Learning Broadcast Protocols (Extended Abstract)

Dana Fisman[1,⋆], **Noa Izsak**[2,⋆], **Swen Jacobs**[2,⋆]
[1]Ben-Gurion University
[2]CISPA Helmholtz Center for Information Security
dana@bgu.ac.il, noa.izsak@cispa.de, jacobs@cispa.de

1 Introduction

We study the problem of learning parameterized concurrent systems in the form of broadcast protocols (Fisman, Izsak, and Jacobs 2024). Unlike prior work, which assumes a fixed number of interacting processes, we consider systems running with an arbitrary number of processes. Such systems are inherently parameterized: a finite protocol description represents an infinite family of systems, one for each $n \in \mathbb{N}$, where n is the number of processes.

2 Broadcast Protocols and Fine BPs

Broadcast Protocols. A *broadcast protocol* (*BP*) is a tuple $B = (S, s_0, L, R)$ where S is a finite set of *states* with *initial state* $s_0 \in S$, $L = \{a!!, a?? \mid a \in A\}$ for a finite set of *actions* A, and $R \subseteq S \times L \times S$ is the transition relation. A transition labeled $a!!$ is a broadcast sending transition, and $a??$ is the corresponding receiving transition. Following Esparza et al. (1999), for each $a \in A$, there exists exactly one state $s_a \in S$ that *enables* $a!!$, and from every state $s \in S$ there is exactly one outgoing $a??$ transition.

Parameterized Systems. Given a BP B and $n \in \mathbb{N}$, the system B^n consists of n identical processes executing B concurrently. A *global transition* of B^n on action $a \in A$ is one in which a single process (in state s_a) performs an $a!!$-transition, and all other $(n-1)$ processes simultaneously perform their respective $a??$ transitions; that is, all n processes participate in every global transition. A word $w \in A^*$ is *feasible* in B^n if it labels an execution of B^n; the language of B^n is denoted $L(B^n)$, and $L(B) = \bigcup_{n \in \mathbb{N}} L(B^n)$.

Fine Broadcast Protocols. We focus on *fine broadcast protocols*, namely BPs that: (i) have no hidden states (i.e., every state enables at least one sending transition), and (ii) admit a *cutoff*, i.e., a number $c \in \mathbb{N}$ such that for all $n \geq c$: $L(B^n) = L(B^c)$. Fine BPs form a subclass with natural restrictions that enable learnability. In particular, the cutoff property captures a semantic stabilization phenomenon: beyond a certain number of processes, no new observable behaviors arise. Importantly, the learning framework does not assume that the cutoff is known a priori; reconstruction must rely solely on observed samples, without knowledge of the cutoff.

⋆These authors contributed equally to this work

3 Learning Framework

A *sample* is a finite set $\mathcal{S} \subseteq A^* \times \mathbb{N} \times \{\mathsf{T}, \mathsf{F}\}$ of labeled examples. A BP B is *consistent* with $\mathcal{S}$, written $B \vDash \mathcal{S}$, if every $(w, n, \mathsf{T}) \in \mathcal{S}$ satisfies $w \in L(B^n)$ and every $(w, n, \mathsf{F}) \in \mathcal{S}$ satisfies $w \notin L(B^n)$. Observing n is essential: feasibility depends on the number of participating processes, making the learning problem inherently parameter-sensitive.

We study four learning problems for the class of fine BPs:

1. **Inference.** Given a sample $\mathcal{S}$ consistent with a fine BP, return a BP consistent with $\mathcal{S}$.
2. **Consistency.** Given a sample $\mathcal{S}$ and a bound $k \in \mathbb{N}$, determine whether there exists a consistent fine BP with at most k states.
3. **Polynomial data.** Determine whether every fine BP B admits a *characteristic set* (a finite sample that suffices for minimal reconstruction) whose size is polynomial in the size of B.
4. **Polynomial predictability.** Determine whether fine BPs are efficiently predictable in an active learning model with membership and draw queries.

Although $L(B)$ is regular whenever a cutoff exists, the learning problem cannot be reduced to regular language inference: feasibility depends on the process parameter n, and the objective is to recover a concise BP rather than an arbitrary language acceptor. The learning problem is thus a representation-theoretic question rather than merely a language-recognition task. Indeed, the BP representation may be exponentially more succinct than a minimal DFA that recognizes the same language.

Figure 1 illustrates the semantic reconstruction setting. In Figure 1**(a)**, the *characteristic set* (CS) *generation procedure* receives a BP B and generates a finite sample CS_B of labeled examples (w, n, b) such that $B \vDash \mathsf{CS}_B$. Any sample subsuming CS_B allows reconstruction of a minimal fine BP equivalent to B under parameterized semantics. In Figure 1**(b)**, the *inference procedure* receives a sample $\mathcal{S}$ and returns a BP B' with $B' \vDash \mathcal{S}$; if additionally $\mathsf{CS}_B \subseteq \mathcal{S}$ for some BP B, then $L(B') = L(B)$ and B' is minimal.

4 Main Results

Inference. We provide a passive learning algorithm that, given a sample consistent with a fine BP, constructs a BP

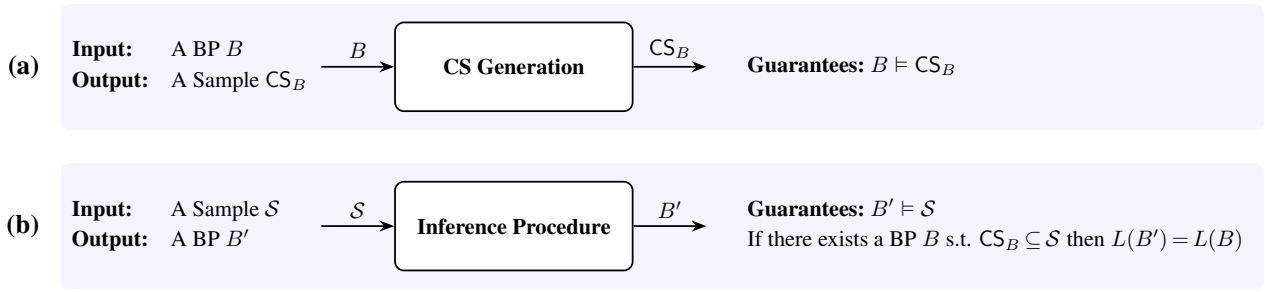

Figure 1: Semantic reconstruction setting. A finite labeled sample $(w, n, \mathsf{T}/\mathsf{F})$ constrains the inferred BP. If the sample subsumes a CS, the inference procedure returns a minimal fine BP B' with $L(B') = L(B)$; otherwise, it returns a BP consistent with the sample.

consistent with the sample. The reconstruction problem is encoded as constraints in the theory of equality with uninterpreted functions (EUF), enabling implementation via an SMT solver. Moreover, if the sample subsumes a suitable characteristic set, the inference procedure is guaranteed to return a minimal fine BP equivalent to the input BP under parameterized semantics.

A central insight underlying minimal reconstruction is that, unlike DFAs, minimal fine broadcast protocols are not canonical: distinct non-isomorphic minimal fine BPs may be semantically equivalent and share the same cutoff. A minimal fine BP is therefore not a unique representative of its semantic equivalence class. Nevertheless, we show that any two such minimal semantically equivalent fine BPs admit a tight structural correspondence between their states preserving enabled actions and reachability patterns, which ensures the correctness of minimal reconstruction despite non-canonicity.

Limits of Learnability. We establish three hardness results concerning the class of fine BPs.

1. **Exponential Characteristic Sets.** There exists a family of fine BPs for which any characteristic set must be of exponential size in the number of states of the BP.
2. **NP-Hard Consistency.** The consistency problem for fine BPs is NP-hard. This is shown via a non-trivial reduction from DFA consistency; a DFA is not a special case of a fine BP, but a fine BP can simulate any DFA.
3. **Non-Polynomial Predictability.** Under standard cryptographic assumptions, fine BPs are not polynomially predictable with membership queries. This is shown via a reduction from predicting intersections of DFAs.

Together, these results delineate a sharp boundary: while a consistent BP can always be inferred from a sample, minimal reconstruction (recovering a fine BP that fully characterizes the parameterized semantics) may inherently require exponentially large distinguishing information, and efficient active prediction is impossible under accepted assumptions.

5 Relevance to Knowledge Representation

Parameterized distributed systems can be viewed as objects of representation learning, where the goal is not merely behavior approximation but recovery of a syntactic structure that captures precise semantics for systems with an arbitrary number of processes. Our results establish the learnability and hardness boundaries for such representations and formalize the reconstruction of models of concurrent systems without assuming a fixed number of interacting processes. The contribution thus falls within the KR line of research on reasoning about concurrent systems, employing methods such as automata theory and computational learning theory.

6 Conclusion and Outlook

We have studied the learnability of fine broadcast protocols, establishing an inference procedure and three hardness results. An implementation of the inference framework, **LeoParDS** (Izsak et al. (2024)), has been developed and evaluated, demonstrating practical viability; a concise overview targeting the verification community appears in Fisman et al. (2025). Ongoing work investigates extensions to alternative synchronization mechanisms and to a broader class of concurrent systems.

Acknowledgments

Noa Izsak carried out this work in part as a member of the Saarbrücken Graduate School of Computer Science.

References

Esparza, J.; Finkel, A.; and Mayr, R. 1999. On the verification of broadcast protocols. In *LICS*, 352–359.

Fisman, D.; Izsak, N.; and Jacobs, S. 2024. Learning broadcast protocols. *Proceedings of the AAAI Conference on Artificial Intelligence* 38(11):12016–12023.

Fisman, D.; Izsak, N.; and Jacobs, S. 2025. Insights into learning broadcast protocols: (short paper). In *International Symposium on Cyber Security, Cryptology, and Machine Learning*, 306–313. Springer.

Izsak, N.; Fisman, D.; and Jacobs, S. 2024. Learning broadcast protocols with leopards. In *International Symposium on Automated Technology for Verification and Analysis*, 220–234. Springer.