



Specification-Guided Path Shortcutting for Efficient Probabilistic Model Checking

Tsubasa Matsumoto¹ Kazuki Watanabe² Masaki Waga^{1,2}

¹Kyoto University ²National Institute of Informatics

EMSOFT 2026, 2026 Oct. 5th

Stochastic Safety-Critical Systems

Waymo robotaxi accident with San Francisco cyclist draws regulatory review

By Reuters

February 9, 2024 5:35 AM GMT+9 · Updated February 9, 2024



A Waymo rider-only robotaxi is seen during a test ride in San Francisco, California, U.S., December 9, 2022. REUTERS/Patrick Dave/File Photo Purchase Licensing Rights

Boeing to plead guilty to fraud in US probe of fatal 737 MAX crashes

By Chris Prentice, Mike Spector and David Shepardson

July 9, 2024 12:19 AM GMT+9 · Updated July 9, 2024



That's according to a court filing by the U.S. government late Sunday.

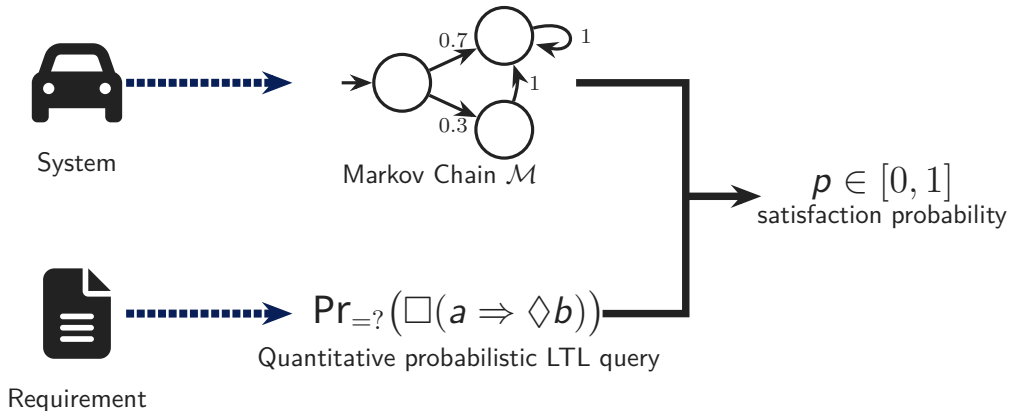
<https://www.reuters.com/world/us/driverless-waymo-car-hits-cyclist-san-francisco-causes-minor-scratches-2024-02-07/>

<https://www.reuters.com/business/aerospace-defense/>

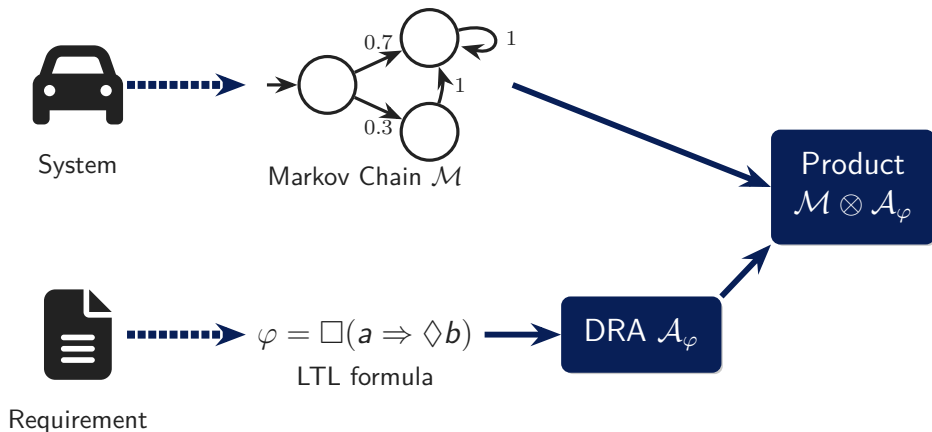
[boeing-plead-guilty-us-probe-fatal-737-max-crashes-says-doj-official-2024-07-08/](https://www.reuters.com/business/aerospace-defense/boeing-plead-guilty-us-probe-fatal-737-max-crashes-says-doj-official-2024-07-08/)

Probabilistic Model Checking

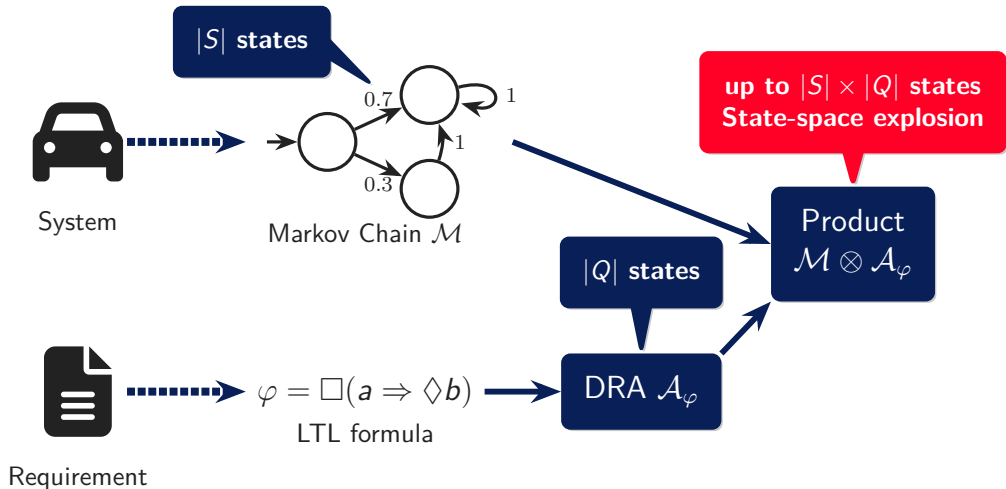
Compute the satisfaction probability of the temporal property by a system



Challenge: State-Space Explosion



Challenge: State-Space Explosion



Specification-Guided Path Shortcutting

[Contribution]

Idea: Shortcut paths equivalent to a shorter path w.r.t. verified spec. φ

Example: $\varphi = \Box(a \Rightarrow \Diamond b)$

Specification-Guided Path Shortcutting

[Contribution]

Idea: Shortcut paths equivalent to a shorter path w.r.t. verified spec. φ

Example: $\varphi = \Box(a \Rightarrow \Diamond b)$

$\{a\}\{b\}$ and $\{b\}$ are equivalent w.r.t. φ

Specification-Guided Path Shortcutting

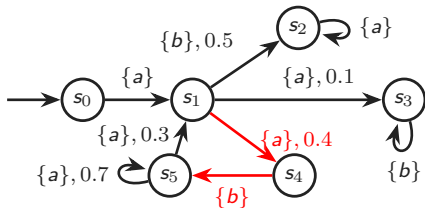
[Contribution]

Idea: Shortcut paths equivalent to a shorter path w.r.t. verified spec. φ

Example: $\varphi = \Box(a \Rightarrow \Diamond b)$

$\{a\}\{b\}$ and $\{b\}$ are equivalent w.r.t. φ

Original MC $\mathcal{M}$



Specification-Guided Path Shortcutting

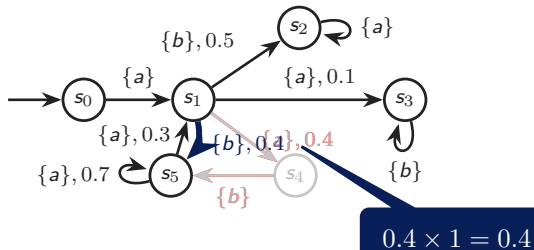
[Contribution]

Idea: Shortcut paths equivalent to a shorter path w.r.t. verified spec. φ

Example: $\varphi = \Box(a \Rightarrow \Diamond b)$

$\{a\}\{b\}$ and $\{b\}$ are equivalent w.r.t. φ

First shortcut



Specification-Guided Path Shortcutting

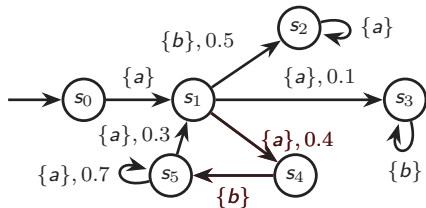
[Contribution]

Idea: Shortcut paths equivalent to a shorter path w.r.t. verified spec. φ

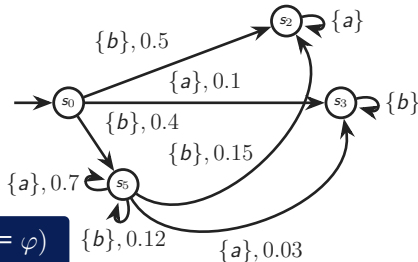
Example: $\varphi = \Box(a \Rightarrow \Diamond b)$

$\{a\}\{b\}$ and $\{b\}$ are equivalent w.r.t. φ

Original MC $\mathcal{M}$



Reduced MC $\mathcal{M}'$



Guarantee: $\mathbb{P}(\mathcal{M} \models \varphi) = \mathbb{P}(\mathcal{M}' \models \varphi)$

Technical Part

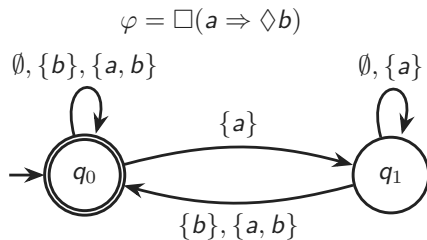
Traces Equivalent w.r.t. Specification φ

Definition (Trace equivalence — informal)

Traces t, c are equivalent w.r.t. $\mathcal{A}$ if for any position q ,

- they end at the same position:
 $\Delta(q, t) = \Delta(q, c)$; and
- shortcutting t to c (and c to t) does not hide visits relevant to Rabin acceptance.

t and c have the same acceptance-relevant effect on $\mathcal{A}$



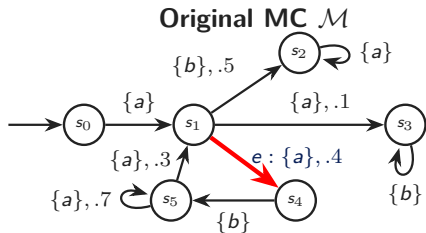
DRA $\mathcal{A}_\varphi$; accept iff q_0 is visited infinitely often.

$$\{a\}\{b\} \sim_{\mathcal{A}_\varphi} \{b\}$$

Evidence of Replaceability

Definition (Replaceable edge — informal)

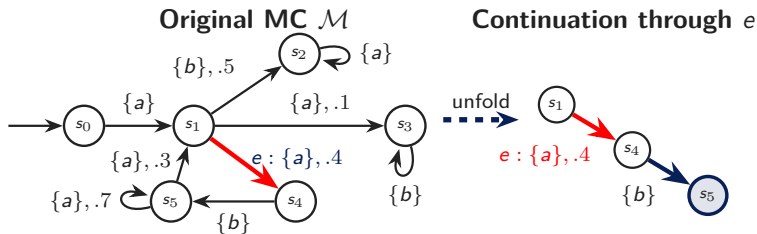
Edge e is **replaceable** if there is a set T of continuations such that each $t \in T$ is equivalent to a single label.



Evidence of Replaceability

Definition (Replaceable edge — informal)

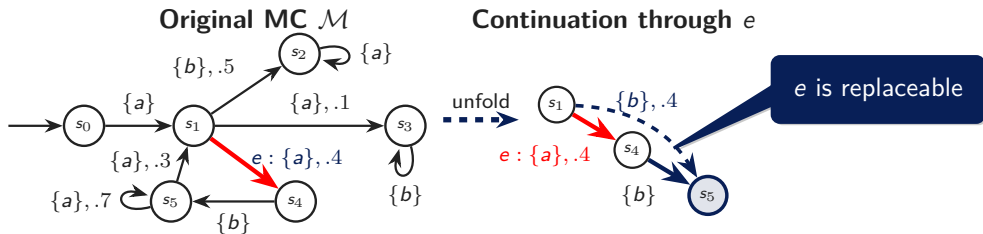
Edge e is **replaceable** if there is a set T of continuations such that each $t \in T$ is equivalent to a single label.



Evidence of Replaceability

Definition (Replaceable edge — informal)

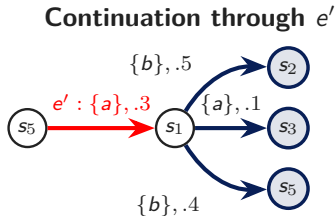
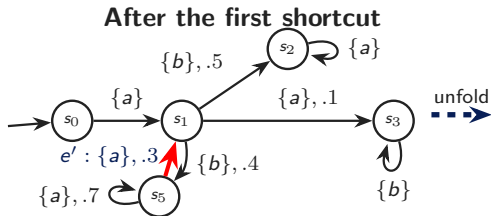
Edge e is **replaceable** if there is a set T of continuations such that each $t \in T$ is equivalent to a single label.



Evidence of Replaceability

Definition (Replaceable edge — informal)

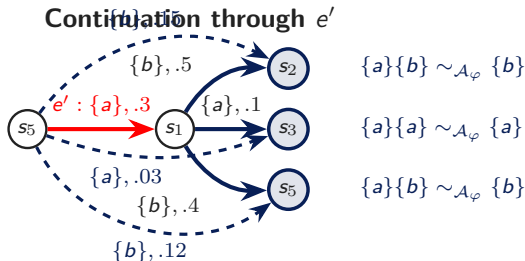
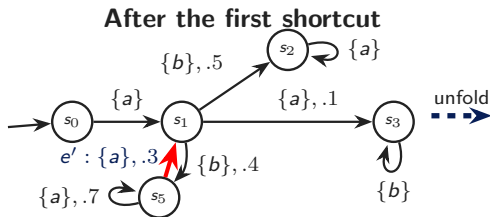
Edge e is **replaceable** if there is a set T of continuations such that each $t \in T$ is equivalent to a single label.



Evidence of Replaceability

Definition (Replaceable edge — informal)

Edge e is **replaceable** if there is a set T of continuations such that each $t \in T$ is equivalent to a single label.



Implementation: search continuations only up to depth K

Theoretical Guarantees

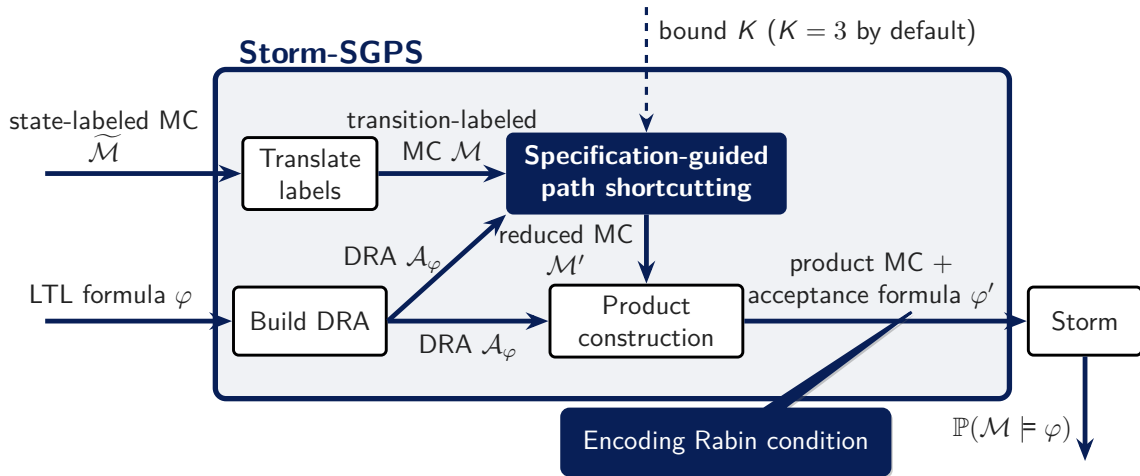
Theorem (Edge replacement)

Replacing a replaceable edge e in $\mathcal{M}$ w.r.t. $\mathcal{A}$ with shortcutting edges does not change the satisfaction probability of $\mathcal{A}$

Corollary (State elimination)

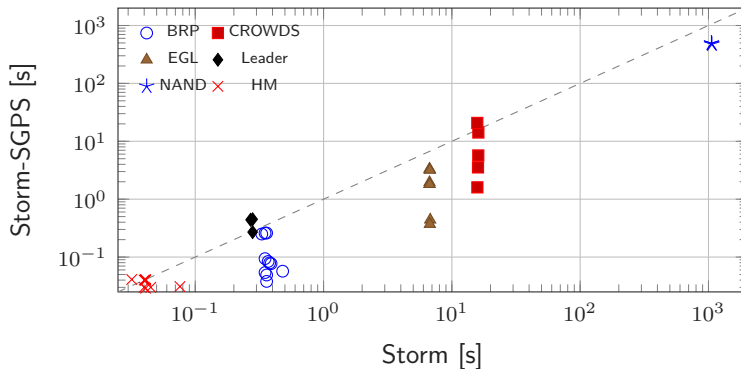
Erasing an unreachable state after edge replacement preserves the satisfaction probability of $\mathcal{A}$.

Implementation



Comparison with Storm

Storm-SGPS is faster in most cases



Largest speedup

EGL φ_7 : 6.74 $\rightarrow$ 0.37 s.

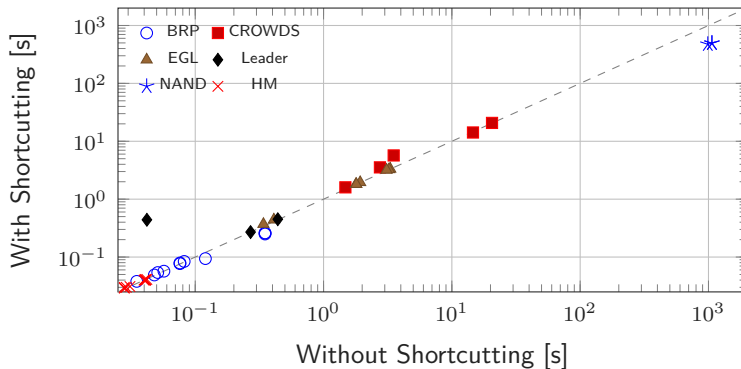
Largest model

NAND φ_1 : 1055 $\rightarrow$ 494 s.
18.8M $\rightarrow$ 7.9M states

Not always faster

CROWDS φ_4 : no state reduction and a harder acceptance encoding.

Ablation: Contribution of Path Shortcutting



Observation

Usually close to diagonal:
Product construction and
label reduction often reduce
time.

But

Substantial state space
reduction can significantly
reduce the execution time.

Ablation: Contribution of Path Shortcutting

Performance depends on the size of the resulting MC

NAND φ_1

- 1064.97 $\rightarrow$ 493.50 s
- 18.8M $\rightarrow$ 7.9M states

Shortcutting clearly improves the performance.

BRP φ_3

- 0.35 $\rightarrow$ 0.25 s
- 5192 $\rightarrow$ 3469 states

Moderate reduction, moderate gain.

Leader φ_1

- 0.042 $\rightarrow$ 0.44 s
- Product transitions: 5268 $\rightarrow$ 132103

Removing states can increase edges.

Conclusions

- Specification-guided notion of replaceable paths and edges
- Sound state elimination by path shortcutting
- Exact preservation of the satisfaction probability
- Prototype implementation and evaluation with Storm
 - Usually reduces end-to-end runtime, especially with substantial state-space reduction
- Property-focused reduced models can also help understanding and debugging

Appendix

Trace Compatibility

[Contribution]

Definition (Acceptance-preserving word)

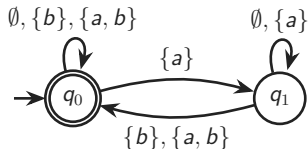
Let $\mathcal{A}$ be a DRA with Rabin condition $(L_i, U_i)_{i \in I}$. A non-empty word t is **acceptance-preserving** if, for every $q \in Q$, every non-empty prefix t' of t , and every $X_i \in \{L_i, U_i\}$,

$$\Delta(q, t') \in X_i \Rightarrow \Delta(q, t) \in X_i.$$

Definition (Compatibility)

For acceptance-preserving t_1, t_2 , we define the compatibility relation $\sim_{\mathcal{A}}$ as follows:

$$t_1 \sim_{\mathcal{A}} t_2 \iff \Delta(q, t_1) = \Delta(q, t_2) \text{ for every } q \in Q.$$



We focus on acceptance-preserving words to prevent hiding a visit relevant to the Rabin acceptance condition.

$$\{a\}\{b\} \sim_{\mathcal{A}_\varphi} \{b\}$$

$$\{a\}\{a\} \sim_{\mathcal{A}_\varphi} \{a\}$$

Replaceable Edge and MCCS

Definition (Replaceable edge)

Let $e = (s, c, s')$ be an edge of $\mathcal{M}$. Let P_e and P_e^ω be the set of finite and infinite paths from e . It is **replaceable** w.r.t. $\mathcal{A}$ if $s \neq s'$ and there exists $T \subseteq P_e$ such that:

1. for every $p^\omega \in P_e^\omega$, some $p \in T$ is a prefix of p^ω ;
2. for every $p \in T$, $O(p)$ is acceptance-preserving and $O(p) \sim_{\mathcal{A}} c'$ for some $c' \in \Sigma$;
3. for every $p \in T$, every strict prefix p' of p with $p' \neq e$, and every $c' \in \Sigma$, we have $O(p') \not\sim_{\mathcal{A}} c'$;
4. $e \notin T$;
5. for every $p \in T$, the target state s' of e appears only once in p .

Minimum compatible covering suffixes

For a replaceable edge, the set T above is unique. We denote it by $MCCS(e, \mathcal{A})$.

Storm-SGPS: Implementation Details

- Prototype implemented in C++ using Spot 2.15.1 [Duret-Lutz et al., 2022]
- Current input format: explicit state-labeled MC in Storm's DRN format
- Internally translates state labels to incoming transition labels
- Exports a state-labeled DRN encoding of $\mathcal{M}' \otimes \mathcal{A}_\varphi$
- Invokes Storm with an LTL formula encoding the DRA acceptance condition

<https://github.com/SoftwareFoundationGroupAtKyotoU/specification-guided-path-shortcutting>

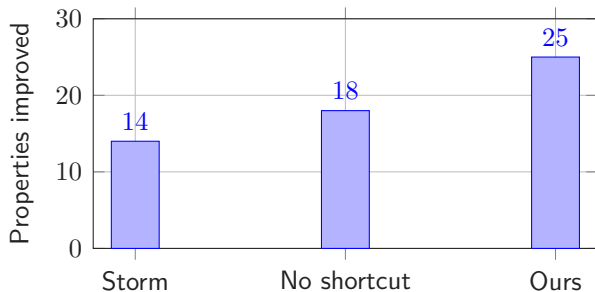
Experimental Setup

- **Baseline:** Storm
1.13.0 [Hensel et al., 2022]
- Main comparison: search-depth
bound $K = 3$
- 30 runs per instance; report
mean runtime
 - End-to-end runtime
preprocessing + Storm
- Ubuntu 24.04; at most 4 CPU
cores / 16 GiB RAM

**6 QComp
benchmarks [Hartmanns et al., 2019]
40 LTL properties, mostly our original**

Benchmark	# states
BRP	5,192
CROWDS	359,622
EGL	115,710
Leader	4,244
NAND	18,826,082
HM	101

Combination with Bisimulation



Complementary

Bisimulation improves Storm-SGPS more often than the other workflows.

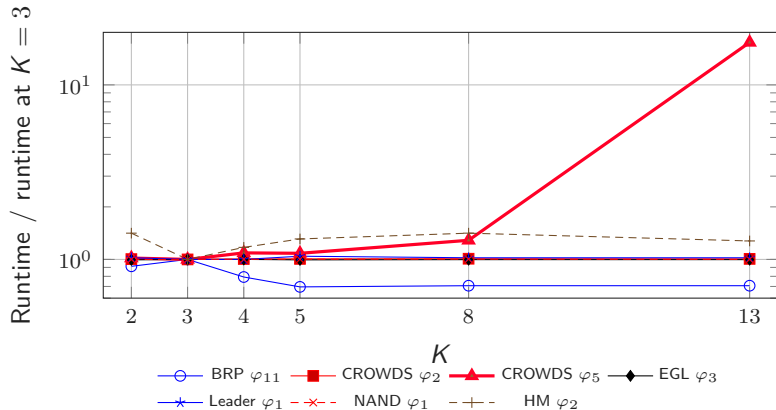
But usually smaller effect

NAND φ_1 :

$$1064.97 \xrightarrow{\text{bisim}} 1058.30 \text{ s}$$

$$1064.97 \xrightarrow{\text{shortcutting}+\text{bisim}} 493.48 \text{ s}$$

Sensitivity to the MCCS Bound K



Result

Reduced states were unchanged for all seven tested properties.

Practical choice

Most runtimes are stable; large K can be expensive.

$K = 2$ or 3 is a reasonable default.

Complete Results (BRP / CROWDS)

Benchmark	Property	Storm [s]	Ours [s]	Reduced states
BRP	φ_1	0.38	0.079	1723
	φ_2	0.35	0.054	1723
	φ_3	0.33	0.25	1723
	φ_4	0.35	0.094	1083
	φ_5	0.35	0.26	1723
	φ_6	0.39	0.077	1723
	φ_7	0.36	0.049	1723
	φ_8	0.36	0.26	1723
	φ_9	0.37	0.084	1723
	φ_{10}	0.36	0.038	1723
	φ_{11}	0.48	0.057	1723
CROWDS	φ_1	16.07	5.69	87360
	φ_2	16.09	14.18	87360
	φ_3	15.99	3.54	30480
	φ_4	15.79	20.75	0
	φ_5	15.81	1.60	0

Complete Results (EGL / Leader)

Benchmark	Property	Storm [s]	Ours [s]	Reduced states
EGL	φ_1	6.70	1.95	1023
	φ_2	6.68	1.81	1023
	φ_3	6.70	3.33	1023
	φ_4	6.71	3.26	1023
	φ_5	6.74	3.19	1023
	φ_6	6.78	0.44	1023
	φ_7	6.74	0.37	1023
	φ_8	6.71	1.81	1023
Leader	φ_1	0.27	0.44	17
	φ_2	0.28	0.27	0
	φ_3	0.28	0.45	0

Complete Results (NAND / HM)

Benchmark	Property	Storm [s]	Ours [s]	Reduced states
NAND	φ_1	1055.41	493.50	10946130
	φ_2	1063.07	473.29	10946130
	φ_3	1055.90	492.91	10946130
HM	φ_1	0.032	0.041	97
	φ_2	0.041	0.029	95
	φ_3	0.041	0.041	97
	φ_4	0.041	0.040	97
	φ_5	0.040	0.030	97
	φ_6	0.040	0.040	97
	φ_7	0.041	0.040	97
	φ_8	0.041	0.040	97
	φ_9	0.045	0.030	97
	φ_{10}	0.076	0.031	97

Ablation / Bisimulation (BRP / CROWDS)

Benchmark	Property	No shortcut	Storm+Bisim	Ours+Bisim	No shortcut+Bisim
BRP	φ_1	.076	.37	.079	.077
	φ_2	.051	.36	.053	.053
	φ_3	.35	.34	.25	.34
	φ_4	.12	.36	.096	.12
	φ_5	.35	.35	.26	.35
	φ_6	.076	.40	.086	.087
	φ_7	.048	.38	.048	.048
	φ_8	.35	.36	.27	.35
	φ_9	.082	.37	.092	.094
	φ_{10}	.035	.37	.036	.036
	φ_{11}	.057	.48	.059	.057
CROWDS	φ_1	3.52	15.61	5.56	3.23
	φ_2	14.61	16.14	14.17	14.44
	φ_3	2.75	16.45	3.46	2.60
	φ_4	20.56	15.77	20.67	20.79
	φ_5	1.48	15.74	1.55	1.43

Ablation / Bisimulation (EGL / Leader)

Benchmark	Property	No shortcut	Storm+Bisim	Ours+Bisim	No shortcut+Bisim
EGL	φ_1	1.93	6.71	1.93	1.92
	φ_2	1.79	6.71	1.79	1.78
	φ_3	3.31	6.69	3.47	3.46
	φ_4	3.24	6.69	3.25	3.22
	φ_5	3.10	6.79	3.11	3.09
	φ_6	.41	6.76	.43	.42
	φ_7	.34	6.92	.35	.34
	φ_8	1.80	6.73	1.79	1.78
Leader	φ_1	.042	.26	.43	.042
	φ_2	.27	.27	.27	.27
	φ_3	.44	.27	.44	.44

Ablation / Bisimulation (NAND / HM)

Benchmark	Property	No shortcut	Storm+Bisim	Ours+Bisim	No shortcut+Bisim
NAND	φ_1	1064.97	1055.98	493.48	1058.30
	φ_2	988.00	1183.87	645.97	1106.76
	φ_3	1062.92	1058.56	494.15	1059.54
HM	φ_1	.041	.033	.040	.040
	φ_2	.028	.041	.029	.028
	φ_3	.042	.041	.041	.042
	φ_4	.041	.040	.040	.040
	φ_5	.029	.040	.030	.030
	φ_6	.040	.040	.040	.040
	φ_7	.041	.040	.041	.041
	φ_8	.040	.041	.040	.041
	φ_9	.028	.044	.028	.028
	φ_{10}	.031	.076	.032	.031

Sensitivity to K : Complete Results

Property	$K = 2$	$K = 3$	$K = 4$	$K = 5$	$K = 8$	$K = 13$	Reduced states (Constant w.r.t. K)
BRP φ_{11}	.084	.092	.073	.064	.065	.065	1723
CROWDS φ_2	14.54	14.48	14.56	14.58	14.61	14.58	87360
CROWDS φ_5	1.70	1.67	1.82	1.81	2.15	29.25	0
EGL φ_3	3.34	3.36	3.36	3.34	3.35	3.35	1023
Leader φ_1	.47	.46	.46	.48	.47	.47	17
NAND φ_1	493.68	493.79	494.68	494.64	495.05	493.97	10946130
HM φ_2	.041	.029	.034	.038	.041	.037	95

Comparison with Bisimulation Minimization

Bisimulation Minimization [Katoen et al., 2007]

- Merge **bisimilar states**
 - $\mathcal{M} \longrightarrow \mathcal{M}/\simeq$
 - property-independent
- We have $\mathbb{P}(\mathcal{M} \models \psi) = \mathbb{P}(\mathcal{M}/\simeq \models \psi)$
for any LTL formula ψ
- Running example:
6 states $\longrightarrow$ 6 states

Specification-guided path shortcutting (**Ours**)

- Shortcut paths equivalent w.r.t. φ and erase states
 - $(\mathcal{M}, \varphi) \longrightarrow \mathcal{M}'$
 - property-specific
- $\mathbb{P}(\mathcal{M} \models \varphi) = \mathbb{P}(\mathcal{M}' \models \varphi)$ **for the given property φ**
- Running example:
6 states $\longrightarrow$ 4 states

References

- ▶ Duret-Lutz, A. et al. (2022).
From spot 2.0 to spot 2.10: What's new?
In *CAV*, pages 174–187.
- ▶ Hartmanns, A., Klauck, M., Parker, D., Quatmann, T., and Ruijters, E. (2019).
The quantitative verification benchmark set.
In *TACAS*, pages 344–350.
- ▶ Hensel, C., Junges, S., Katoen, J.-P., Quatmann, T., and Volk, M. (2022).
The probabilistic model checker storm.
International Journal on Software Tools for Technology Transfer, 24(4):589–610.
- ▶ Katoen, J.-P., Kemna, T., Zapreev, I. S., and Jansen, D. N. (2007).
Bisimulation minimisation mostly speeds up probabilistic model checking.
In *TACAS*, volume 4424 of *Lecture Notes in Computer Science*, pages 87–101. Springer.