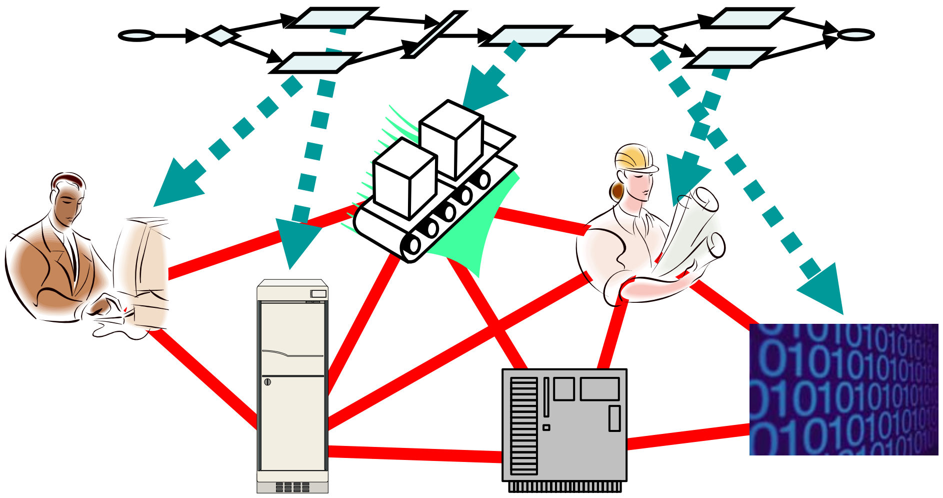
Performance Evaluation of Workflows Using Continuous Approximation of Discrete Sets and Probability Distributions

Kunihiko Hiraishi
School of Information Science
Japan Advanced Institute of Science and Technology

To make information systems dependable,

- Qualitative $\Rightarrow$ Verification of logical correctness
  - Formal verification (e.g., Model checking, Theorem proving)
- Quantitative $\Rightarrow$ Guaranteeing performance index
  - Performance modeling (e.g., queuing theory)

Workflows / Business Processes



Workflows integrate people, systems and information

Process definition

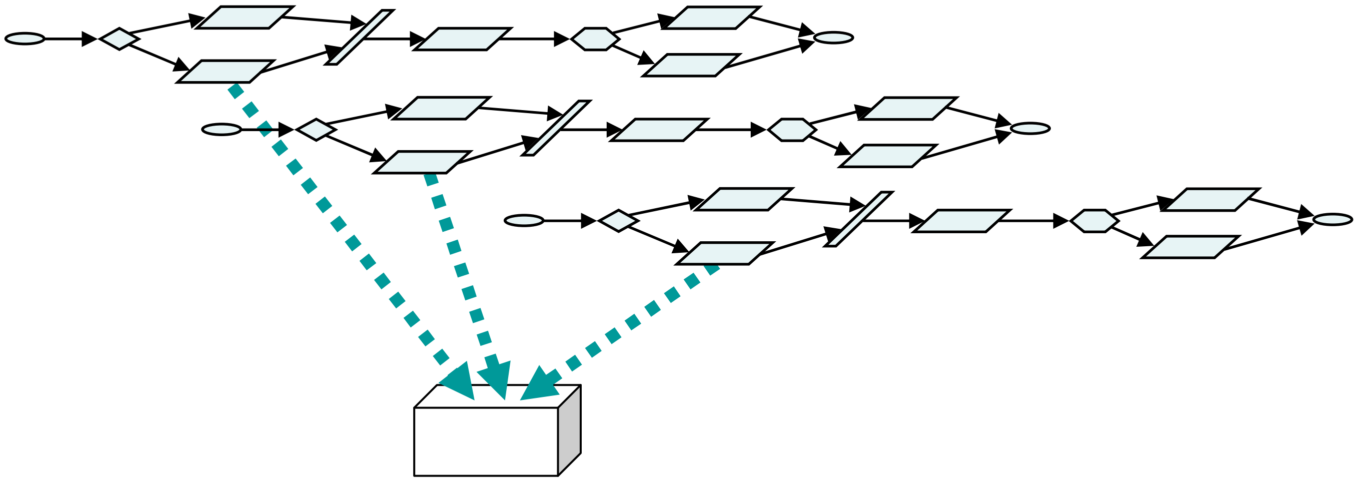
WF Management System

Activities in the Real World

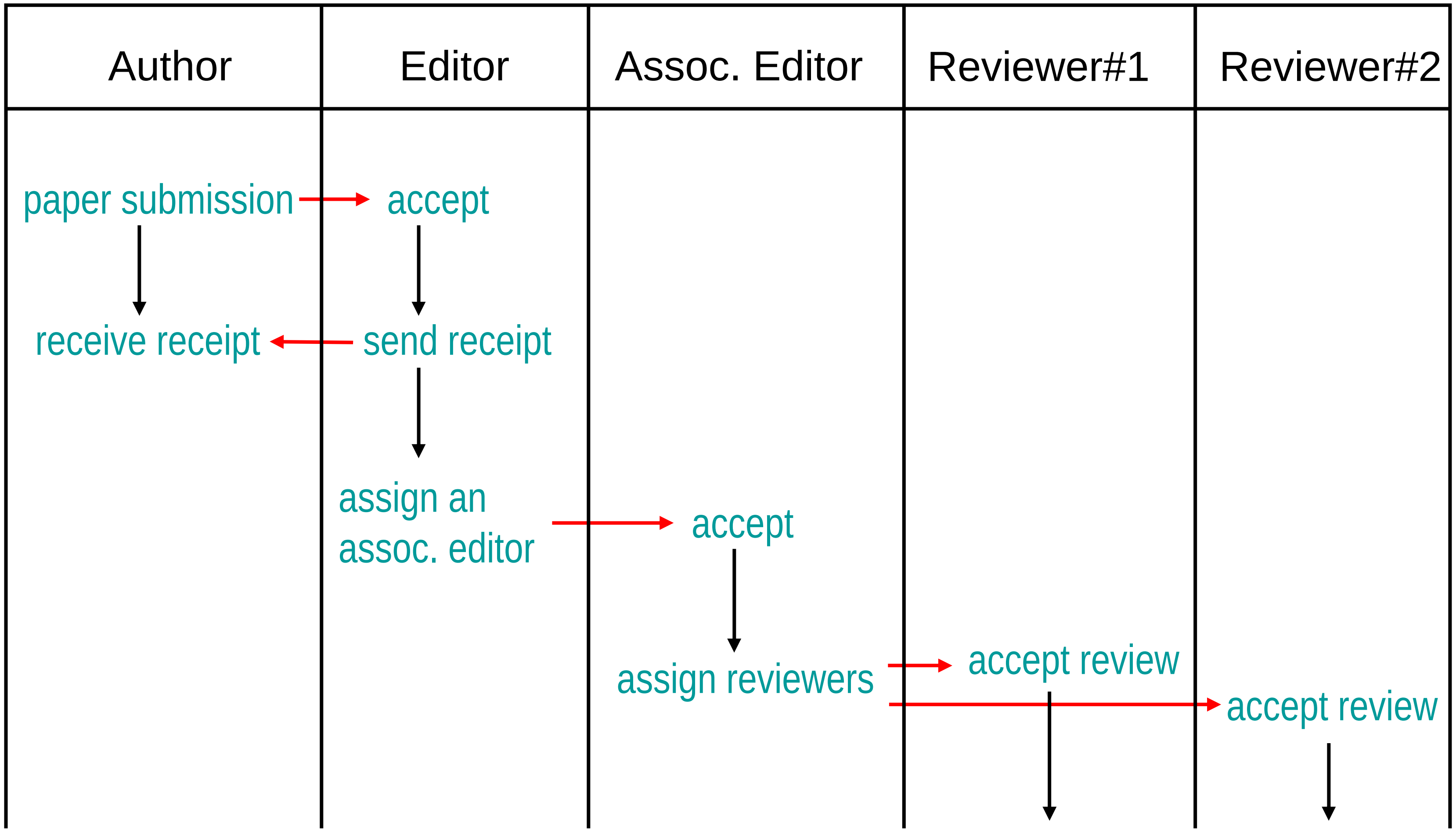
Workflow: The automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules. (Def. by WFMC)

Performance Evaluation of Workflows

- Each workflow is a template of a business process.
- Many instances of workflows are running simultaneously in the information system.
- Optimal resource (people, machines, time, ...) assignment is a crucial issue.



Example:
Review Process of an Academic Journal

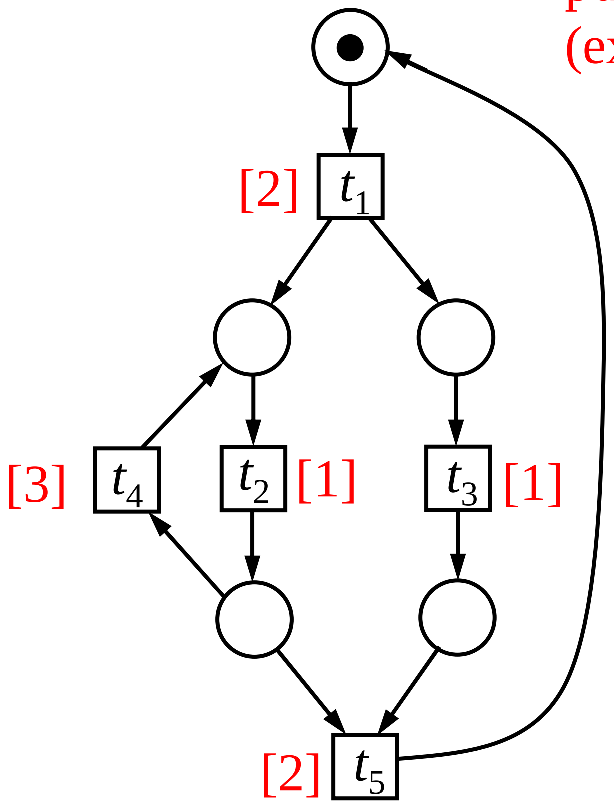


Example:
Review Process of an Academic Journal

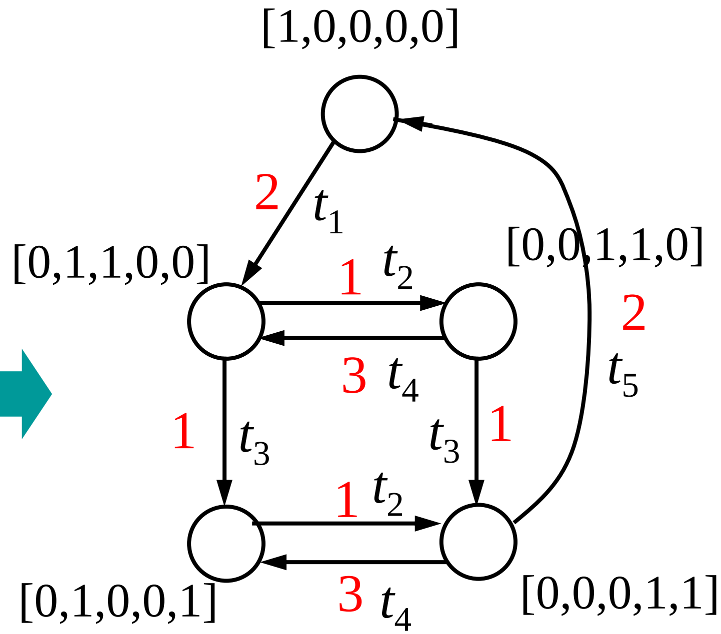
- Given:
  - Workflow
  - Statistical data on paper submission
  - An upper bound of the number of papers each associate editor can handle
- Find:
  - The optimal number of associate editors
- Method:
  - Generalized Stochastic Petri net
  - Approximation by Extended Continuous Petri Nets

Analysis of Stochastic Petri Nets

[λ_i] firing rate.
pdf of firing delay : $\lambda e^{-\lambda_i t}$
(exponentially distributed)



Stochastic Petri Net

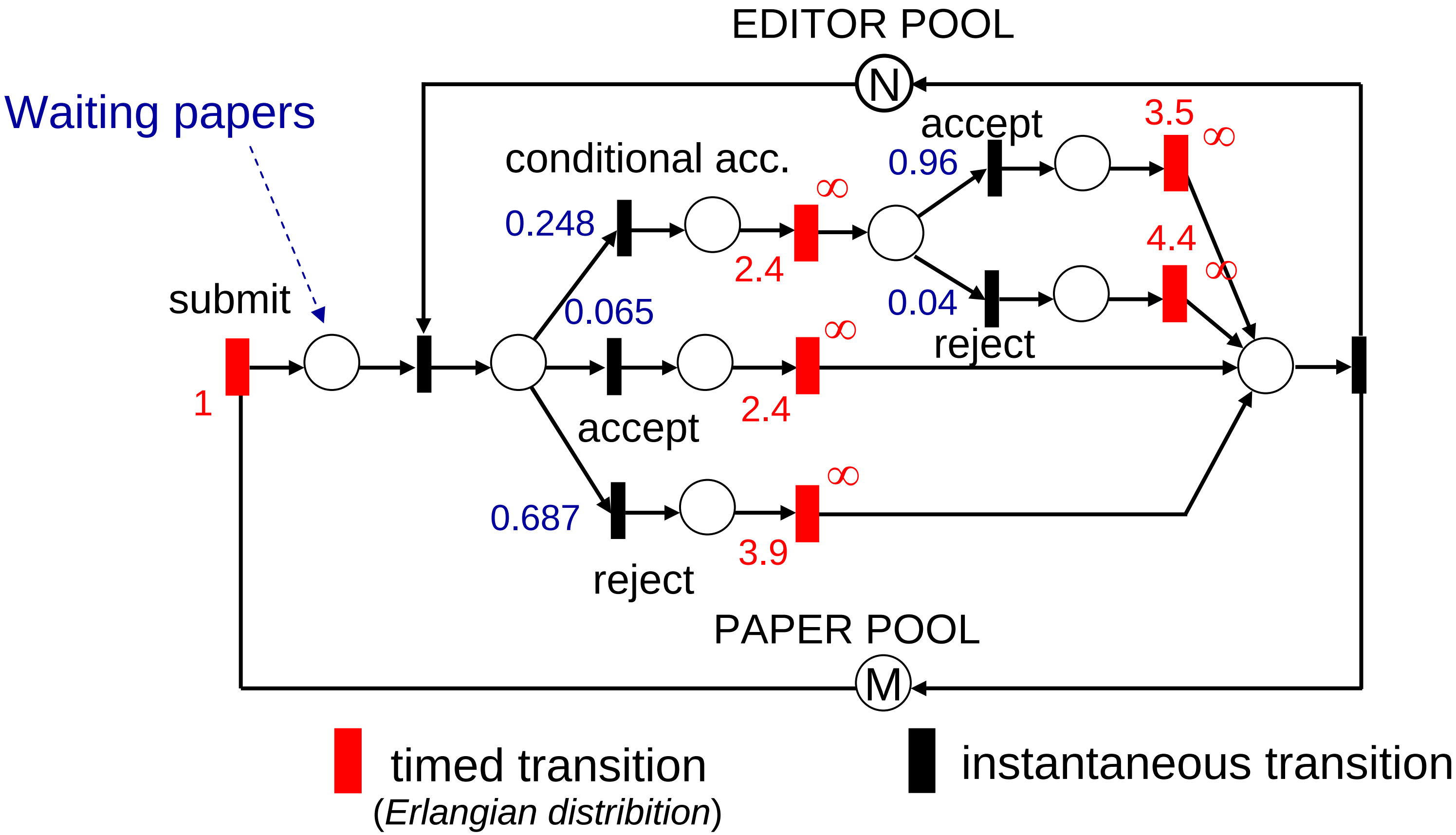


Continuous Time
Markov Chain

Statistics

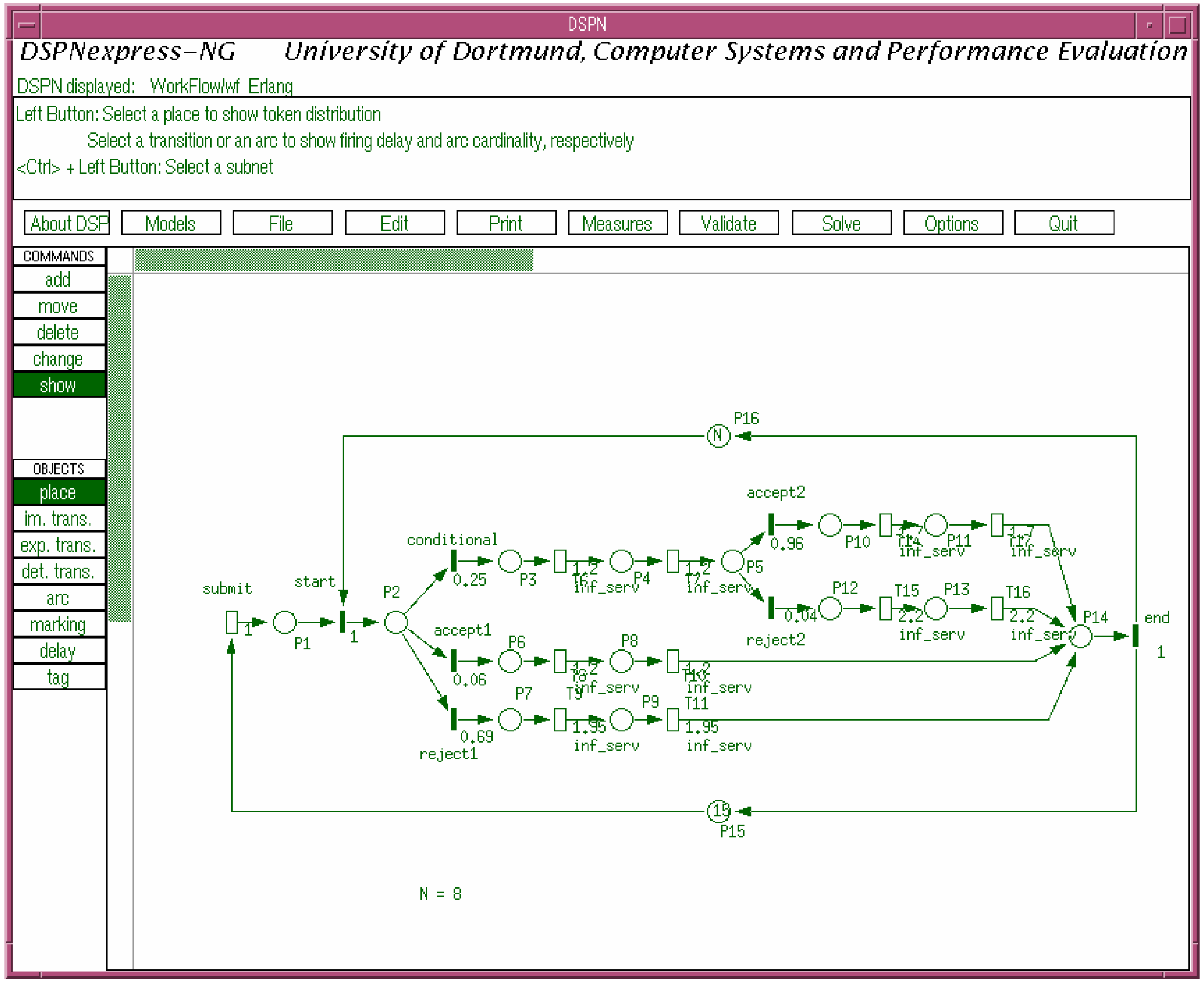
- Duration between submission and final judgment
  - Accept at 1st review: 2.4 month
  - Reject at 1st review: 3.9 month
  - Accept at 2nd review: 5.9 month
  - Reject at 2nd review: 6.8 month
- Ratio of acceptance and rejection
  - Accept at 1st review: 0.065
  - Reject at 1st review: 0.687
  - Accept at 2nd review: 0.238
  - Reject at 2nd review: 0.010
- Average number of paper submissions: 16.9 / month

GSPN Model



PAPER POOL is necessary for the model to have finite state space.

Tool (DSPNexpress)



Result

N	#states	CPU Time (sec.)	#Waiting papers	p(#paper pool = 0)
3	2926	0.3	10.18	0.30
4	8866	0.7	5.94	0.094
5	23023	2.3	1.99	0.013
6	53053	6.2	0.63	0.0021
7	110968	15	0.21	0.00049
8	213928	29	0.08	0.00020
9	384098	58	0.03	0.00010

Itanium2 1.6GHz/9MBCache, 16GB Memory

1 token = 16.9 papers

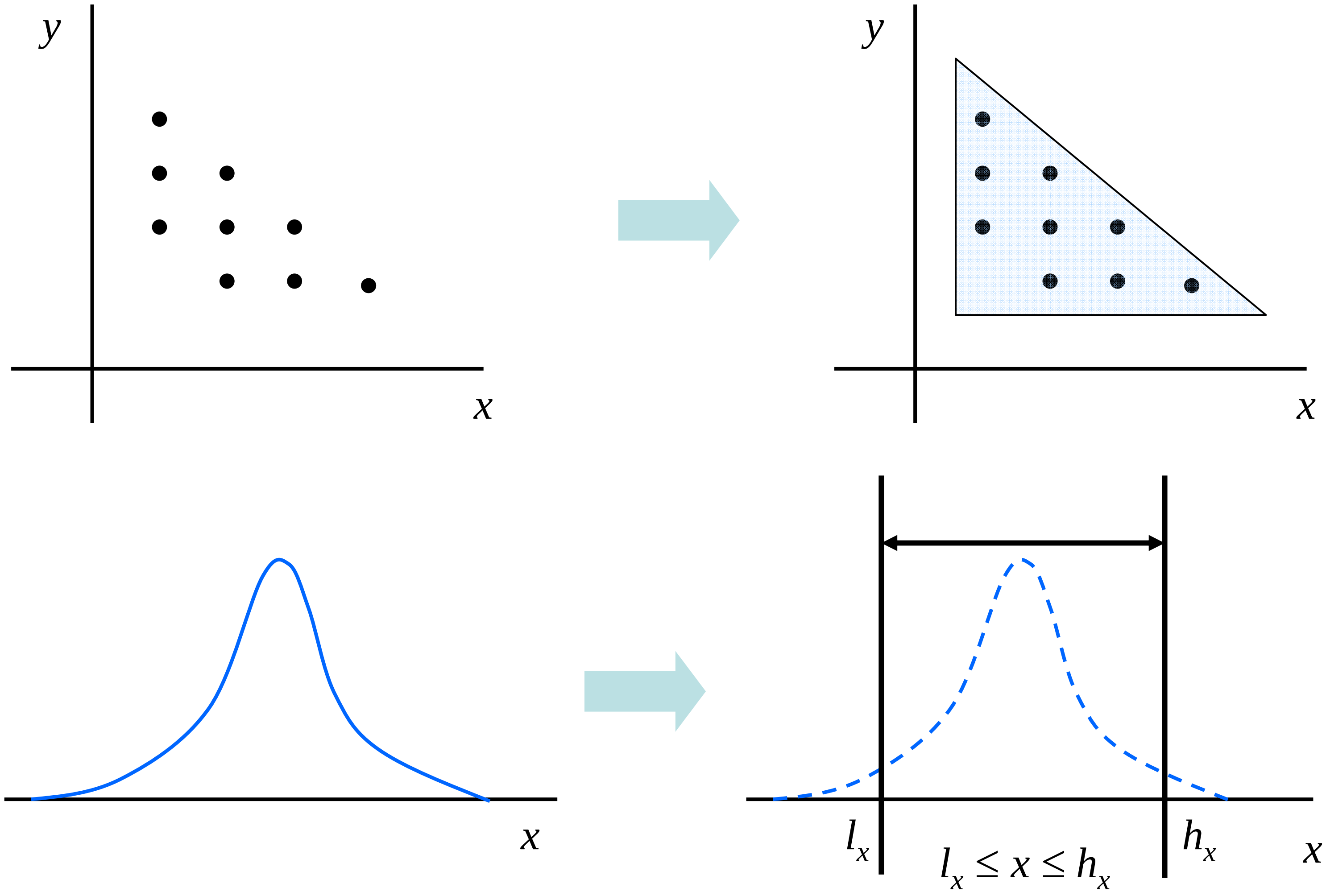
N = 6 ⇒ processing power : P = 6×16.9 = 101.4 papers simultaneously

5 papers / person ⇒ 101.4/5 = 20.28 associate editors are necessary

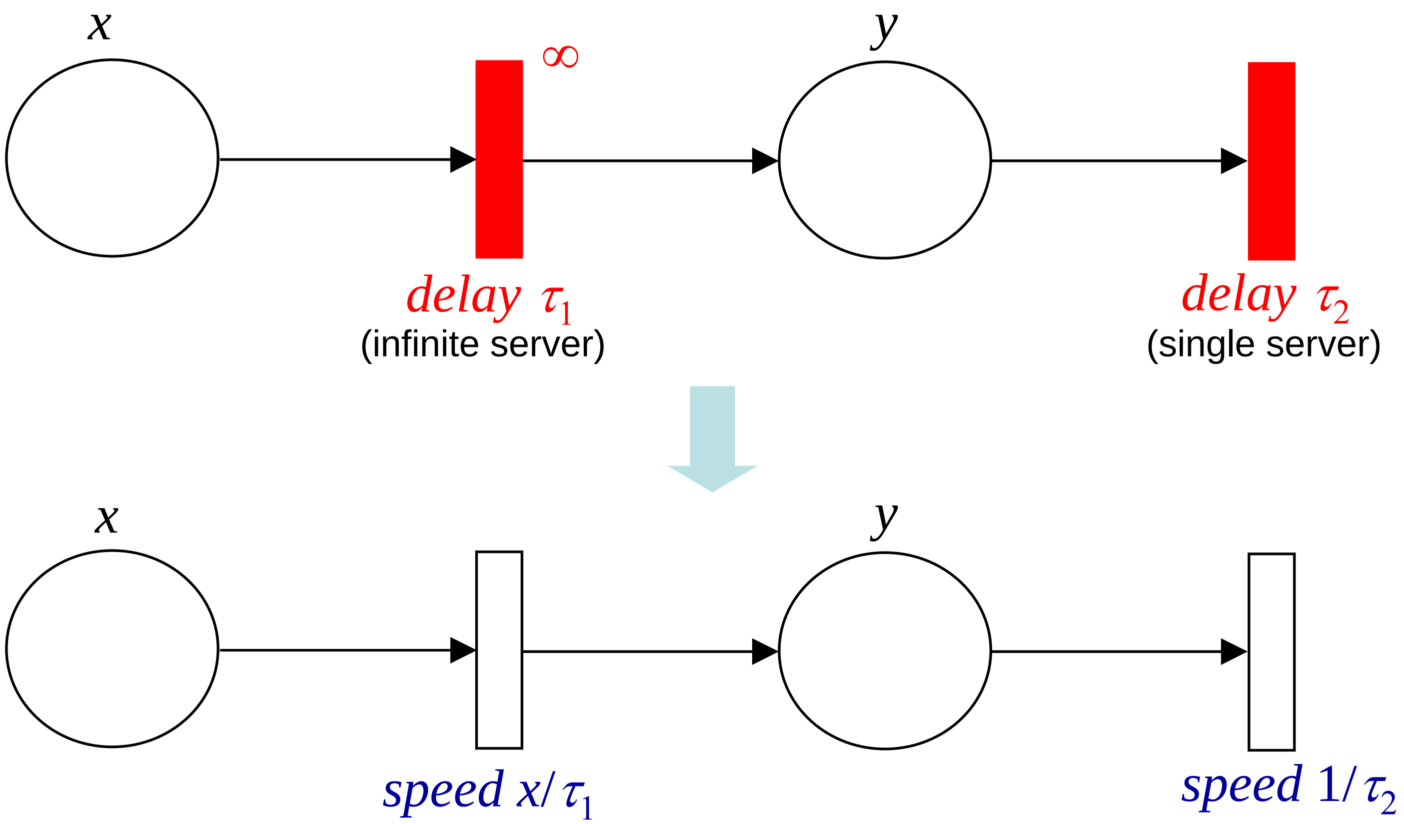
Continuous Approximation

- Analysis by GSPN is costly, because of *state space explosion*.
- A large number of discrete resources can be approximated by *a continuous quantity*.
- We first make a hybrid Petri net model from the GSPN model as follows:
  - tokens in a place ⇒ a continuous variable
  - state space ⇒ polyhedral approximation
  - firing delay of a timed transition ⇒ firing speed of a continuous transition
  - probability distribution of firing delay ⇒ interval of firing speed
- Then we derive differential equation from the HPN model.
- Finally, we compute an approximated state space by symbolic computation.

Continuous Approximation

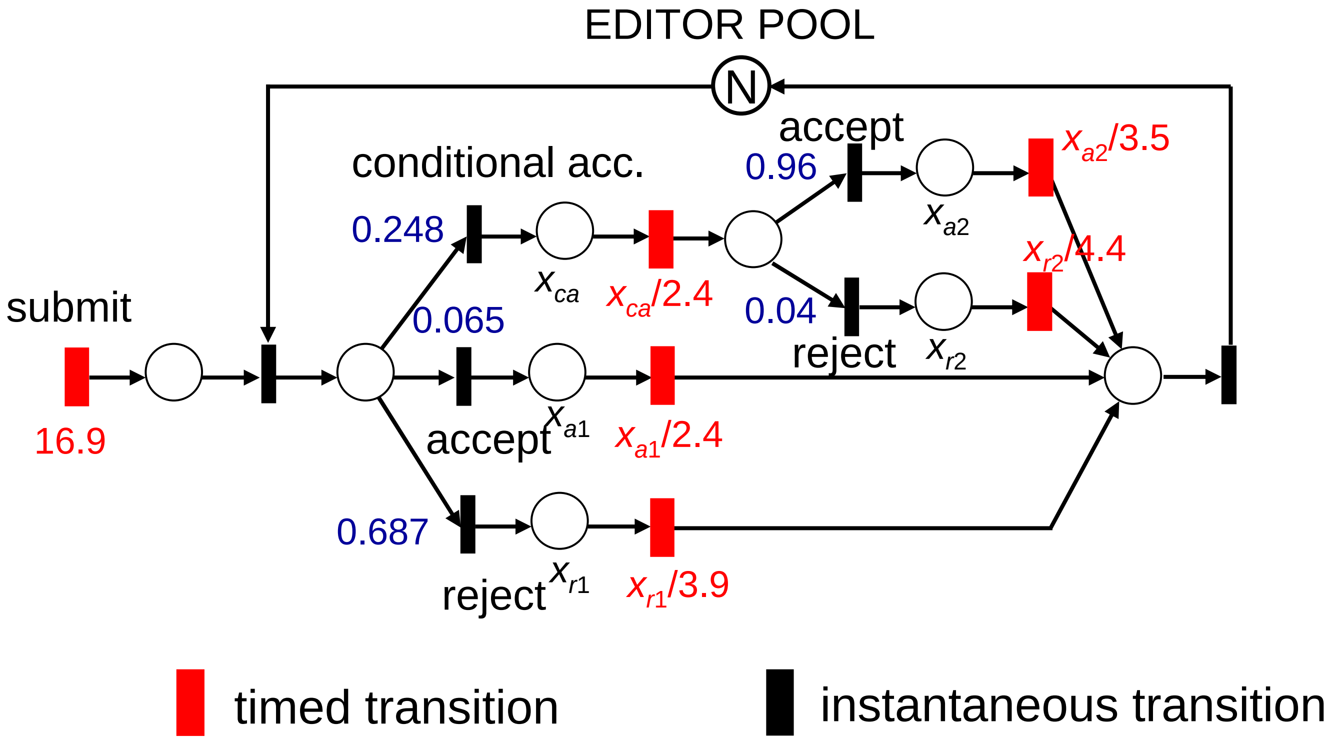


Firing Delay ⇒ Firing Speed



Linear differential equation $\dot{y} = \frac{1}{\tau_1} x - \frac{1}{\tau_2} \quad (\tau_1 \in [L_1, U_1],\ \tau_2 \in [L_2, U_2],\ y \geq 0)$

Extended Continuous Petri Net Model



Computation of the State Space

1. Differential equations (continuous time) $\Rightarrow$ *Difference equations* (discrete time).
2. *Rectangular approximation* of reachable regions at each step. We use *place invariants* for avoiding divergence of intervals.
3. *Symbolic computation* by KCLP-HS (a rapid prototyping tool for algorithms on hybrid systems).

1-Step State Transition

$$\underline{x}_p \le x_p \le \overline{x}_p, \underline{x}_s \le x_s \le \overline{x}_s, \underline{x}_{a1} \le x_{a1} \le \overline{x}_{a1}, \underline{x}_{r1} \le x_{r1} \le \overline{x}_{r1},$$
$$\underline{x}_{ca} \le x_{ca} \le \overline{x}_{ca}, \underline{x}_{a2} \le x_{a2} \le \overline{x}_{a2}, \underline{x}_{r2} \le x_{r2} \le \overline{x}_{r2},$$

Approximation by Rectangular Sets

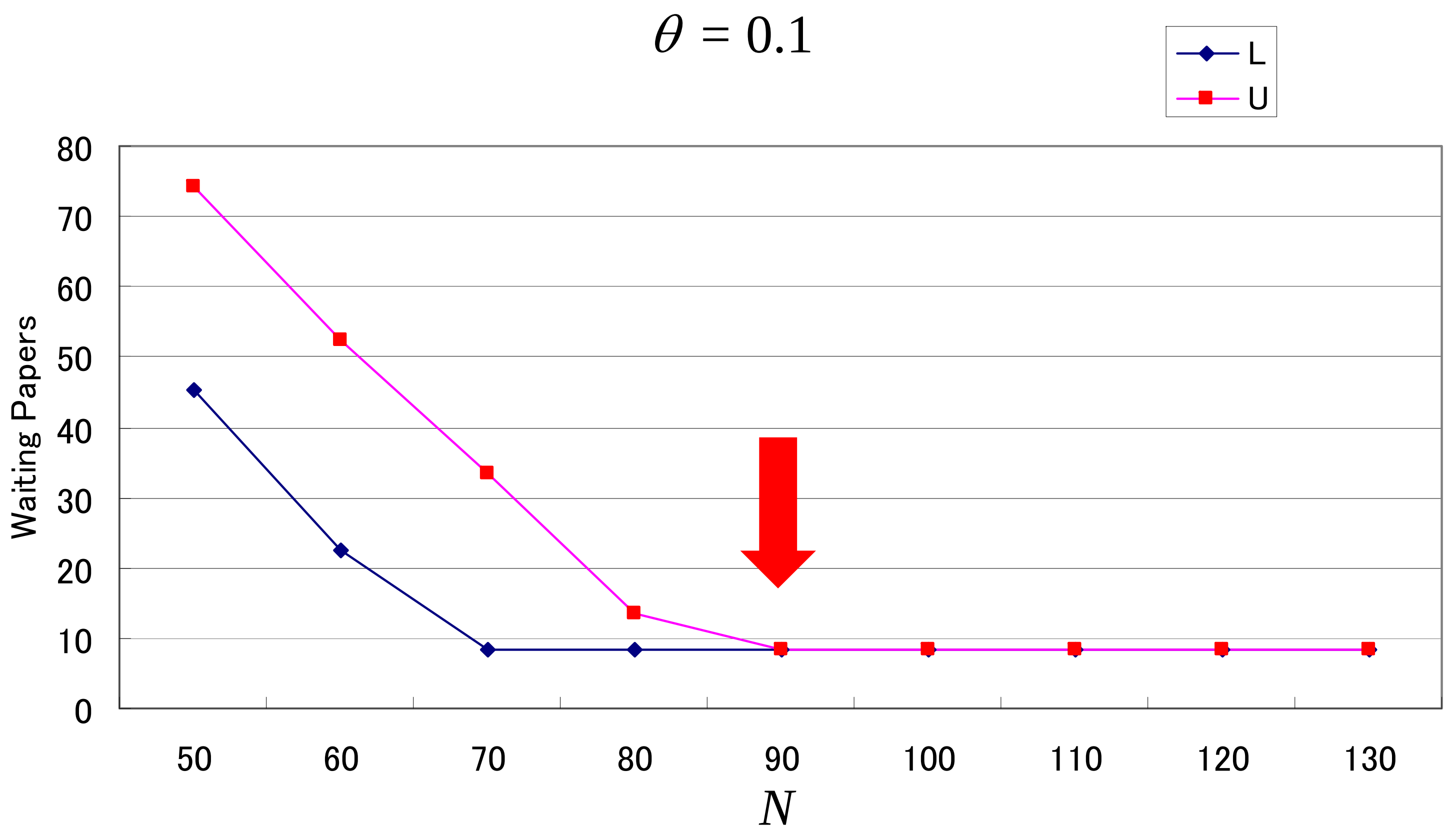
if $x_p \ge x_s$, **then** $x_{tmp} := x_p - x_s$, $in := x_s$, $x_s' := r_s$;
else $x_{tmp} := 0$, $in := x_p$, $x_s' := x_s - x_p + r_s$;
if $x_{a1} \ge r_{a1}$, **then** $x_{a1}' := x_{a1} - r_{a1} + in \cdot p_{a1}$, $rel_0 := r_{a1}$;
else $x_{a1}' := in \cdot p_{a1}$, $rel_0 := x_{a1}$;
if $x_{r1} \ge r_{r1}$, **then** $x_{r1}' := x_{r1} - r_{r1} + in \cdot p_{r1}$, $rel_1 := rel_0 + r_{r1}$;
else $x_{r1}' := in \cdot p_{r1}$, $rel_1 := rel_0 + x_{r1}$;
if $x_{ca} \ge r_{ca}$, **then** $x_{ca}' := x_{ca} - r_{ca} + in \cdot p_{ca}$, $in_2 := r_{ca}$;
else $x_{ca}' := in \cdot p_{ca}$, $in_2 := x_{ca}$;
if $x_{a2} \ge r_{a2}$, **then** $x_{a2}' := x_{a2} - r_{a2} + in_2 \cdot p_{a2}$, $rel_2 := rel_1 + r_{a1}$;
else $x_{a2}' := in_2 \cdot p_{a2}$, $rel_2 := rel_1 + x_{a2}$;
if $x_{r2} \ge r_{r2}$, **then** $x_{r2}' := x_{r2} - r_{r2} + in_2 \cdot p_{r2}$, $rel_3 := rel_2 + r_{r2}$;
else $x_{r2}' := in_2 \cdot p_{r2}$, $rel_3 := rel_2 + x_{r2}$;
$x_p' := x_{tmp} - rel_3$;

➡ Compute Min and Max of each x_i'

$$x_p + x_{a1} + x_{r1} + x_{ca} + x_{a2} + x_{r2} = N$$

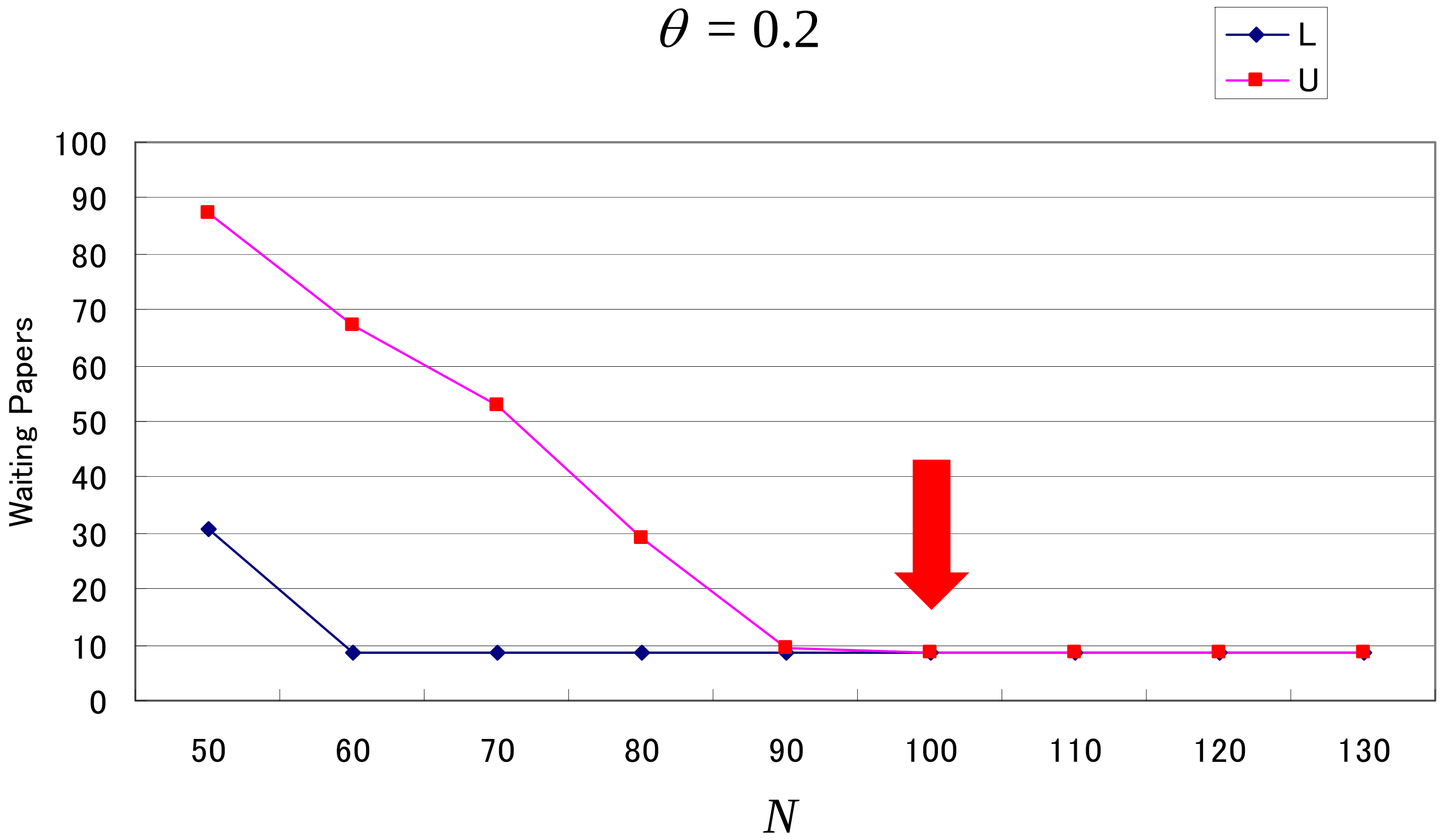
Place-invariant

Result



The number of waiting papers at 6 month later (sampling interval = 0.5 month).
Firing speeds may change ±10%. CPU time < 0.1 sec. for each P.

Result



The number of waiting papers at 6 month later (sampling interval = 0.5 month).
Firing speeds may change ±20%. CPU time < 0.1 sec. for each P.

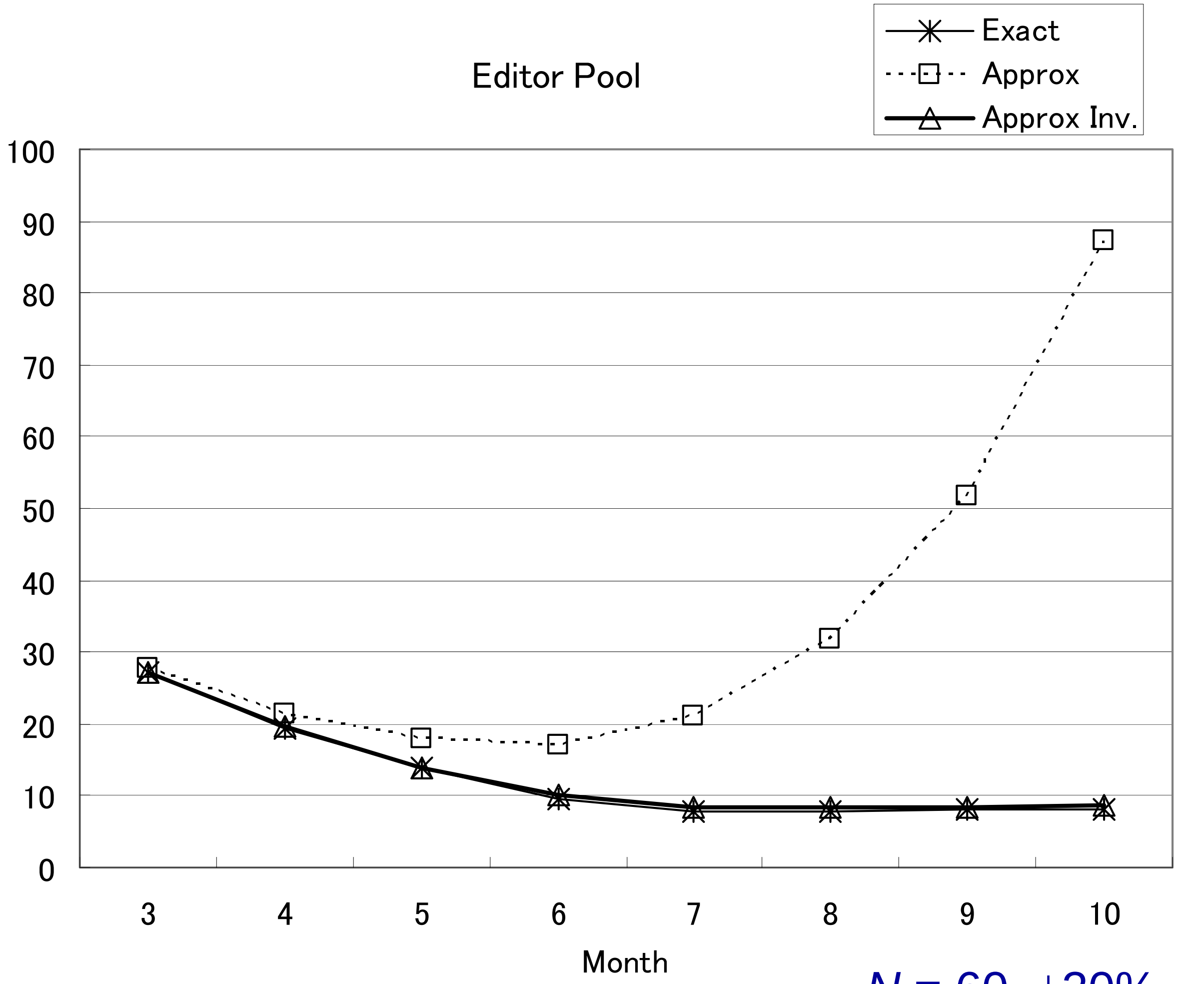
Comparison

CPU Time (sec.)

Duration	(1) Exact	(2) Approx	(3) Approx with inv.
3	0.02	0.01	0.02
4	0.09	0.02	0.03
5	0.27	0.03	0.03
6	3.3	0.04	0.05
7	17	0.05	0.07
8	48	0.06	0.08
9	106	0.08	0.09
10	212	0.09	0.10

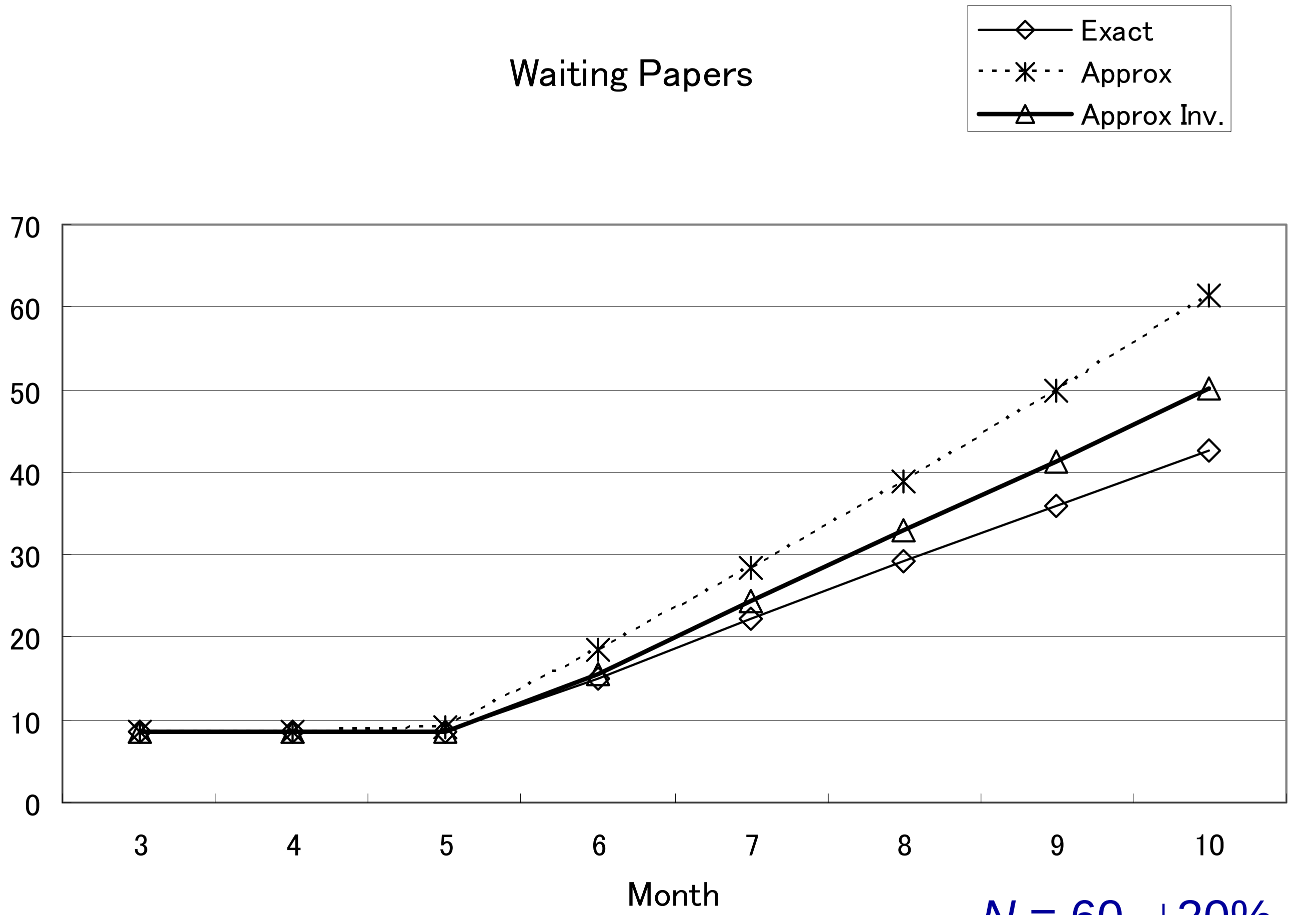
$N = 60$, ±20%.

Comparison



N = 60, ±20%.

Comparison



N = 60, ±20%.

Conclusion

- For performance evaluation of workflows, we have tried to methods, GSPN and continuous approximation by hybrid systems.
- The later method derives a similar result in a much shorter time.
- We expect that the continuous approximation by hybrid systems is applicable to larger workflows for which GSPN is infeasible to compute the solution.