

# Active Learning from Blackbox to Timed Connectors

Yi Li\*, Yiwu Wang\* and Meng Sun\*

\*Department of Informatics, School of Mathematical Sciences, Peking University, Beijing, China  
liyi\_math@pku.edu.cn, yiwuwang@126.com, summeng@math.pku.edu.cn

**Abstract**—Coordination models and languages play a key role in formally specifying the communication and interaction among different components in large-scale distributed and concurrent systems. In this paper, we propose an active learning framework to extract timed connector models from black-box system implementation. We first introduce parameterized Mealy machine(PMM) as an operational semantic model for channel-based coordination language Reo. Parameterized Mealy machine serves as a bridge between Reo connectors and mealy machines. With the product operator, complex connectors can be constructed by joining basic channels and transformed into mealy machines. Moreover, we adapt L\*, a well-known learning algorithm, to timed connectors (in the form of mealy machines). The new algorithm has shown its efficiency in multiple case studies. Implementations of this framework is provided as a package in Golang.

**Index Terms**—Active Learning, Coordination Languages, Timed Connectors

## I. INTRODUCTION

Distributed real-time embedded systems (DRES) are reforming our lives with the Internet of Things(IoT), wherein individual components are composed via a *connector* to build a system. Such systems could be distributed logically or physically, which makes coordination processes even more complicated. In this case, we need to specify the coordination processes with *coordination languages*, such that formal techniques can be applied to guarantee their reliability.

*Timed Reo* is a real-time extension of the coordination language *Reo*, by which coordination process in DRES can be depicted clearly and intuitively. Different formal semantics are proposed to specify the behavior of timed Reo. For example, in [14], a UTP based (*Unifying Theories of Programming*) semantics is provided to verify connectors as designs. An operational semantics based on *constraint automata* was raised by Baier et al. [3], where a variant of LTL was also proposed to describe the properties of timed connectors.

Formal verification and validation techniques have also been proved applicable for timed connectors, e.g. a conformance testing framework on timed connectors in [13]. A SAT-based bounded model checking approach was adapted in [12] for verifying connectors. Although these solutions are practical and impressive, there is one important question faced by most of them: *how to obtain the formal models?*

*Correctness* of connectors is closely related to some low-level implementation details. For example, well-written code may behave dramatically weird with an improper set of concurrency primitives. Such *scenarios* happen frequently in

embedded systems with different hardware platforms or operating systems. Consequently, manually modeling an existing connector seems rather unreliable, even with reference to its source code.

As a branch of *machine learning* technique, active learning offers a way to obtain models from low-level models. Works in [1], [7], [17] shows interesting examples where active learning is used to extract *Mealy machines* or *regular languages* without time domain.

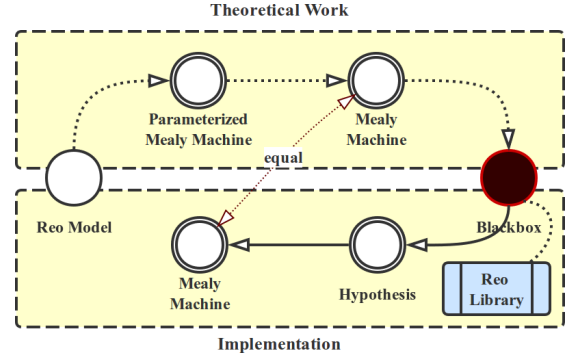


Fig. 1. The Active Learning Framework

In this paper we address this question by proposing a active learning framework as shown in Figure 1) to automatically extract timed connector models from blackbox implementations. To achieve this goal, we first introduce *parameterized Mealy machine* as a *parameterized* semantics for timed Reo connectors, which can be transformed into concrete Mealy machines with a given alphabet. Then the L\* algorithm [1] is adapted and optimized to extract Mealy machines with timed action from the connector implementations as blackboxes.

The rest of the paper is organized as follows: In Section II we briefly illustrate some basic concepts, including the coordination language Reo, Mealy machine, and active automata learning. Section III defines an operational semantics of Reo, which is used in Section IV to show how to extract Reo models from blackboxes by means of active learning. In Section V, we discuss the implementation and optimization of the approach in Golang.

## II. PRELIMINARIES

### A. Reo Coordination Language

We provide here a brief overview of the main concepts in Reo. Further details can be found in [2], [4].

Reo is a channel based exogenous coordination language proposed by F. Arbab in [2]. A coordinator in Reo, also called *connector*, provides the protocol that controls and organizes the communication, synchronization and cooperation among the components which communicate through the connector. Connectors can be defined without dependence on components, which makes Reo a powerful “glue language” in component-based development [9].

In Reo, complex connectors are made up of simpler ones. The atomic connectors are called *channels*, and each of them has two *channel ends*. There are two types of channel ends: *source* and *sink*. Source channel ends accept data into the channel, while sink channel ends dispense data out of the channel.

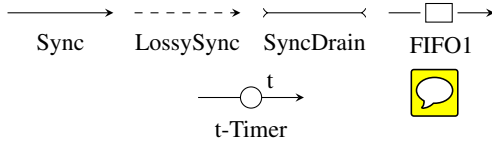


Fig. 2. Basic Reo Channels

Figure 2 shows the graphical representation of some basic channel types in Reo, whose behavior are described as follows:

- A *Sync* channel accepts a data item from its source end iff the data item can be dispensed through its sink end simultaneously.
- A *LossySync* channel is always ready to accept data items. These items will be dispensed through its sink end  $\lambda$  if possible, otherwise they will be lost.
- A *SyncDrain* channel has two source ends and no sink end. It can accept a data item through one of its source end iff a data item is also available to be accepted simultaneously through the other end. Then both items will be dropped.
- A *FIFO1* channel is an asynchronous channel with a buffer cell. It accepts a data item whenever the buffer is empty.
- A *t-timer* channel accepts any data item from its source end, and later dispense it to its sink end after a delay of  $t$  time units  $\lambda$ .

Channels are joint together in *nodes*. There are three types of nodes: *source*, *sink* and *mixed* nodes, depending on whether all channel ends that coincide on a node are source ends, sink ends or a combination of both. (see in Figure 3)

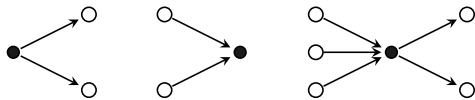


Fig. 3. Source, Sink and Mixed Nodes in Reo

Components can be linked to source nodes or sink nodes. A component can write data items to its corresponding source node only if the data item can be dispensed simultaneously to all source ends on this node. Similarly, a component can read a data item only if there is at least one readable sink end on its corresponding node. A mixed node non-deterministically selects and takes a data item from one of its coincident sink ends and replicates it into all of its coincident source ends.

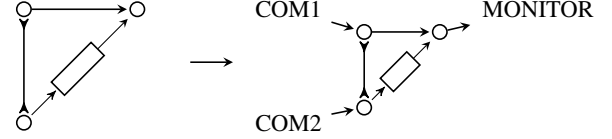


Fig. 4. Coordination with Reo Connectors

Figure 4 provides a simple example to illustrate how complex connectors are constructed and used in coordination. In this example, we have three components COM1, COM2 and MONITOR. With an *alternator* connector as shown in Figure 4, MONITOR can obtain data items from COM1 and COM2 alternately. This example has been implemented in *Golang* as shown in Section V-C. Besides, all the connectors are supposed to have *deterministic* behavior hereinafter to meet the requirement of active learning algorithm  $L^*$ .

### B. Mealy Machines

In this work, we use *Mealy machines* to model Reo connectors. Mealy machine was first proposed by George. H. Mealy in [8] as an extension of finite state machine. Compared with other variants, Mealy machines are designed to model reactive systems, where outputs of a machine are determined by not only its current state but also the current inputs. Besides, Mealy machines are supposed to be *input enabled*, which means that all possible inputs should be acceptable in all states. In other words, if an input is invalid for some state, we need to manually define an additional state to describe such exceptions.

Various forms of Mealy machines have been defined in the literature. Since active automata learning requires the system-under-learn to be deterministic, in this paper we take the deterministic Mealy machine model as defined in [19].

**Definition 1 (Mealy machine):** A Mealy machine is a 6-tuple  $(S, s_0, I, O, \delta, \lambda)$  consisting of:

- a finite set of states  $S$
- an initial  $s_0 \in S$
- a finite set of inputs  $I$
- a finite set of outputs  $O$
- an output function  $\delta : S \times I \rightarrow O$  mapping a pair of a state and an input to the corresponding output.
- a transition function  $\lambda : S \times I \rightarrow S$  mapping a pair of a state and an input to the corresponding successor state.

We say that a Mealy machine is *finite* if the set  $S$  of states and the set  $I$  of inputs are both finite. Hereinafter, Mealy machines are always supposed to be finite in this paper.

### C. Active Learning

In this section, we briefly introduce the main ideas of active automata learning.

Active learning [18] is a special case of semi-supervised machine learning where a learning algorithm is able to interactively query the target systems to obtain the desired outputs under certain inputs. By well-designed query strategy, active learning is able to obtain more accurate models with smaller dataset.

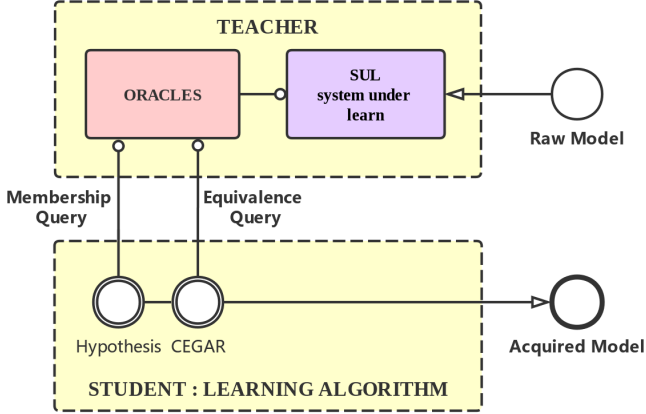


Fig. 5. Active Automata Learning

Figure 5 shows the sketch of active learning, wherein:

- *Teacher* and *Student*: Active learning is an interactive process where students ask questions and teachers answer. Here learning algorithm plays the role of student.
- *Oracles* is an interface specifying which kind of questions can be answered by the teacher.
- *SUL* is an abbreviation of System Under Learn. In this paper, we take blackbox models as our SULs.
- *CEGAR* indicates Counter-Example Guided Abstraction Refinement [6]. In active learning, we need counter-examples to guide us on further queries and cover the undistinguished states.

When applying active automata learning on some model, we assume that the model should be equivalent to some *Mealy machine*, i.e., a deterministic model accepting a finite set of inputs and mapping them to a finite set of outputs.

Such a model is encapsulated as a *teacher* by the *Oracle*, which handles all communication with the model. The oracle also serves as a so-called *Minimal Adequate Teacher* interface, which is responsible for two types of queries.

- **Membership Query** (hereinafter referred to as *mq*) In grammar-learning [1], *mq* checks if a word is a member of certain language defined by the given grammar. When it comes to automata learning, *mq* is supposed to provide *simulation results* for given input sequences.
- **Equivalence Query** (hereinafter referred to as *eq*) Given a hypothesis (usually constructed by the learning algorithm), *eq* checks whether the hypothesis is equivalent to

the system-under-learn and generates a counter-example if needed. Generally, *eq* is unrealizable when SUL is a blackbox. So we tactically use *mq* to achieve the approximate results.

These queries are given by *learning algorithms*, or so-called *students* in Figure 5. From the *mq* results, a learning algorithm constructs a *hypothesis* and then check it with *eq*. If counter-examples are found, we turn back and repeat the hypothesis construction until the equivalence query returns *true*.

More details on the active automata learning algorithm will be presented in Section IV.

### III. TIMED CONNECTORS AS MEALY MACHINES

In this section, we illustrate how to transform a timed connector into a Mealy machine with timed action. Since time is not involved in original Mealy machine, we first discuss how to formalize the time dimension in Mealy machine. After that, we present the notion of *parameterized Mealy machine* as a bridge between connectors and Mealy machines.

#### A. Time Domain

Time has been investigated in several extensions of Reo. For example, timed Reo [3], [15], hybrid Reo [5], etc. Generally, these models are designed to handle real-time behavior where time is defined on the real number field  $\mathbb{R}$ . However, there are also ~~found~~ some works like [16] where rational time indeed makes things easier. In this paper, we choose the rational number field  $\mathbb{Q}$  as our time domain, which simplifies discretization of timed behavior greatly.

As presented in section II-A, all real-time behavior in timed Reo comes with a finite set of timed channels, ~~and the number of these channels are apparently finite~~. We use  $t_i \in \mathbb{Q}, i = 1, 2, \dots, n$  to denote the delays of these timed channels, and then we can define a precision function *prec* which calculate a gcd-style *maximal time precision*:

$$prec(t_1, \dots, t_n) = \max\{T \in \mathbb{Q} | \forall t_i. \exists n_i \in \mathbb{N}. t_i = n_i \cdot T\}$$

It's easy to prove that such a *T* always exists.

In embedded systems, the maximal time precision is also called “clock-period”, which is the basic time unit provided by an oscillator. For example, a widely-used Intel-80C51 microcontroller works under the frequency 16Mhz, where the clock-period is 62.5ns. A reasonable assumption comes that such a maximal time precision  $dt = prec(t_1, \dots, t_n)$  should always be provided to make sure different components are able to work together. Then all  $t_i$ -Timers can be seen as  $n_i dt$ -Timers for some  $n_i$ . ~~In the following sections, we'll use  $n_i$ -Timers instead.~~

Besides, we're going to add a “T” action in Mealy machines. It indicates that the corresponding transition will take one time unit to finish.

#### B. Parameterized Mealy Machine

We present a model named *parameterized Mealy machine* to model timed connectors. PMM is supposed to behave as a bridge between Mealy machines and timed connectors.

Connectors are firstly defined as PMMs, and composed by production operator and link operator. Then, concrete Mealy-machine model can be generated from the PMM model. Following the formal definition of Mealy machine in Section II, we define the *parameterized Mealy machine* as follows.

**Definition 2 (Parameterized Mealy Machine):** A *Parameterized Mealy Machine* with a parameter  $\Sigma$  is defined as a 6-tuple  $\mathcal{PM} = \langle S, s_0, I, O, \delta, \lambda \rangle$  where

- The value of  $\Sigma$  is a *finite* data set (hereinafter referred as alphabet), e.g.  $\{a, b\}$
- $S$  maps an alphabet to a *finite* set of states.
- $s_0$  is the initial state. It satisfies  $\forall \Sigma. s_0 \in S(\Sigma)$
- $I$  is a finite set of source-ends.
- $O$  is a finite set of sink-ends.
- $\delta$  maps an alphabet to an *output function*. We use  $\delta(\Sigma) : S(\Sigma) \times Input(\Sigma, I, O) \rightarrow Output(\Sigma, O)$  to denote the output function.
- $\lambda$  maps an alphabet to a *transition function*. We use  $\lambda(\Sigma) : S(\Sigma) \times Input(\Sigma, I, O) \rightarrow S(\Sigma)$  to denote the transition function.

In the definition above, *Input* and *Output* are used to generate the set of input actions and output actions from the corresponding alphabets and source/sink ends.  $Input(\Sigma, I, O)$  is defined as a set of functions on  $I \cup O$  and an additional *time action*  $T$ , where for all  $f \in Input(\Sigma, I, O) \setminus \{T\}$  we have

$$\forall i \in I. f(i) \in \Sigma \cup \{\perp\} \wedge \forall o \in O. f(o) \in \{\triangleright, \ominus\}$$

where we use  $\perp$  to indicate that there is *no* data item on a channel end. Besides,  $\triangleright$  means that a sink end is ready for writing,  $\ominus$  otherwise. In the same way,  $Output(\Sigma, O)$  is defined as a set of functions on  $O$ , with an additional symbol  $\ominus$  indicating an input failure. Similarly, for all  $f \in Output(\Sigma, O) \setminus \{\ominus\}$ ,

$$\forall o \in O. f(o) \in \Sigma \cup \{\perp\}$$

**Example 1 (Input and Output):** If we have a simple alphabet with only one item  $d_0$ , a source-end named  $A$ , and a sink-end named  $B$ , the input actions would be

$$\begin{aligned} & Input(\{d_0\}, \{A\}, \{B\}) \\ = & \{ \{A \mapsto d_0, B \mapsto \triangleright\}, \{A \mapsto d_0, B \mapsto \ominus\}, \\ & \{A \mapsto \perp, B \mapsto \triangleright\}, \{A \mapsto \perp, B \mapsto \ominus\}, T \} \end{aligned}$$

and its output actions

$$Output(\{d_0\}, \{A\}, \{B\}) = \{ \{B \mapsto d_0\}, \{B \mapsto \perp\}, \ominus \}$$

We use  $\emptyset$  to denote the empty output when no data item is written to any sink end. For example, in this case  $\{B \mapsto \perp\}$  can be briefly denoted as  $\emptyset$ .

Parameterized Mealy machines can be seen as an abstraction of Mealy machines, and we still need a convert function between the two models.

**Definition 3 (Concretize Mapping):** We use  $[\mathcal{PM}]_\Sigma$  to denote the concrete Mealy machine which is determined by an

abstract parameterized Mealy machine  $\mathcal{PM}$  and the alphabet  $\Sigma$ . Apparently  $[\mathcal{PM}]_\Sigma$  can be described as

$$\begin{aligned} [\mathcal{PM}]_\Sigma = & \langle \mathcal{PM}.S(\Sigma), \mathcal{PM}.s_0, \\ & Input(\Sigma, \mathcal{PM}.I, \mathcal{PM}.O), Output(\Sigma, \mathcal{PM}.O), \\ & \mathcal{PM}.\delta(\Sigma), \mathcal{PM}.\lambda(\Sigma), \rangle \end{aligned}$$

Now we can use PMMs to specify timed Reo. Here we take the asynchronous channel (FIFO1), the synchronous channel (Sync) and the timed channel (Timer) as examples. For simplicity, we use  $A = \_$  to indicate that there is no constraint on channel end  $A$ .

**Example 2 (FIFO1):** The PMM of a FIFO1 with source end  $A$  and sink end  $B$  can be defined as follows.

- $S(\Sigma) = \{q_0\} \cup \{q_d | d \in \Sigma\}$
- $I = \{A\}$ ,  $O = \{B\}$ ,  $s_0 = q_0$
- output function

$$\delta(\Sigma)(s, i) = \begin{cases} \emptyset & s = q_0 \wedge i = \{A \mapsto \_, B \mapsto \ominus\} \\ \ominus & s = q_0 \wedge i = \{A \mapsto d', B \mapsto \triangleright\} \\ \ominus & s = q_0 \wedge i = \{A \mapsto \perp, B \mapsto \triangleright\} \\ \emptyset & s = q_d \wedge i = \{A \mapsto \perp, B \mapsto \ominus\} \\ \{B \mapsto d\} & s = q_d \wedge i = \{A \mapsto \_, B \mapsto \triangleright\} \\ \ominus & s = q_d \wedge i = \{A \mapsto d', B \mapsto \ominus\} \end{cases}$$

- transition function

$$\lambda(\Sigma)(s, i) = \begin{cases} q_{d'} & s = q_0 \wedge i = \{A \mapsto d', B \mapsto \_ \} \\ q_{d'} & s = q_d \wedge i = \{A \mapsto d', B \mapsto \triangleright \} \\ q_d & s = q_d \wedge i = \{A \mapsto \_, B \mapsto \ominus \} \\ q_0 & \text{otherwise} \end{cases}$$

Besides, the concrete Mealy machine (where  $\Sigma = \{d\}$ ) is shown in Figure 6, where labels of edges are in form of  $\frac{input}{output}$ . All  $\ominus$  edges are ignored for clearance.

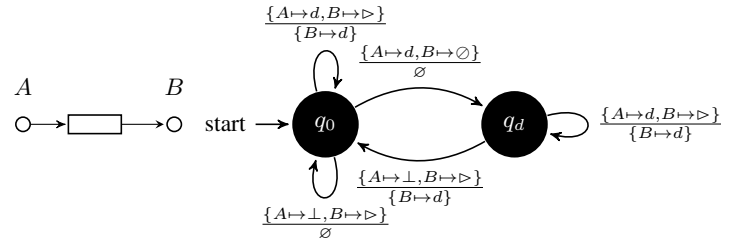


Fig. 6. PMM-based Semantics of  $[FIFO1]_\Sigma$ , where  $\Sigma = \{d\}$

**Example 3 (Sync):** The PMM of a Sync with a source-end  $A$  and a sink-end  $B$  can be defined as:

- $S(\Sigma) = \{q_0\}$ ,  $I = \{A\}$ ,  $O = \{B\}$ ,  $s_0 = q_0$
- output function

$$\delta(\Sigma)(s, i) = \begin{cases} \emptyset & i = \{A \mapsto \perp, B \mapsto \ominus\} \\ \{B \mapsto d\} & i = \{A \mapsto d, B \mapsto \triangleright\} \\ \ominus & i = \{A \mapsto \perp, B \mapsto \triangleright\} \\ \ominus & i = \{A \mapsto d, B \mapsto \ominus\} \end{cases}$$

- transition function  $\lambda(\Sigma)(s, i) = q_0$ .

**Example 4 (n-Timer):** The PMM of an n-Timer with a source-end  $A$  and a sink-end  $B$  can be defined as:

- $S(\Sigma) = \{q_{i,d} | 0 \leq i \leq n, d \in \Sigma\} \cup \{q_0\}$



-  $I = \{A\}, O = \{B\}, s_0 = q_0$

- output function

$$\delta(\Sigma)(s, i) = \begin{cases} \{B \mapsto d\} & s = q_{n,d} \wedge i = \{A \mapsto \neg, B \mapsto \triangleright\} \\ \ominus & s = q_{n,d} \wedge (i = T \vee i = \{A \mapsto \neg, B \mapsto \odot\}) \\ \ominus & s = q_{j,d} \wedge 0 \leq j < n \wedge (i = \{A \mapsto \neg, B \mapsto \triangleright\} \vee i = \{A \mapsto d, B \mapsto \neg\}) \\ \emptyset & \text{otherwise} \end{cases}$$

- transition function

$$\lambda(\Sigma)(s, i) = \begin{cases} q_{0,d} & s = q_0 \wedge i = \{A \mapsto d, B \mapsto \neg\} \\ q_{j+1,d} & s = q_{j,d} \wedge i = T \wedge 0 \leq j < n \\ q_0 & s = q_{n,d} \wedge i = \{A \mapsto \neg, B \mapsto \triangleright\} \\ q_{0,d'} & s = q_{n,d} \wedge i = \{A \mapsto d', B \mapsto \triangleright\} \\ q_0 & s = q_{n,d} \wedge i = T \\ s & \text{otherwise} \end{cases}$$

Similarly, we can use PMMs to describe the behavior of other basic timed Reo channels. Now we show how to compose these channels into more complicated connectors.

**Definition 4 (Production):** The *production* operator *prod* of two *parameterized Mealy machines*

$$\mathcal{PM}' = \text{prod}(\mathcal{PM}_1, \mathcal{PM}_2) \text{ if } \mathcal{PM}_2.O \cap \mathcal{PM}_1.I = \emptyset$$

is defined as follows.

- $\forall \Sigma. \mathcal{PM}'.S(\Sigma) = \mathcal{PM}_1.S(\Sigma) \times \mathcal{PM}_2.S(\Sigma)$
- $\mathcal{PM}'.I = (\mathcal{PM}_1.I \cup \mathcal{PM}_2.I) \setminus \mathcal{PM}_1.O$
- $\mathcal{PM}'.O = (\mathcal{PM}_1.O \cup \mathcal{PM}_2.O) \setminus \mathcal{PM}_2.I$

(Here we assume that sink ends of  $\mathcal{PM}_1$  can be connected to source ends  $\mathcal{PM}_2$ , but not vice versa)

- $\mathcal{PM}'.s_0 = (\mathcal{PM}_1.s_0, \mathcal{PM}_2.s_0)$
- $\forall \Sigma. \mathcal{PM}'.\delta(\Sigma)((s_1, s_2), i) =$

$$\begin{cases} (Out_1 \cup Out_2) \downarrow_{\mathcal{PM}'.O} & Out_1 \neq \ominus \wedge Out_2 \neq \ominus \\ \emptyset & i = T \\ \ominus & \text{otherwise} \end{cases}$$

where we have

- \*  $In_1 = i \downarrow_{\mathcal{PM}_1.I}$
- \*  $Out_1 = \mathcal{PM}_1.\delta(\Sigma)(s_1, In_1)$
- \*  $In_2 = (Out_1 \cup i) \downarrow_{\mathcal{PM}_2.I}$
- \*  $Out_2 = \mathcal{PM}_2.\delta(\Sigma)(s_2, In_2)$
- $\forall \Sigma. \mathcal{PM}'.\lambda(\Sigma)((s_1, s_2), i) = (s'_1, s'_2)$  where we have
  - \*  $s'_1 = \mathcal{PM}_1.\lambda(\Sigma)(s_1, In_1)$
  - \*  $s'_2 = \mathcal{PM}_2.\lambda(\Sigma)(s_2, In_2)$

The idea in *production* is quite simple. The output of  $\mathcal{PM}_1$  would be provided as part of the input of  $\mathcal{PM}_2$ . Time actions T is only executed simultaneously between  $\mathcal{PM}_1$  and  $\mathcal{PM}_2$ .

As mentioned in the notation above, we cannot connect a sink end of  $\mathcal{PM}_2$  to a source end of  $\mathcal{PM}_1$ , which makes it difficult to construct connectors like alternator in Figure 4. To address the problem, we define a *link* operation to connect sink ends and source ends in the same connector.

**Definition 5 (Link):** The *link* operator constructs connector by connecting a sink end OUT to a source end IN within the same connector.

$$\mathcal{PM}' = \text{link}(\mathcal{PM}, \text{OUT}, \text{IN})$$

Here we assume that  $\text{IN} \in \mathcal{PM}.I$  and  $\text{OUT} \in \mathcal{PM}.O$ .

- $\forall \Sigma. \mathcal{PM}'.S(\Sigma) = \mathcal{PM}.S(\Sigma)$
- $\mathcal{PM}'.I = \mathcal{PM}.I \setminus \{\text{IN}\}$
- $\mathcal{PM}'.O = \mathcal{PM}.O \setminus \{\text{OUT}\}$
- $\mathcal{PM}'.s_0 = \mathcal{PM}.s_0$
- $\forall \Sigma, i \in \text{Input}(\Sigma, \mathcal{PM}'.I, \mathcal{PM}'.O). \mathcal{PM}'.\delta(\Sigma)(s, i) =$ 

$$\begin{cases} \emptyset & i = T \\ \mathcal{PM}.\delta(\Sigma)(s, i \cup \{\text{IN} : d\}) \downarrow_{\mathcal{PM}'.O} & (*) \\ \ominus & \text{otherwise} \end{cases}$$

where the condition (\*) is defined as

$$\exists d. \mathcal{PM}.\delta(\Sigma)(s, i \cup \{\text{IN} : d\})(\text{OUT}) = d$$

- $\forall \Sigma, i \in \text{Input}(\Sigma, \mathcal{PM}'.I, \mathcal{PM}'.O).$

$$\mathcal{PM}.\lambda(\Sigma)(s, i) = \begin{cases} \mathcal{PM}.\lambda(\Sigma, T) & i = T \\ \mathcal{PM}.\lambda(\Sigma, i \cup \{\text{IN} : d\}) & (*) \\ s & \text{otherwise} \end{cases}$$

With *link* and *prop* defined, a connector can be formed by simpler ones in form of *parameterized Mealy machines*. It will be transformed into a concrete Mealy machine later, once the alphabet  $\Sigma$  is provided.

#### IV. FROM BLACKBOX TO TIMED CONNECTORS

In this section, we show how to use L\* algorithm to extract models from blackboxes in 3 steps: *constructing hypothesis*, *enclosing hypothesis* and *checking hypothesis*. Besides, we assume that input symbols and output symbols of the blackbox have been provided as  $\mathcal{I}$  and  $\mathcal{O}$ .

##### A. Observation Table

As mentioned above, membership queries provide us information with the form  $mq(s) = o$ , where  $s$  is an input sequence in  $\Sigma^+$ . Now the question is, *how to find the connection between such query results and Mealy machines?*

Generally, Mealy machines are composed of *states* and *transitions*. Provided with a Mealy machine  $m$ , any input sequence in  $m.I^+$  leads to a unique state. Such a sequence is called an *access sequence* of this state. Particularly, we use  $\varepsilon$ , the *empty sequence*, to denote the initial state. Similarly, if a blackbox model has the same behavior with a Mealy machine, we can use access sequences to label its states and, in the same way, input actions to label its transitions. For example, we consider a Mealy machine which is staying in some state labelled by  $s$ . With an action  $a$  provided, the Mealy machine will jump to a new state labelled by  $s' = sa$  and generate an output  $mq(sa)$ .

Different access sequences may lead to similar states. For example, a 2-Timer channel will dispense any accepted data item in 2 time units. After that, the state of this channel is exactly same as its initial state. Here *suffixes* are used to distinguish different states as shown in the following definition, where  $D$  is initialized as  $\mathcal{I}$ . Under such a suffix set, two states are considered different only when they have different one-step behaviors.

**Definition 6 (Equivalence under Suffixes):** provided with two access sequences  $s_1, s_2$  and a set of suffixes  $D \subset \mathcal{I}^+$ , we

say the corresponding states of  $s_1$  and  $s_2$  are equivalent under  $D$  (denoted as  $s_1 \sim_D s_2$ ) iff  $\forall d \in D, mq(s_1 d) = mq(s_2 d)$ .

Since all this *inputs* and *outputs* can be observed outside the model, we can use the same notations to describe a black box as a hypothetical Mealy machine. Following this idea, *Observation Tables*, proposed in [1], shows a tabular form of such hypothetical models. In observation tables, rows represent states with label of their access sequences and columns are labelled with suffixes. Successors of a state  $s$  is, intuitively, denoted as extended access sequences  $sa$  where  $a$  is a single input action. A cell with column-label  $d$  and row-label  $s$  should contain the value  $mq(sd)$ . Algorithm 1 shows how to build such an observation table.

**Algorithm 1:** CloseTable

**Input:** Oracle interface  $mq$ , Input actions  $\mathcal{I}$ , suffix set  $D$   
**Output:** Observation table  $obs$

```

1  $obs$  initialized as empty;
2  $unclosed = \{\varepsilon\}$ ;
3 repeat
4    $next = \{st | \forall s \in unclosed, \forall t \in \mathcal{I};$ 
5      $append\ unclosed\ to\ obs;$ 
6    $unclosed = \{seq \in next | \forall a \in obs, seq \not\sim_D a\}$ ;
7 until  $unclosed = \emptyset$ ;
8 return  $obs$ ;
```

Taking a 2-Timer channel as an example, we now illustrate how this algorithm works. We assume that the source end of this 2-Timer channel is  $A$ , the sink end is  $B$  and the alphabet is  $\{a\}$ . We briefly denote  $\{A : a, B : \oslash\}$  and other input/output actions in form of  $a, \oslash$ .

Firstly,  $obs$  is initialized as empty and the empty sequence  $\varepsilon$  is pushed in  $unclosed$ . Then we will explore all the access sequences in  $unclosed$  and calculate its successors. Here the  $unclosed$  set consists of access sequences that has no equivalent fellow with less length. For example, figure 7 shows the  $obs$  of a 2-Timer channel after the first iteration. The five successors of  $\varepsilon$  is presented as the bottom part of the table, where four are equivalent with  $\varepsilon$  but  $a, \oslash$  is not. Therefore, we take  $a, \oslash$  as a brand-new state and all of its successors need further exploration (see in the hypothesis Mealy machine presented in the following figure).

	$a, \triangleright$	$\perp, \triangleright$	$a, \oslash$	$\perp, \oslash$	$T$	AccessSequence
$\varepsilon$	$\ominus$	$\ominus$	$\emptyset$	$\emptyset$	$\emptyset$	$\varepsilon$
$a, \triangleright$	$\ominus$	$\ominus$	$\emptyset$	$\emptyset$	$\emptyset$	$\varepsilon$
$\perp, \triangleright$	$\ominus$	$\ominus$	$\emptyset$	$\emptyset$	$\emptyset$	$\varepsilon$
$a, \oslash$	$\ominus$	$\ominus$	$\ominus$	$\emptyset$	$\emptyset$	NO MATCH
$\perp, \oslash$	$\ominus$	$\ominus$	$\emptyset$	$\emptyset$	$\emptyset$	$\varepsilon$
$T$	$\ominus$	$\ominus$	$\emptyset$	$\emptyset$	$\emptyset \varepsilon$	$\varepsilon$

Fig. 7. Observation Table and Corresponding Hypothesis

In the second iteration, we explore the successors of  $a, \oslash$ . Fortunately we find that every successor has a shorted equivalence fellow (*itself*). Consequently, the algorithm terminates

with a closed hypothesis where all the unclosed edge shown in Figure 7 turn into self-loop.

	$a, \triangleright$	$\perp, \triangleright$	$a, \oslash$	$\perp, \oslash$	$T$	AccessSequence
$\varepsilon$	$\ominus$	$\ominus$	$\emptyset$	$\emptyset$	$\emptyset$	$\varepsilon$
$a, \triangleright$	$\ominus$	$\ominus$	$\emptyset$	$\emptyset$	$\emptyset$	$\varepsilon$
$\perp, \triangleright$	$\ominus$	$\ominus$	$\emptyset$	$\emptyset$	$\emptyset$	$\varepsilon$
$a, \oslash$	$\ominus$	$\ominus$	$\ominus$	$\emptyset$	$\emptyset$	$a, \oslash$
$\perp, \oslash$	$\ominus$	$\ominus$	$\emptyset$	$\emptyset$	$\emptyset$	$\varepsilon$
$T$	$\ominus$	$\ominus$	$\emptyset$	$\emptyset$	$\emptyset$	$\varepsilon$
$a, \oslash - a, \triangleright$	$\ominus$	$\ominus$	$\ominus$	$\emptyset$	$\emptyset$	$a, \oslash$
$a, \oslash - \perp, \triangleright$	$\ominus$	$\ominus$	$\ominus$	$\emptyset$	$\emptyset$	$a, \oslash$
$a, \oslash - a, \oslash$	$\ominus$	$\ominus$	$\ominus$	$\emptyset$	$\emptyset$	$a, \oslash$
$a, \oslash - \perp, \oslash$	$\ominus$	$\ominus$	$\ominus$	$\emptyset$	$\emptyset$	$a, \oslash$
$a, \oslash - T$	$\ominus$	$\ominus$	$\ominus$	$\emptyset$	$\emptyset$	$a, \oslash$

Fig. 8. A Closed Observation Table

### B. Counter-Examples' Analysis

Apparently, the closed hypothesis presented in Figure 8 is not equal to the 2-Timer channel. It's easy to find a counter-example  $s = a, \oslash - T - T - \perp, \triangleright$  where  $mq(s) = a$  while according to the hypothesis, the result should have been  $\emptyset$ . In this section, we show how to find and analyze counter-examples using the method in [19].

Firstly, we give a formal definition of *counter-examples*. With a observation table  $obs$  and a sequence  $s$ , we use  $acc(s)$  to denote the access sequence of  $obs \xrightarrow{s}$  and  $hq(s)$  to denote the corresponding output.

**Definition 7 (Counter-Example):** We use  $hq$  to denote execution result of the hypothetical Mealy machine, sequence  $s$  is called a counter-example iff  $mq(s) \neq hq(s)$ .

Now we consider the reason that lead to existence of counter examples. Since the suffix set  $D$  is used to distinguish different states, the existence of counter-examples shows that the current suffix set  $D$  is not powerful enough to distinguish all different states.

Considering a counter example  $s$ , we have

$$\text{MaxSuffix}(s) = \max_d \{s = s'd, mq(acc(s')d) \neq mq(s)\}$$

Since  $mq(acc(s)) = hq(s) \neq mq(s)$  and  $mq(s) = mq(s)$ , the existence of such  $d$  is easy to prove. After appending  $d$  in  $D$ , we can rebuild a closed hypothesis where  $s$  is no longer a counter-example.

The whole  $L^*$  algorithm can be concluded as follows.

## V. EXPERIMENTS

Both *Reo Coordination Models* and *Adapted  $L^*$  Algorithm* are implemented in Golang [10], which is widely known for its elegant design and impressive efficiency. Moreover, with a CSP-style [11] cocurrency model, Golang shares many similar ideas with Reo, and in turn makes our implementation much more natural.

All the following experiments are coded under Golang 1.2.1 and executed on a laptop with 8GB of RAM and a Core i7-3630 CPU. The source code is available at <https://github.com/liyi-david/reo-learn>.

**Algorithm 2: L\***

**Input:** Oracle interface  $mq, eq$ , Input actions  $\mathcal{I}$

**Output:** Observation table  $obs$

```

1  $D = I$ ;
2 repeat
3    $obs = \text{CloseTable}(mq, \mathcal{I}, D)$ ;
4    $ce = eq(obs)$ ;
5   if  $ce! = \text{true}$  then
6      $D.append(\text{MaxSuffix}(s))$ ;
7 until  $ce = \text{true}$ ;
8 return  $obs$ ;

```

### A. Case Study

A simple example of timed connector is presented to show how L\* works.

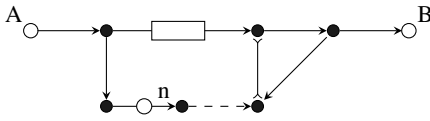


Fig. 9. Expiring FIFO1 Channel (ExpFIFO1)

Informally speaking, an expiring FIFO1 channel with *time-out*  $n$  is able to accept a data item and stored it in the buffer cell for  $n$  time units. If a read operation on  $B$  is performed within  $n$  time units, it will obtain the data item successfully and clear the buffer. However, the data item would be dropped if no read operation comes.

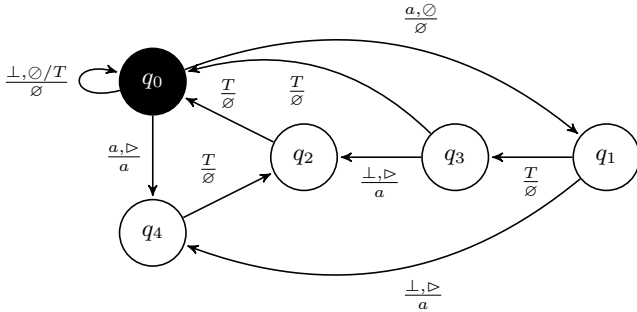


Fig. 10. Learn Result of The ExpFIFO1 where  $n = 2, \Sigma = \{a\}$

Figure 10 shows the learning result of this example. To simplify the graph, we ignore the all the trivial transitions  $\frac{\perp, \odot}{\emptyset}$  and block transitions. More details of this case can be found in our *github repo*.

### B. Performance Optimization

As a well-known learning algorithm, L\* has proved its efficiency in models without time. However, when dealing with timed connectors, the algorithm failed to meet our expectation.

As shown in Table I, time consumption mainly comes from membership queries. With time involved, every single membership query takes a lot of time inevitably. After reviewing

TABLE I  
TIME-COST ANALYSIS

	FIFO	Alternator	Gate
Membership Query(s)	41.571	126.468	169.161
Hypothesis Query(s)	0.001	0.003	0.004
Total Time(s)	41.715	165.114	247.098
Membership Query(%)	99.6	76.6	68.5

our algorithm, we found that simulations on similar sequences were invoked frequently:

- When constructing *Obs* tables, there are lots of redundant calls to membership queries. For example, a sequence with prefix 'aa' and suffix 'b' is exactly same as another one with prefix 'a' and suffix 'ab'.
- Simulation on mealy machines can provide multi-step output. Consequently, if we have simulated an 'abc' sequence, there's no reason to perform simulation on an 'ab' sequence.

If previous simulation results are stored in a well-maintained cache, the time-cost in simulation process could be reduced significantly. In this work, we use a multiway tree to buffer these results.

TABLE II  
REDUCTION OF MEMBERSHIP QUERIES

	FIFO	Alternator	Gate
Original Algorithm	93	880	1034
Cached Algorithm	90	725	707
Reduction Rate	3.2%	21.4%	31.6%

With cache applied, we have made considerable reduction on the calls of membership queries. The results can be found in Table II.

### C. An Example of the Reo Package

Our implementation in Golang is well-prepared not only for academic use but also for practical concurrent programming. The following code shows how to compose an alternator connector (see in Figure 4) in Golang. An intact version of this example can be found in our *github repo*.

```

1 func alternator(A, B, C Port) {
2   M := MakePorts(6)
3
4   // definition of channels
5   go ReplicatorChannel(A, M[0], M[1])
6   go ReplicatorChannel(B, M[2], M[3])
7   go MergerChannel(M[4], M[5], C)
8
9   go SyncdrainChannel(M[1], M[2])
10  go SyncChannel(M[0], M[4])
11  go FifoChannel(M[3], M[5])
12 }

```

Provided with the source and sink nodes, *alternator* function creates a series of basic channels and mixed nodes (named *Port*) to serve as the alternator connector we need. Now we

can activate the components and using *alternator* function to coordinate them.

```

1 A,B,C := MakePorts(3)
2 alternator(A, B, C)
3
4 go sender(A, "MSG_A")
5 go sender(B, "MSG_B")
6 go monitor(C)

```

In this case, *sender* are goroutines (basic parallel units in Golang) that keep sending certain messages to some given port (A and B). A *monitor* keep trying to read data items from the sink end C and print them on the screen. Finally, we have an interleaved sequence of “MSG\_A” and “MSG\_B”.

## VI. CONCLUSION AND FUTURE WORK

In this paper, we come up with an approach to learn timed connectors from blackboxes. We define the *parameterized Mealy machine* to serve as an operational semantics of Reo. Then we describe a set of basic channels with parameterized Mealy machine, which can be used to construct complex connectors by means of the production operator and link operator. Besides, we show that every connector can be transformed into a Mealy machine with time action T, and such Mealy machines can be extracted by an improved version of well-known active learning algorithm L\*.

The original L\* algorithm run into the bottleneck when time is involved. By a tree-style cache, we make significant reduction on membership queries and, in turn, improve the performance of learning algorithm. As a by-product, we also encapsulate the Reo connector as a distributable package which could contribute to concurrent programming.

Our future works would mainly focus on better support of dense time. To describe dense time behavior, the Mealy machine model needs a lot of changes instead of a simple T action. Besides, we will also try to improve our algorithm to handle non-deterministic behaviors. This is not a brand new topic [20], but we believe that there is still room for improvement.

## REFERENCES

- [1] D. Angluin. Learning regular sets from queries and counterexamples. *Inf. Comput.*, 75(2):87–106, 1987.
- [2] F. Arbab. Reo: a channel-based coordination model for component composition. *Mathematical Structures in Computer Science*, 14(3):329–366, 2004.
- [3] F. Arbab, C. Baier, F. S. de Boer, and J. J. M. M. Rutten. Models and temporal logics for timed component connectors. In *2nd International Conference on Software Engineering and Formal Methods (SEFM 2004)*, 28-30 September 2004, Beijing, China, pages 198–207. IEEE Computer Society, 2004.
- [4] C. Baier, M. Sirjani, F. Arbab, and J. J. M. M. Rutten. Modeling component connectors in reo by constraint automata. *Sci. Comput. Program.*, 61(2):75–113, 2006.
- [5] X. Chen, J. Sun, and M. Sun. A hybrid model of connectors in cyber-physical systems. In S. Merz and J. Pang, editors, *Formal Methods and Software Engineering - 16th International Conference on Formal Engineering Methods, ICFEM 2014, Luxembourg, Luxembourg, November 3-5, 2014. Proceedings*, volume 8829 of *Lecture Notes in Computer Science*, pages 59–74. Springer, 2014.
- [6] E. M. Clarke, O. Grumberg, S. Jha, Y. Lu, and H. Veith. Counterexample-guided abstraction refinement. In E. A. Emerson and A. P. Sistla, editors, *Computer Aided Verification, 12th International Conference, CAV 2000, Chicago, IL, USA, July 15-19, 2000, Proceedings*, volume 1855 of *Lecture Notes in Computer Science*, pages 154–169. Springer, 2000.
- [7] W. Daelemans. Colin de la higuera: Grammatical inference: learning automata and grammars - cambridge university press, 2010, iv + 417 pages. *Machine Translation*, 24(3-4):291–293, 2010.
- [8] H. George. A method for synthesizing sequential circuits. *Bell System Technical Journal*, 34(5):1045–1079, 1955.
- [9] N. S. Gill. Reusability issues in component-based development. *ACM SIGSOFT Software Engineering Notes*, 28(4):4, 2003.
- [10] Google. Google go. <https://golang.org/>.
- [11] C. A. R. Hoare. *Communicating Sequential Processes*. Prentice-Hall, 1985.
- [12] S. Kemper. Sat-based verification for timed component connectors. *Sci. Comput. Program.*, 77(7-8):779–798, 2012.
- [13] S. Li, X. Chen, Y. Wang, and M. Sun. A framework for off-line conformance testing of timed connectors. In *2015 International Symposium on Theoretical Aspects of Software Engineering, TASE 2015, Nanjing, China, September 12-14, 2015*, pages 15–22. IEEE Computer Society, 2015.
- [14] S. Meng. Connectors as designs: The time dimension. In T. Margaria, Z. Qiu, and H. Yang, editors, *Sixth International Symposium on Theoretical Aspects of Software Engineering, TASE 2012, 4-6 July 2012, Beijing, China*, pages 201–208. IEEE Computer Society, 2012.
- [15] S. Meng and F. Arbab. On resource-sensitive timed component connectors. In M. M. Bonsangue and E. B. Johnsen, editors, *Formal Methods for Open Object-Based Distributed Systems, 9th IFIP WG 6.1 International Conference, FMOODS 2007, Paphos, Cyprus, June 6-8, 2007, Proceedings*, volume 4468 of *Lecture Notes in Computer Science*, pages 301–316. Springer, 2007.
- [16] P. Prabhakar, P. S. Duggirala, S. Mitra, and M. Viswanathan. Hybrid automata-based CEGAR for rectangular hybrid systems. *Formal Methods in System Design*, 46(2):105–134, 2015.
- [17] H. Raffelt and B. Steffen. Learnlib: A library for automata learning and experimentation. In L. Baresi and R. Heckel, editors, *Fundamental Approaches to Software Engineering, 9th International Conference, FASE 2006, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2006, Vienna, Austria, March 27-28, 2006, Proceedings*, volume 3922 of *Lecture Notes in Computer Science*, pages 377–380. Springer, 2006.
- [18] B. Settles. Active learning literature survey. *University of Wisconsin, Madison*, 52(55-66):11, 2010.
- [19] B. Steffen, F. Howar, and M. Merten. Introduction to active automata learning from a practical perspective. In M. Bernardo and V. Issarny, editors, *Formal Methods for Eternal Networked Software Systems - 11th International School on Formal Methods for the Design of Computer, Communication and Software Systems, SFM 2011, Bertinoro, Italy, June 13-18, 2011. Advanced Lectures*, volume 6659 of *Lecture Notes in Computer Science*, pages 256–296. Springer, 2011.
- [20] M. Volpato and J. Tretmans. Approximate active learning of nondeterministic input output transition systems. *ECEASST*, 72, 2015.