Analyzing an information spread for a gossip-based protocol

Rena Bakhshi

Theoretical Computer Science Section
Department of Computer Science
Vrije Universiteit Amsterdam

RWTH Aachen,
November 20, 2009

Acknowlegments

Talk is based on:

R. Bakhshi, D. Gavidia, W. Fokkink, and M. van Steen.
An analytical model of information dissemination for a gossip-based protocol.
Computer Networks, 53(13):2288–2303, 2009.

R. Bakhshi, and W. Fokkink.
How Probable is it to Discard an Ace of Hearts?
In *Liber Amicorum for Roel de Vrijer* , pages 1–9, 2009.

R. Bakhshi, and A. Fehnker.
On the impact of modelling choices for distributed information spread.
In *Proc. of QEST*, pages 41–50. IEEE Computer Society, 2009.

The Shuffle Protocol

- Large-scale networks
- Information dissemination
- Gossip-based protocol
- Simple push-pull protocol

How the shuffle protocol works?

Each node

- holds list of data items
- communicates with random peer
- exchanges random subset of items

The Shuffle Protocol

The algorithm

wait (some t time units)
$B :=$ selectPeer();
$s_A :=$ itemsToSend(c_A);
sendTo(s_A, B);
$s_B :=$ receive($\cdot$);
$c_A :=$ itemsToKeep(c_A, s_B);

(a) initiating node

$s_A :=$ receive($\cdot$);
$s_B :=$ itemsToSend(c_B);
sendTo(s_B, A);
$c_B :=$ itemsToKeep(c_B, s_A);

(b) contacted node

Example

c_A
a
b
c
d
e

The Shuffle Protocol

The algorithm

wait (some t time units)
$B :=$ selectPeer();
$s_A :=$ itemsToSend(c_A);
sendTo(s_A, B);
$s_B :=$ receive($\cdot$);
$c_A :=$ itemsToKeep(c_A, s_B);

(a) initiating node

$s_A :=$ receive($\cdot$);
$s_B :=$ itemsToSend(c_B);
sendTo(s_B, A);
$c_B :=$ itemsToKeep(c_B, s_A);

(b) contacted node

Example

c_A	c_B
a	a
b	g
c	f
d	e
e	h

The Shuffle Protocol

The algorithm

wait (some t time units)
$B := \text{selectPeer}();$
$s_A := \text{itemsToSend}(c_A);$
$\text{sendTo}(s_A, B);$
$s_B := \text{receive}(\cdot);$
$c_A := \text{itemsToKeep}(c_A, s_B);$

(a) initiating node

$s_A := \text{receive}(\cdot);$
$s_B := \text{itemsToSend}(c_B);$
$\text{sendTo}(s_B, A);$
$c_B := \text{itemsToKeep}(c_B, s_A);$

(b) contacted node

Example

c_A	s_A	c_B
a	c	a
b	d	g
c	a	f
d		e
e		h

The Shuffle Protocol

The algorithm

wait (some t time units)
B := selectPeer();
s_A := itemsToSend(c_A);
sendTo(s_A, B);
s_B := receive($\cdot$);
c_A := itemsToKeep(c_A, s_B);

(a) initiating node

s_A := receive($\cdot$);
s_B := itemsToSend(c_B);
sendTo(s_B, A);
c_B := itemsToKeep(c_B, s_A);

(b) contacted node

Example

c_A	$\rightarrow$	c_B
a	s_A	a
b	c	g
c	d	f
d	a	e
e		h

The Shuffle Protocol

The algorithm

wait (some t time units)
$B := \text{selectPeer}();$
$s_A := \text{itemsToSend}(c_A);$
$\text{sendTo}(s_A, B);$
$s_B := \text{receive}(\cdot);$
$c_A := \text{itemsToKeep}(c_A, s_B);$

(a) initiating node

$s_A := \text{receive}(\cdot);$
$s_B := \text{itemsToSend}(c_B);$
$\text{sendTo}(s_B, A);$
$c_B := \text{itemsToKeep}(c_B, s_A);$

(b) contacted node

Example

c_A	$\rightarrow$	c_B
a	s_A	a
b	c	g
c	d	f
d	a	e
e		h

The Shuffle Protocol

The algorithm

wait (some t time units)
B := selectPeer();
s_A := itemsToSend(c_A);
sendTo(s_A, B);
s_B := receive($\cdot$);
c_A := itemsToKeep(c_A, s_B);

(a) initiating node

s_A := receive($\cdot$);
s_B := itemsToSend(c_B);
sendTo(s_B, A);
c_B := itemsToKeep(c_B, s_A);

(b) contacted node

Example

c_A	$\rightarrow$		c_B
a	s_A	s_B	a
b	c	f	g
c	d	e	f
d	a	a	e
e			h

The Shuffle Protocol

The algorithm

(a) initiating node	(b) contacted node
wait (some t time units)	s_A := receive$(\cdot)$;
B := selectPeer();	s_B := itemsToSend(c_B);
s_A := itemsToSend(c_A);	sendTo(s_B, A);
sendTo(s_A, B);	c_B := itemsToKeep(c_B, s_A);
s_B := receive$(\cdot)$;	
c_A := itemsToKeep(c_A, s_B);	

Example

c_A	$\rightarrow$	$\leftarrow$	c_B
a	s_A	s_B	a
b	c	f	g
c	d	e	f
d	a	a	e
e			h

The Shuffle Protocol

The algorithm

wait (some t time units)
B := selectPeer();
s_A := itemsToSend(c_A);
sendTo(s_A, B);
s_B := receive($\cdot$);
c_A := itemsToKeep(c_A, s_B);

(a) initiating node

s_A := receive($\cdot$);
s_B := itemsToSend(c_B);
sendTo(s_B, A);
c_B := itemsToKeep(c_B, s_A);

(b) contacted node

Example

c_A	$\leftarrow$	$\rightarrow$	c_B
a	s_B	s_A	a
b	f	c	g
c	e	d	f
d	a	a	e
e			h

The Shuffle Protocol

The algorithm

wait (some t time units)
B := selectPeer();
s_A := itemsToSend(c_A);
sendTo(s_A, B);
s_B := receive(·);
c_A := itemsToKeep(c_A, s_B);

(a) initiating node

s_A := receive(·);
s_B := itemsToSend(c_B);
sendTo(s_B, A);
c_B := itemsToKeep(c_B, s_A);

(b) contacted node

Example

c_A	s_B	s_A	c_B
a			a
b	f	c	g
c	e	d	f
d	a	a	e
e			h

The Shuffle Protocol

The algorithm

wait (some t time units)
$B := \text{selectPeer}()$;
$s_A := \text{itemsToSend}(c_A)$;
$\text{sendTo}(s_A, B)$;
$s_B := \text{receive}(\cdot)$;
$c_A := \text{itemsToKeep}(c_A, s_B)$;

(a) initiating node

$s_A := \text{receive}(\cdot)$;
$s_B := \text{itemsToSend}(c_B)$;
$\text{sendTo}(s_B, A)$;
$c_B := \text{itemsToKeep}(c_B, s_A)$;

(b) contacted node

Example

c_A	s_B	s_A	c_B
a			a
b	f	c	g
c	e	d	f
d	a	a	e
e			h

The Shuffle Protocol

The algorithm

wait (some t time units)
B := selectPeer();
s_A := itemsToSend(c_A);
sendTo(s_A, B);
s_B := receive($\cdot$);
c_A := itemsToKeep(c_A, s_B);

(a) initiating node

s_A := receive($\cdot$);
s_B := itemsToSend(c_B);
sendTo(s_B, A);
c_B := itemsToKeep(c_B, s_A);

(b) contacted node

Example

c_A	c_B
a	a
b	g
f	c
d	d
e	h

Protocol Analysis

Assumptions

- No failures
- Uniform distribution
- Insertion of items at once (validated experimentally)
- Shuffle is atomic operation

System parameters

n # different items in the network

c local storage size

s exchange message size

Protocol Analysis

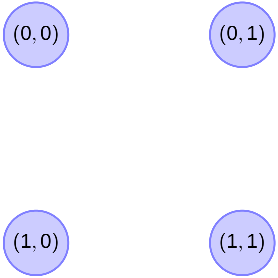
- Analysing the spread of an item d
  - State of a node:

    0 d is not in cache
    1 node posses d
- Pairwise node interaction
  - Pairwise state update
    - $10 \rightarrow 01$
  - Calculating transition probabilities

$$A \underset{s_B}{\overset{s_A}{\rightleftarrows}} B$$

- Modelling protocol properties
  - differential equations (scalable!)
  - model checking using PRISM
  - comparison of different models

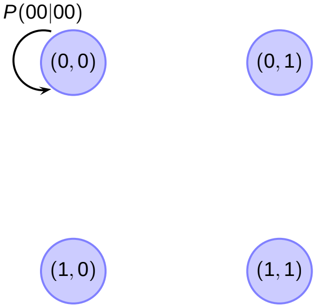
State Transitions



Transition probability matrix

Out-	In-state			
	00	01	10	11
00	1	0	0	0
01	0	$P(01\|01)$	$P(01\|10)$	$P(01\|11)$
10	0	$P(10\|01)$	$P(10\|10)$	$P(10\|11)$
11	0	$P(11\|01)$	$P(11\|10)$	$P(11\|11)$

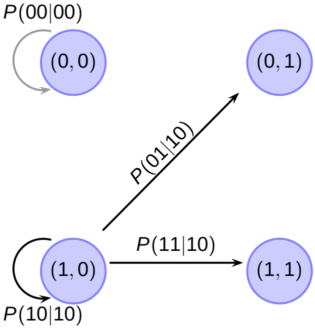
State Transitions



Transition probability matrix

Out-	In-state			
	00	01	10	11
00	1	0	0	0
01	0	$P(01\|01)$	$P(01\|10)$	$P(01\|11)$
10	0	$P(10\|01)$	$P(10\|10)$	$P(10\|11)$
11	0	$P(11\|01)$	$P(11\|10)$	$P(11\|11)$

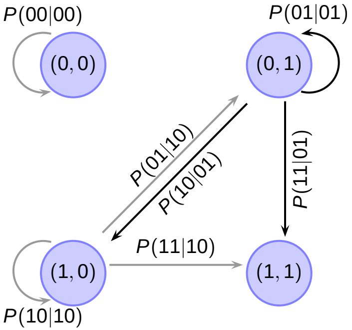
State Transitions



Transition probability matrix

Out-	In-state			
	00	01	10	11
00	1	0	0	0
01	0	$P(01\|01)$	$P(01\|10)$	$P(01\|11)$
10	0	$P(10\|01)$	$P(10\|10)$	$P(10\|11)$
11	0	$P(11\|01)$	$P(11\|10)$	$P(11\|11)$

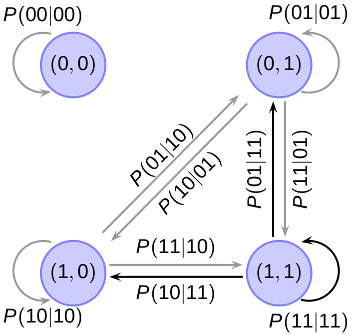
State Transitions



Transition probability matrix

Out-	In-state			
	00	01	10	11
00	1	0	0	0
01	0	$P(01\|01)$	$P(01\|10)$	$P(01\|11)$
10	0	$P(10\|01)$	$P(10\|10)$	$P(10\|11)$
11	0	$P(11\|01)$	$P(11\|10)$	$P(11\|11)$

State Transitions



Transition probability matrix

Out-	In-state			
	00	01	10	11
00	1	0	0	0
01	0	$P(01\|01)$	$P(01\|10)$	$P(01\|11)$
10	0	$P(10\|01)$	$P(10\|10)$	$P(10\|11)$
11	0	$P(11\|01)$	$P(11\|10)$	$P(11\|11)$

Transition Probabilities

- $P(a_2b_2|a_1b_1) = P(b_2a_2|b_1a_1)$ by symmetry;
- 5 transition probabilities to calculate;
- Depend on P_{select} and P_{drop}.

Select/drop an item

- P_{select} : a node selects an item during exchange
- P_{drop} : a node drops (selected) item after exchange

Example

- $P(10|10) = 1 - P_{select}$
- $P(01|10) = P_{select} \cdot P_{drop}$

- $P_{select} = \dfrac{s}{c}$
- $P_{drop} = \dfrac{n-c}{n-s} \cdot \left(1 - \dfrac{1}{\binom{n}{s}}\right) \approx \dfrac{n-c}{n-s}$

Protocol Properties

Generic item d inserted in a network of gossiping nodes.

Properties

Replication # copies of d at round r?

Coverage # nodes seen d after r rounds?

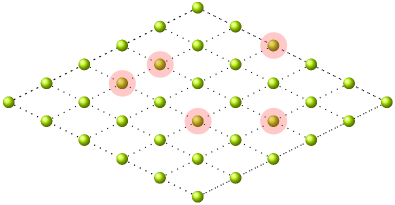


Figure: Replication

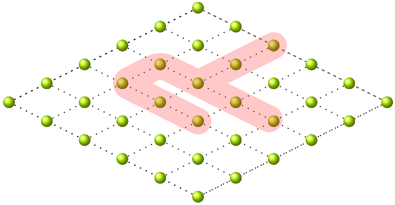


Figure: Coverage

Experimental validation

- event-based simulator PeerSim
- 2500 nodes, grid topology

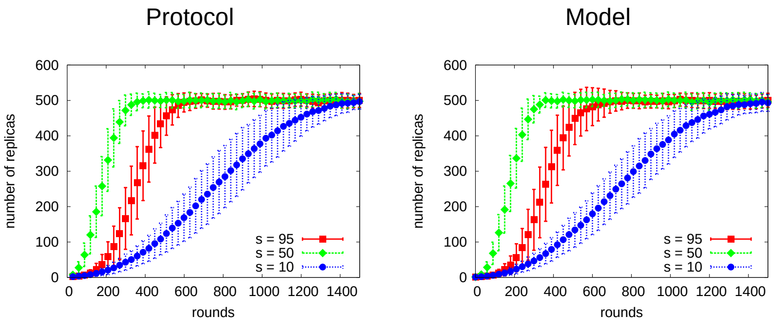


Figure: Replication results for $n = 500$, $c = 100$

Experimental validation

- event-based simulator PeerSim
- 2500 nodes, grid topology

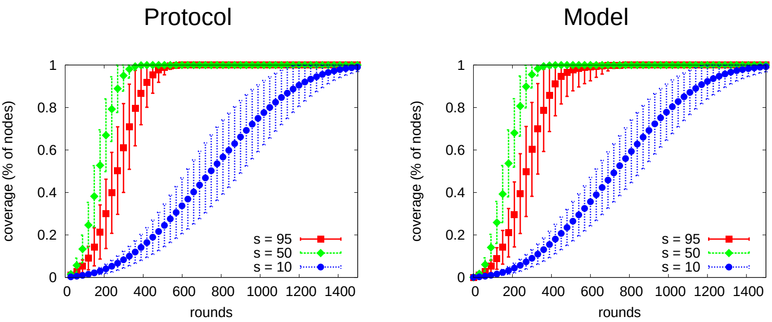


Figure: Coverage results for $n = 500$, $c = 100$

Model checking

Observations:

- order in which nodes initiate gossip is unspecified
- different notion of rounds

Setup:

- 10 nodes
- $c = 100$, $n = 500$, $s = 50$

Basic PRISM model:

- module per node
- implements state transition diagram

ST diagram

PRISM models:

- Non-deterministic: all possible schedulers
- Probabilistic: a fair scheduler
- Deterministsic: round-robin scheduler

Model checking

Observations:

- order in which nodes initiate gossip is unspecified
- different notion of rounds

Setup:

- 10 nodes
- $c = 100$, $n = 500$, $s = 50$

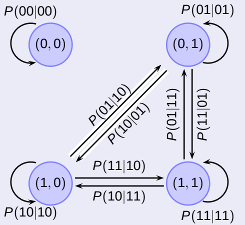
Basic PRISM model:

- module per node
- implements state transition diagram

PRISM models:

- Non-deterministic: all possible schedulers
- Probabilistic: a fair scheduler
- Deterministsic: round-robin scheduler

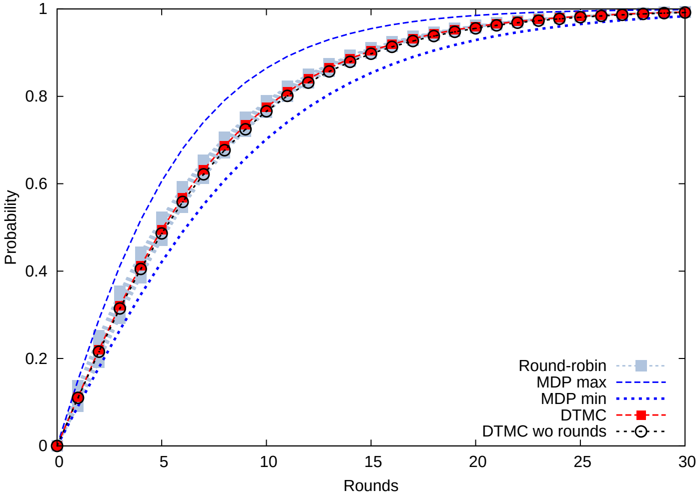
ST diagram



PRISM models

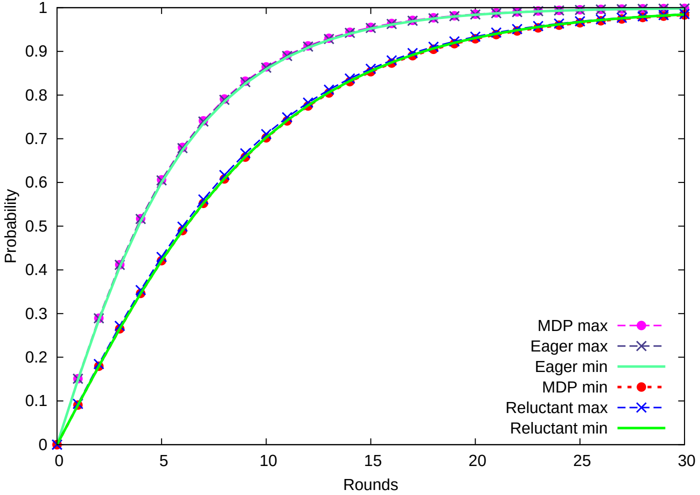
Coverage property in pCTL

```
P=?  [(true) U<=k*(N+1) (ni)]
```



PRISM models

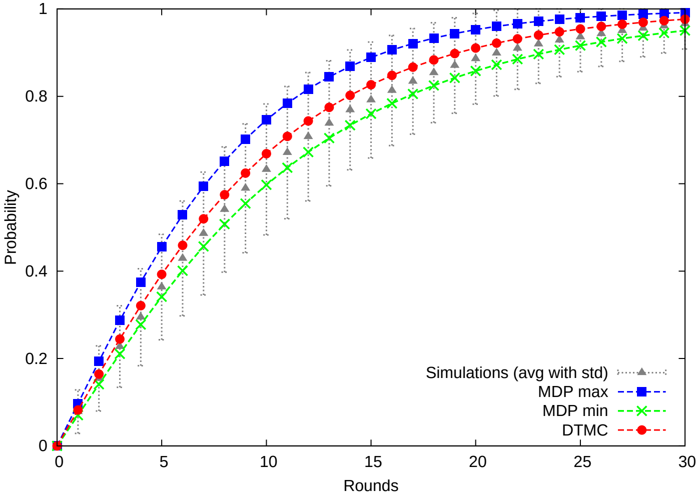
- MDP is a combination of several schedulers
- “unfair” schedulers: optimistic and pessimistic



PRISM models

Coverage property in pCTL

```
P=?  [(true) U<=k*(N+1) (ni & sent)]
```



Properties modelling

Assumptions:

- Full connectivity
- Node initiates gossip once per round

Let

- x – fraction of nodes possesing item d

Replication

$$\frac{dx}{dt} = [P(11|10) + P(11|01)] \cdot (1 - x) \cdot x - [P(10|11) + P(01|11)] \cdot x \cdot x$$

Properties modelling

Assumptions:

- Full connectivity
- Node initiates gossip once per round

Let

- x – fraction of nodes possesing item *d*

Replication

$$\frac{dx}{dt} = \underbrace{[P(11|10) + P(11|01)] \cdot (1 - x) \cdot x}_{\text{duplication of } d} - [P(10|11) + P(01|11)] \cdot x \cdot x$$

Properties modelling

Assumptions:

- Full connectivity
- Node initiates gossip once per round

Let

- x – fraction of nodes possesing item d

Replication

$$\frac{dx}{dt} = \underbrace{[P(11|10) + P(11|01)] \cdot (1 - x) \cdot x}_{\text{duplication of } d} - \underbrace{[P(10|11) + P(01|11)] \cdot x \cdot x}_{\text{dropping } d}$$

Properties modelling

Assumptions:

- Full connectivity
- Node initiates gossip once per round

Let

- x – fraction of nodes possesing item d
- y – fraction of nodes have seen d

Coverage

$$\frac{dy}{dt} = (1-y) \cdot \sum_{i=0}^{k} C(i) \cdot \Phi_{i+1},$$

where

$$\Phi_i = \sum_{m=0}^{i-1} (1-\Phi_m) \cdot x \cdot P(1*|01) \cdot (x \cdot P(1*|11) + (1-x) \cdot P(1*|10))^{(i-m)-1}$$

Properties modelling

Assumptions:

- Full connectivity
- Node initiates gossip once per round

Let

- x – fraction of nodes possesing item d
- y – fraction of nodes have seen d

Coverage

$$\frac{dy}{dt} = \underbrace{(1-y)}_{\text{not seen } d} \cdot \sum_{i=0}^{k} C(i) \cdot \Phi_{i+1},$$

where

$$\Phi_i = \sum_{m=0}^{i-1} (1-\Phi_m) \cdot x \cdot P(1*|01) \cdot (x \cdot P(1*|11) + (1-x) \cdot P(1*|10))^{(i-m)-1}$$

Properties modelling

Assumptions:

- Full connectivity
- Node initiates gossip once per round

Let

- x – fraction of nodes possesing item d
- y – fraction of nodes have seen d

Coverage

$$\frac{dy}{dt} = \underbrace{(1-y)}_{\text{not seen } d} \cdot \sum_{i=0}^{k} \underbrace{C(i) \cdot \Phi_{i+1},}_{\text{shuffle } i \text{ times and get } d}$$

where

$$\Phi_i = \sum_{m=0}^{i-1} (1-\Phi_m) \cdot x \cdot P(1*|01) \cdot (x \cdot P(1*|11) + (1-x) \cdot P(1*|10))^{(i-m)-1}$$

Properties modelling

Assumptions:

- Full connectivity
- Node initiates gossip once per round

Let

- x – fraction of nodes possesing item d
- y – fraction of nodes have seen d

Coverage

$$\frac{dy}{dt} = (1-y) \cdot \sum_{i=0}^{k} C(i) \cdot \Phi_{i+1},$$

where

$$\Phi_i = \sum_{m=0}^{i-1} (1-\Phi_m) \cdot \underbrace{x \cdot P(1*|01)}_{\text{obtain } d} \cdot (x \cdot P(1*|11) + (1-x) \cdot P(1*|10))^{(i-m)-1}$$

Properties modelling

Assumptions:

- Full connectivity
- Node initiates gossip once per round

Let

- x – fraction of nodes possesing item d
- y – fraction of nodes have seen d

Coverage

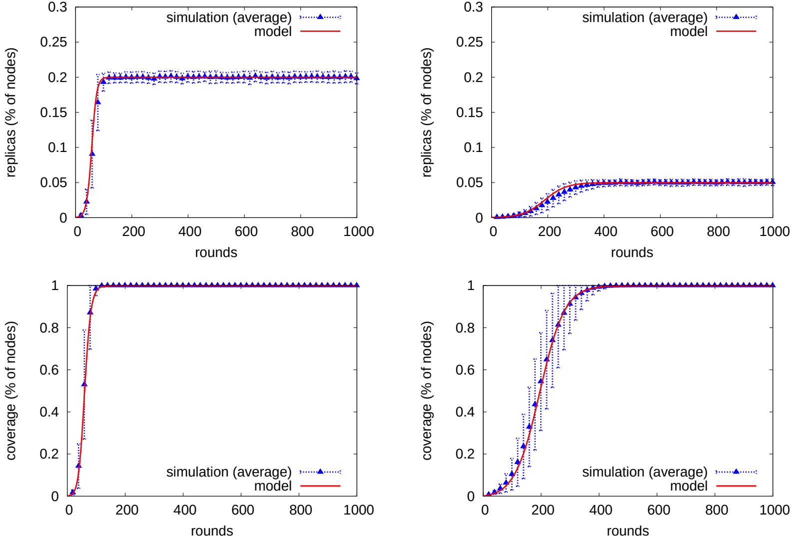
$$\frac{dy}{dt} = (1-y) \cdot \sum_{i=0}^{k} C(i) \cdot \Phi_{i+1},$$

where

$$\Phi_i = \sum_{m=0}^{i-1} (1-\Phi_m) \cdot \underbrace{x \cdot P(1*|01)}_{\text{obtain } d} \cdot \underbrace{(x \cdot P(1*|11) + (1-x) \cdot P(1*|10))}_{\text{keep } d}{}^{(i-m)-1}$$

Validation results

$N = 2500$, $c = 100$, $s = 50$, 100 independent runs



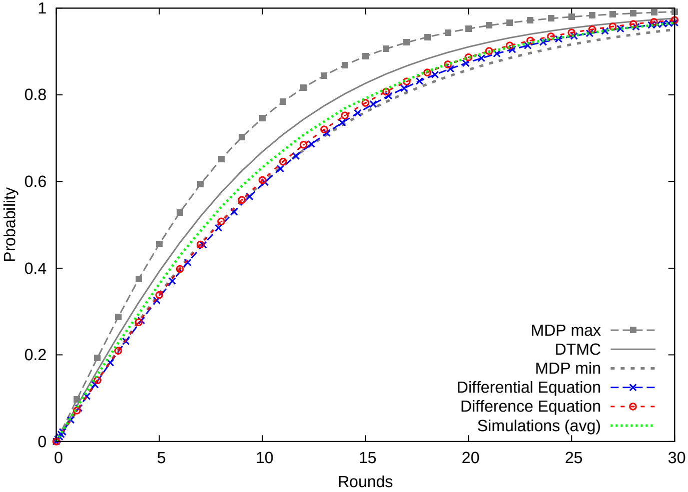
500 items

2000 items

Equational vs PRISM models

Difference equation for coverage

$$y(t+1) = y(t) + (1 - y(t)) \cdot \sum_{i=0}^{k} C(i) \cdot \Phi(i+1)$$



Equational vs PRISM models

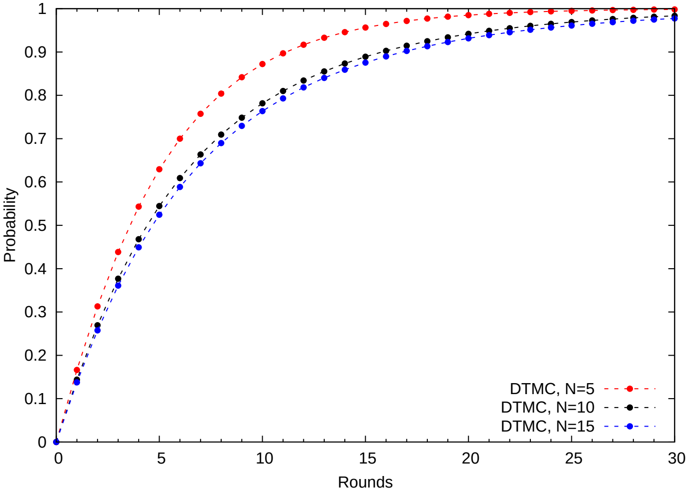


Figure: DTMCs and DEqns for different N

Equational vs PRISM models

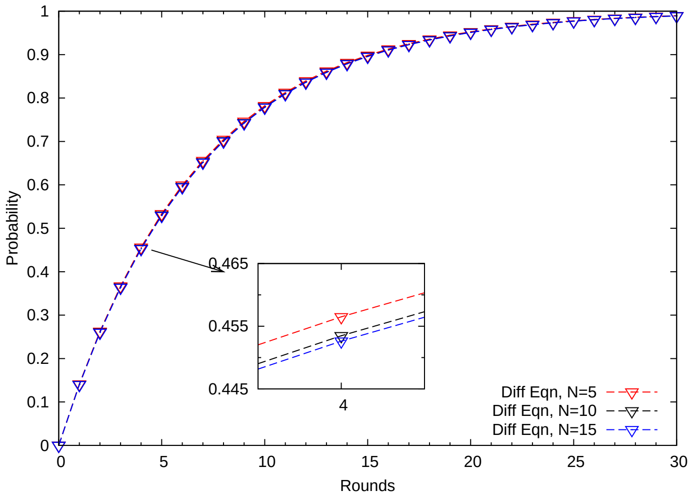


Figure: DTMCs and DEqns for different N

And finally!

Thank you for your attention!