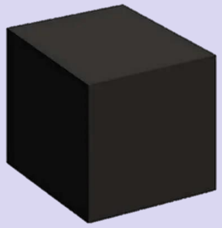
Learning Global Synchronous Protocols: A Broadcast Automata Perspective

Dana Fisman, Noa Izsak, Swen Jacobs

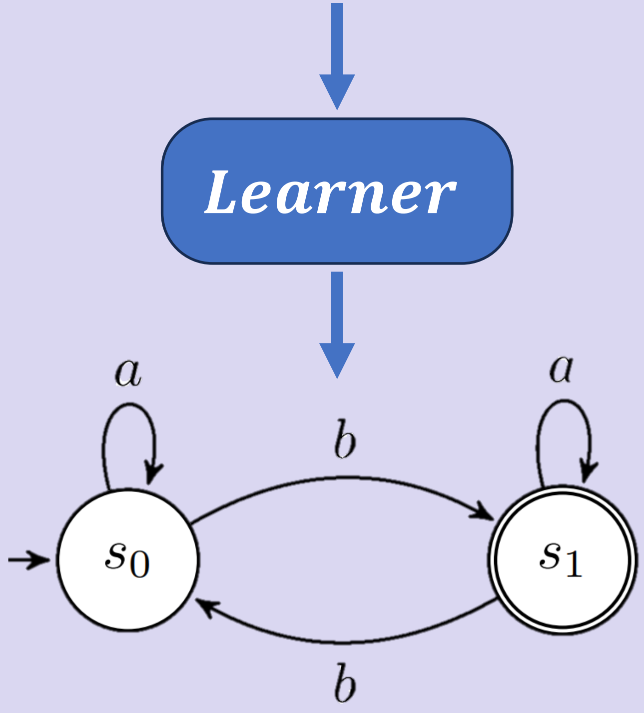
AUTOMATA & WAN 2025

Learning Concurrent systems

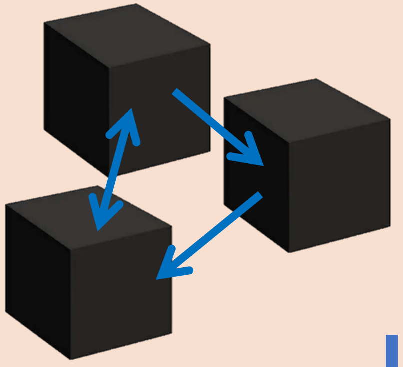
Automata Learning



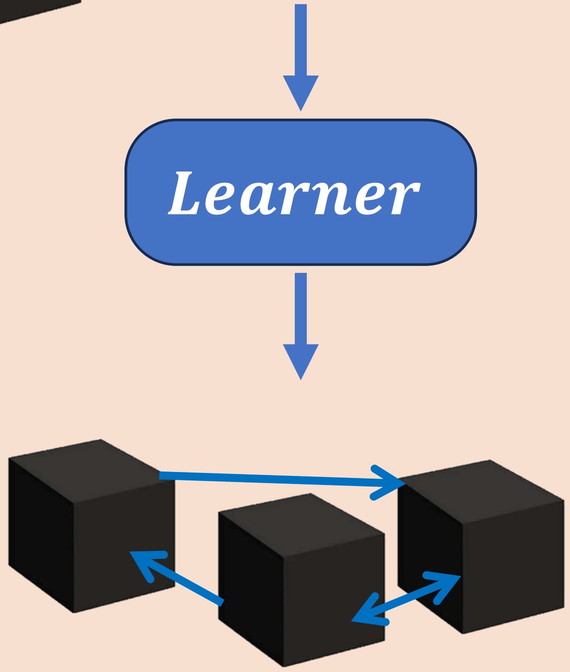
+	-
aba	aa
ba	bba
$abbb$	abb



Fixed size (state of the art)

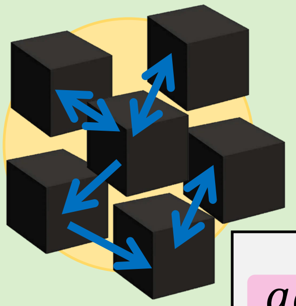


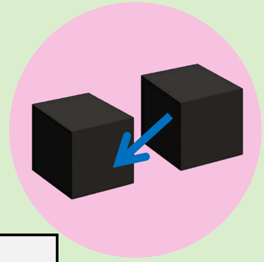
+	-
aba	aa
ba	bba
$abbb$	abb



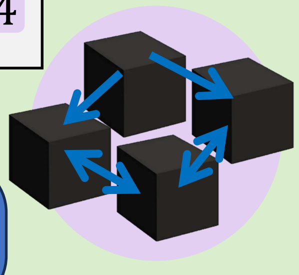
Arbitrary size

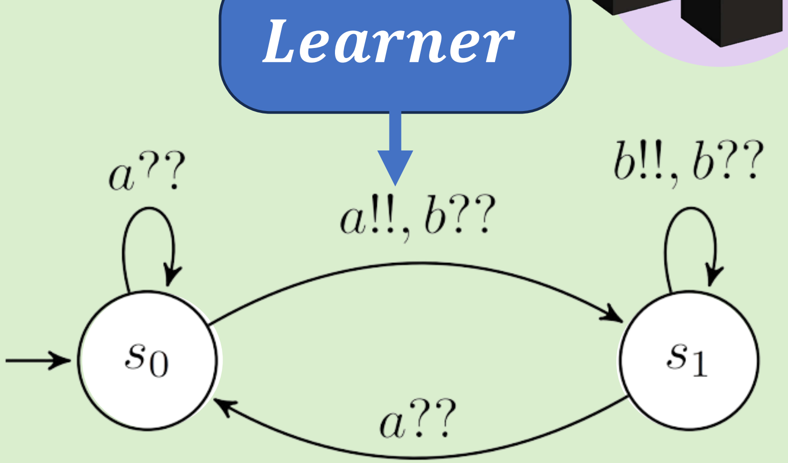
Learning Broadcast Protocols, AAAI24





+	-
$aab, 2$	$aa, 1$
$ab, 1$	$bba, 2$
$abbb, 6$	$abba, 4$





Broadcast Protocol (BP)

BP = Finite-State Machine + Global Synchronous Steps

Given a protocol $P, n \in \mathbb{N}$.

A parallel running of protocol P for n processes: $P^n = \underbrace{P \parallel P \parallel \cdots \parallel P}_{n}$

The “language” of P^n: $\mathcal{L}(P^n)$

The “language” of a protocol P: $\mathcal{L}(P) = \bigcup_{n \in \mathbb{N}} \mathcal{L}(P^n)$

Note that $\mathcal{L}(P^1) \subseteq \mathcal{L}(P^2) \subseteq \mathcal{L}(P^3) \subseteq \cdots$

Are these inclusions strict, or does there exist an n s.t. adding more processes does not change the language?

Cutoff

If $\exists n \in \mathbb{N}$ s.t. $\forall m > n$ $\mathcal{L}(P^n) = \mathcal{L}(P^m)$

If such an n exists, then the system has a **cutoff**, n.
Otherwise, we say there is no cutoff.

Fine BPs

A BP that:

1. Has no **hidden** states

2. A **cutoff exists**

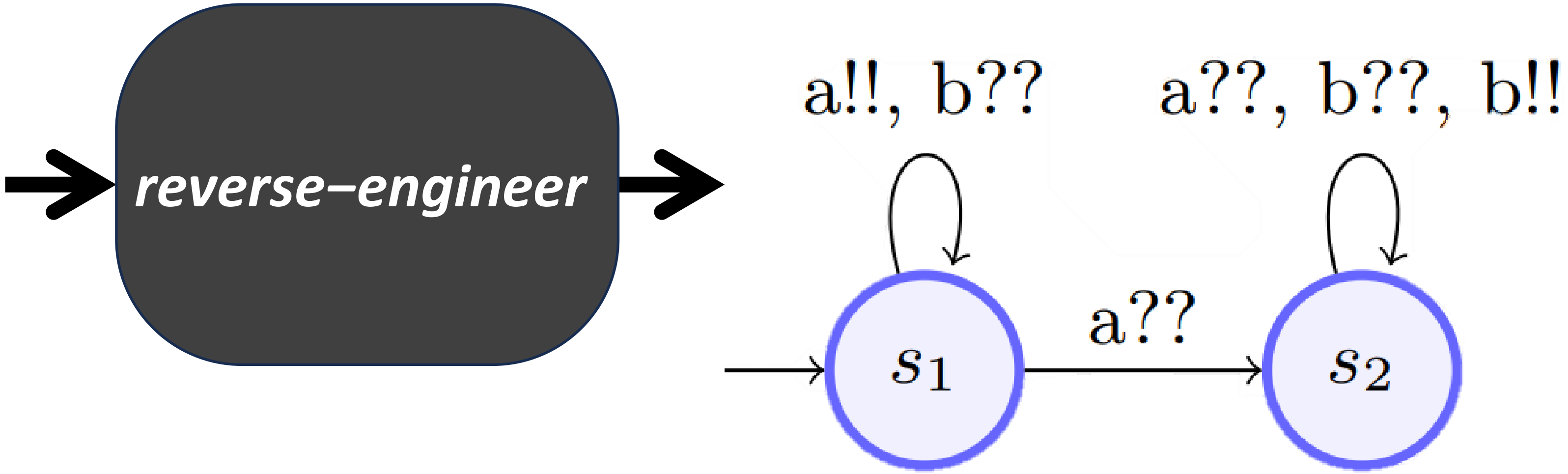
Motivation - Why Learn These Rules?

- Understand distributed behavior
- Debug faulty systems
- Recover legacy protocols
- Support verification

Given traces, can we recover the global rule?

Feasible	Infeasible
(trace,#procs)	(trace,#procs)
$aab, 2$	$ab, 1$
$aaa, 1$	$bba, 2$
$abbb, 6$	$b, 1$
…	…

Consistent sample

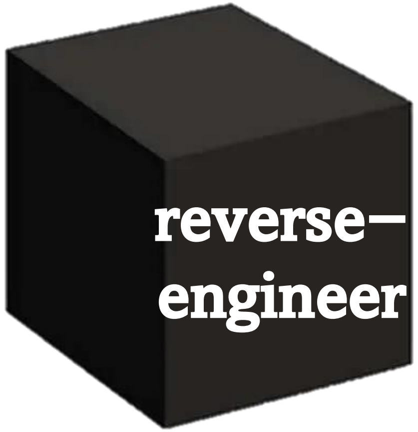


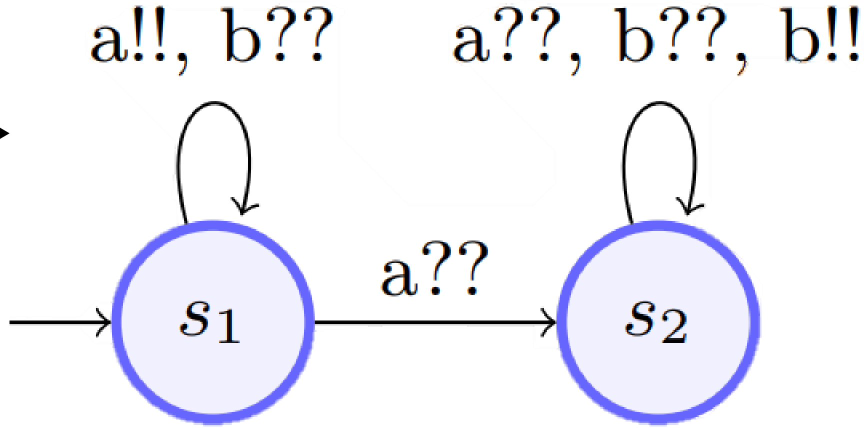
Yes, we can!

Protocols Inference

Feasible	**Infeasible**
(trace,#procs)	(trace,#procs)
$aab, 2$	$ab, 1$
$aaa, 1$	$bba, 2$
$abbb, 6$	$b, 1$
…	…

Consistent sample





Inference

We provide an inference algorithm for BPs,
given a sample of words that are consistent with a BP,
infers a correct BP.

Why is it Hard?

* **Consistency** is **NP-complete**

* **Characteristic sets** can be exponentially large

* **Predictability** is impossible under cryptographic assumptions

No free lunch!

Summary & Takeaways

- Inference → **LeoParDS**
- Consistency
- Polynomial Data
- Polynomial Predictability

[AAAI`24] Learning Broadcast Protocols – Fisman et al.

[ATVA`24] Learning Broadcast Protocols with LeoParDS – Izsak et al.

Thanks for your attention

Contact information

Noa Izsak

izsak@post.bgu.ac.il

https://noa-izsak.github.io/