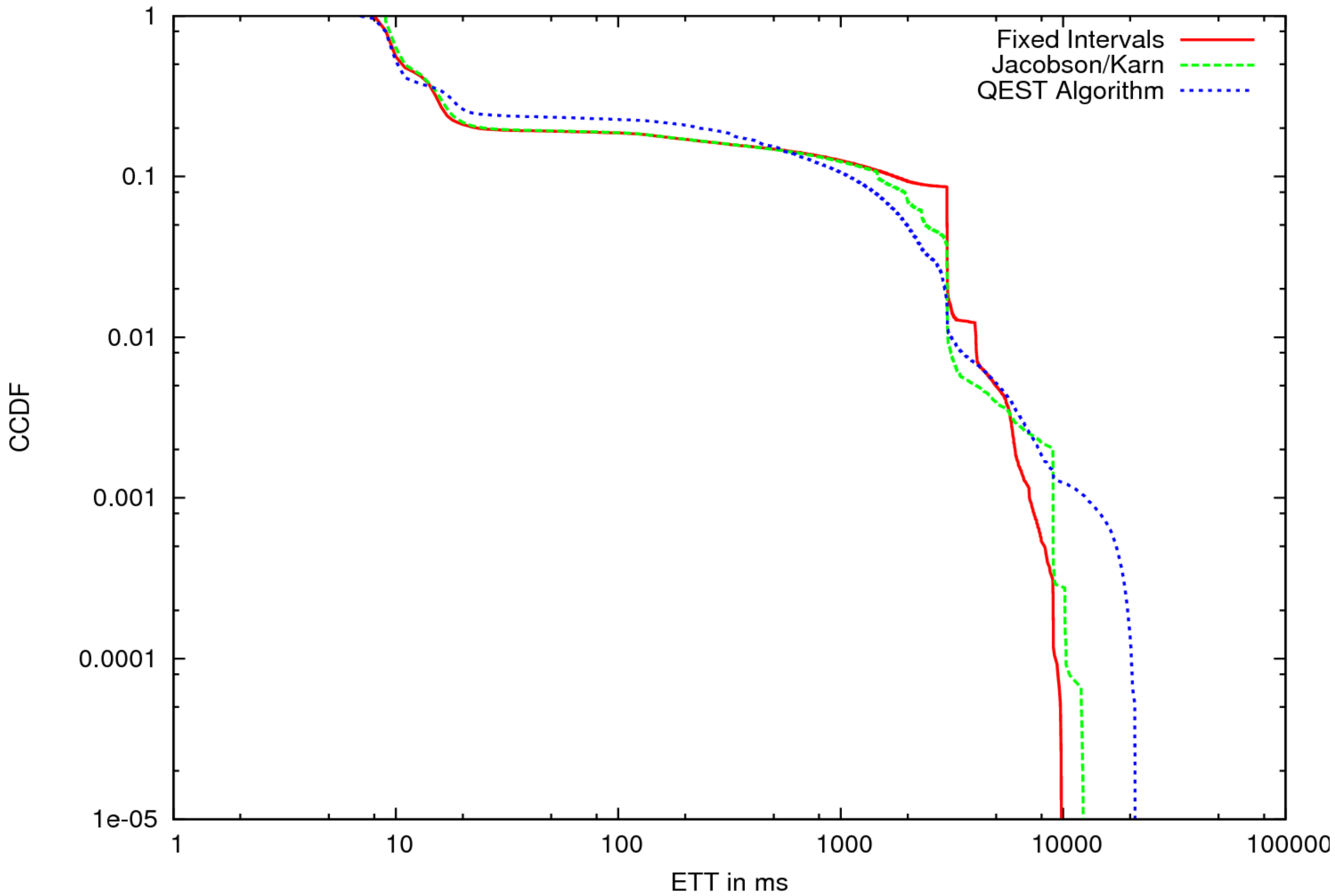
- Experiment setup:
  - real-world WSRM implementation (Sandesha)
  - two-state Gilbert model to inject IP faults
  - consider three loss levels
- Measure message transmission times (ETT)
- Restart strategies:
  - Fixed Intervals (4s)
  - Jacobson/Karn
  - QEST Algorithm

Properties of the data

	S1FI	S1JK	S1Q	S2FI	S2JK	S2Q	S3FI	S3JK
Mean	19.38	20.32	19.22	120.68	117.1	105.37	403.33	275.83
Std. Dev.	63.41	49.96	22.86	538.25	466.86	399.14	1021.36	692.76
Min.	9	12	12	9	12	12	8	9
Median	13	16	15	10	16	15	10	13
95% q	25	23	23	161	267	495	3009	1879
99% q	149	160	162	3013	2551	2516	4059	3017
Max.	3017	2106	328	4465	3863	3366	9018	5467
CoV	10.71	6.04	1.41	19.89	15.9	14.35	6.41	6.31

Do not look like exponential distributions

Distribution of the data (S3)



- Long tails
- `Steps' at the values of the TCP RTO timeouts (3s, 6s, 9s, ...)

ACPH Approximations

- Bi-diagonal representation of the CTMC:

$$\hat{Q} := \begin{bmatrix} Q & q \\ 0 & 0 \end{bmatrix}$$

$$\hat{\alpha} := (\alpha_{1,} \alpha_{2,} \ldots, \alpha_N, \alpha_{N+1})$$

$$Q = \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & -\lambda_2 & \lambda_2 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & -\lambda_N \end{bmatrix}$$

Special ACPHs

- Second-Order ACPH (ACPH(2)):

$$\alpha = (\alpha, 1-\alpha)$$

$$Q = \begin{bmatrix} -\lambda_1 & \lambda_1 \\ 0 & -\lambda_2 \end{bmatrix}$$

- Hyper-Erlang Distribution (HErD):
  - mixture of M Erlang distributions with different rate and shape parameters

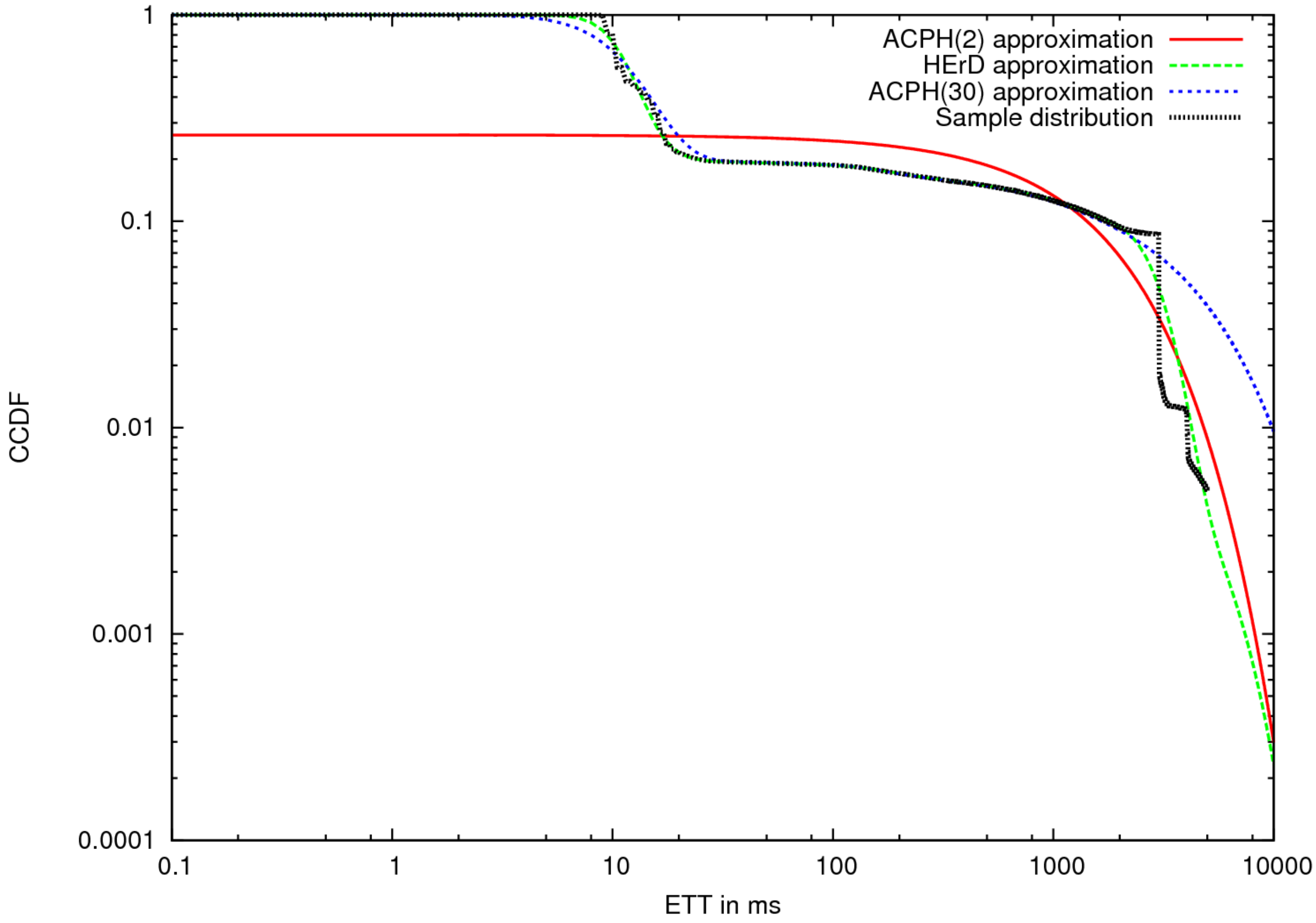
$$\alpha = (\alpha_{1,} 0, \ldots, 0, \alpha_{2,} 0, \ldots, 0, \alpha_M, 0, \ldots, 0)$$

$$Q = \begin{bmatrix} \ddots & \ddots & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\lambda_r & \lambda_r & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & \ddots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\lambda_r & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 & 0 \end{bmatrix}$$

Fitting procedures

- ACPH(2): Moment-fitting (Heindl/Telek):
  - fit first three moments
  - use approximate for third moment if exact fit is impossible
- HErD: G-Fit (Thuemmler/Buchholz/Telek):
  - logarithmic aggregation (Panchenko/Thuemmler)
  - 15 branches, increasing shape parameters (1, 2, ..., 15)
- ACPH(30): PhFit (Horvath/Telek):
  - logarithmic aggregation
  - 30 phases, no special tail

Results: Approximations



Evaluation: Error Measures

- Relative error in the first three moments:

$$e_i = \frac{\left| c_i(\hat{F}) - c_i(F) \right|}{c_i(F)}$$

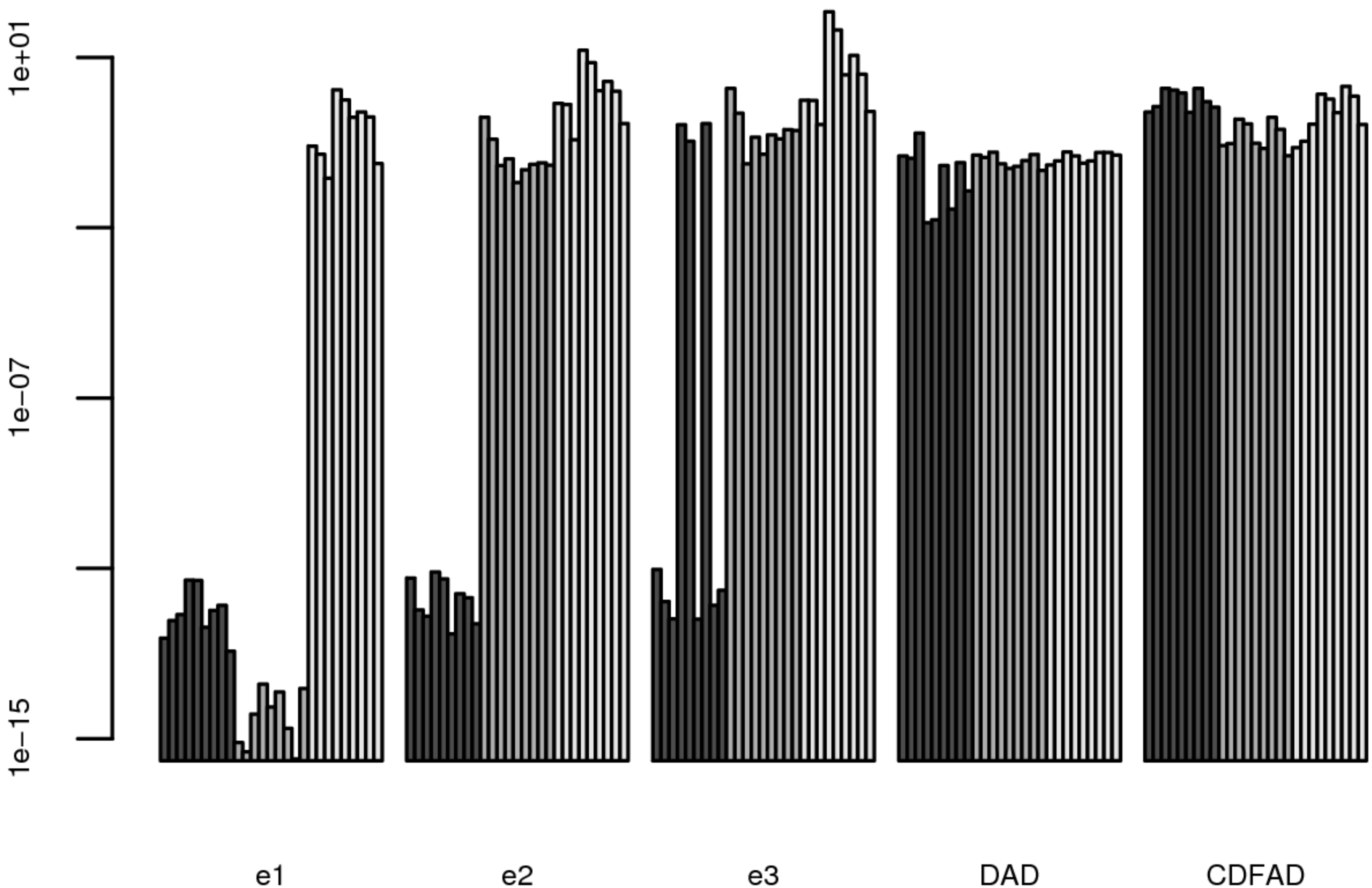
- Absolute PDF area distance:

$$PDFAD = \int_0^\infty \left| \hat{f}(t) - f(t) \right| dt$$

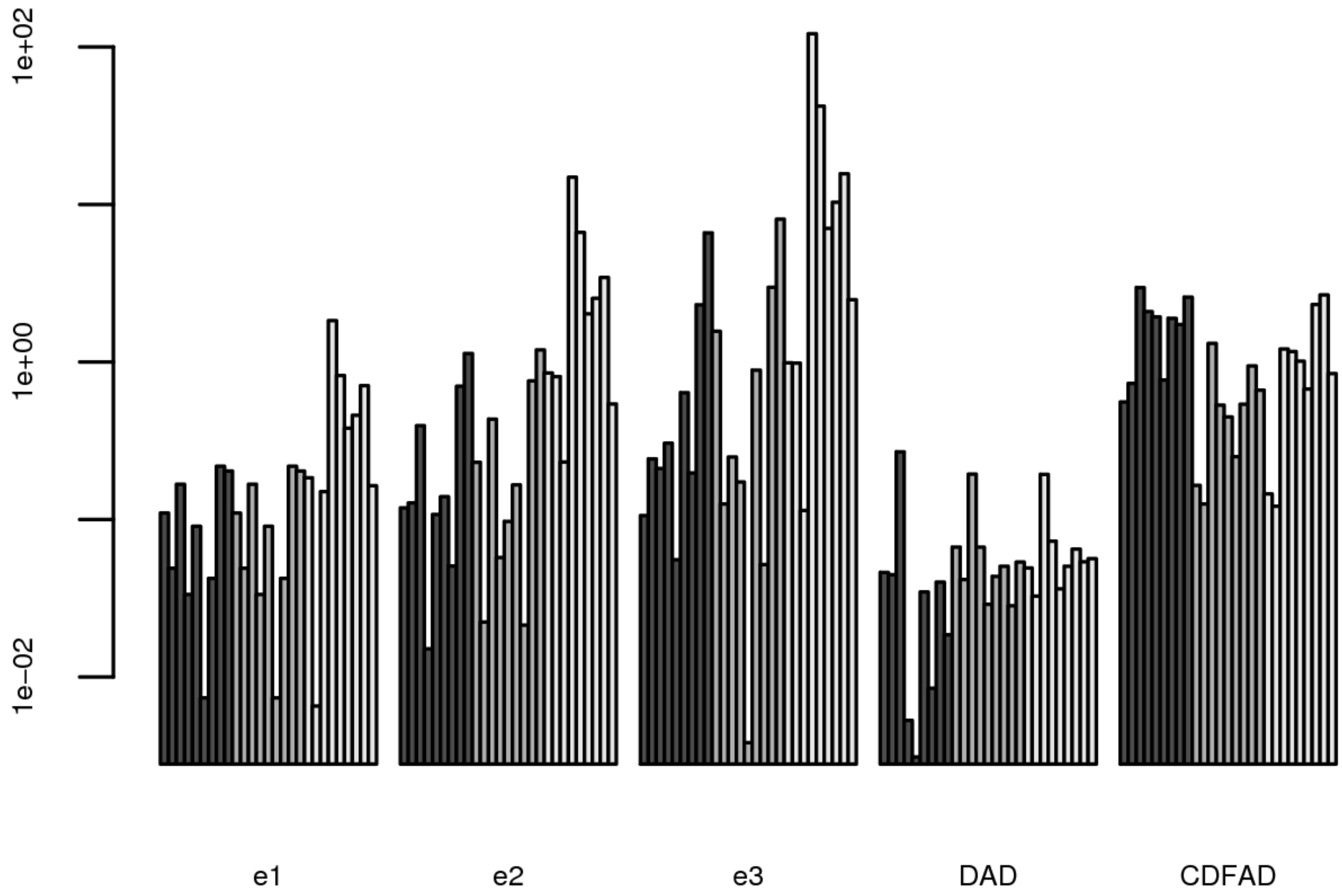
- Absolute CDF area distance:

$$CDFAD = \int_0^\infty \left| \hat{F}(t) - F(t) \right| dt$$

Evaluation using original samples



Cross-Evaluation



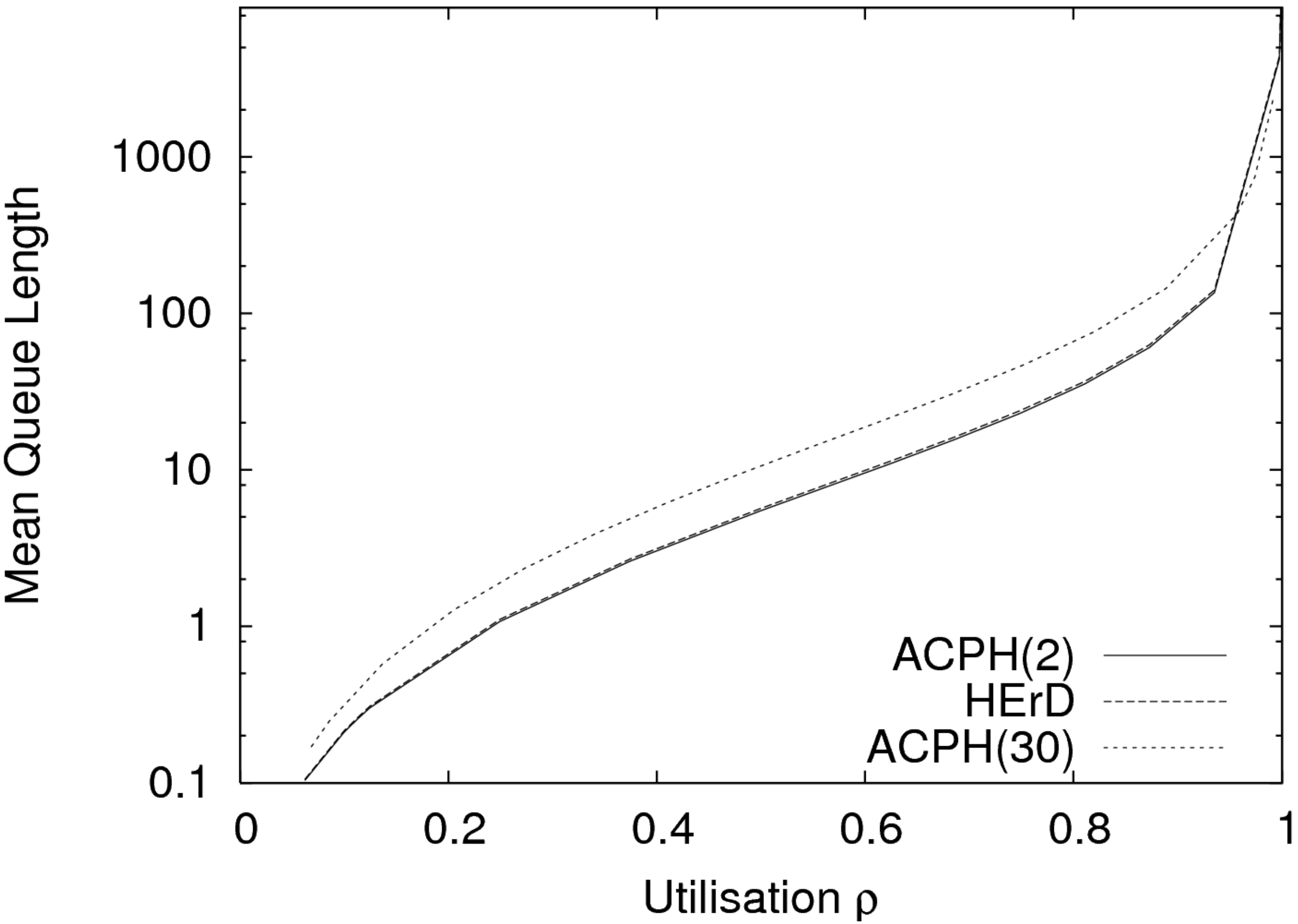
Example Application

- Model SOAP transmissions over WSRM using an M|PH|1 queue
- Matrix-geometric solution for the expected queue length:

$$E[N] = z(I - R)^{-2} 1$$

- Vary arrival rate λ to obtain various utilisations ρ

Mean queue length



Conclusion

- We fit ACPH models to large data sets in WSRM
- Small ACPH(2) models may be sufficient
- Models are available for download (Matlab format) at http://www.informatik.hu-berlin.de/~preineck/acphmodels
- Future work:
  - Fit ADPH models to the data sets (may fit steps and gaps better)
  - Consider special tail fitting with PhFit
  - Apply the approximations in some analyses

Fin.