Reasoning about Connectors in Coq

Xiyue Zhang, Weijiang Hong, Yi Li and Meng Sun

Department of Informatics and LMAM, School of Mathematical Sciences,
Peking University
{zhangxiyue,wj.hong,liyi_math,sunm}@pku.edu.cn

Abstract. Reo is a channel-based exogenous coordination model in which complex coordinators, called connectors, are compositionally built out of simpler ones. In this paper, we present a new approach to model connectors in Coq which is a proof assistant based on higher-order logic and λ-calculus. The model reflects the original structure of connectors simply and clearly. In our framework, basic connectors (channels) are interpreted as axioms and composition operations are specified as inference rules. Furthermore, connectors are interpreted as logical predicates which describe the relation between inputs and outputs. With such definitions provided, connector properties, as well as equivalence and refinement relations between different connectors, can be naturally formalized as *goals* in Coq and easily proved using pre-defined *tactics*.

Keywords: Coordination language, Reo, Coq, Reasoning

1 Introduction

Modern software systems are typically distributed over large networks of computing devices, and usually the components that comprise a system do not exactly fit together as pieces of a jigsaw puzzle, but leave significant interfacing gaps that must somehow be filled with additional code. Compositional coordination models and languages provide a formalization of the "glue code" that interconnects the constituent components and organizes the mutual interactions among them in a distributed processing environment, and played a crucial role for the success of component-based systems in the past decades.

As an example, Reo [3], which is a channel-based model for exogenous coordination, offers a powerful language for implementation of coordinating component connectors. Connectors provide the protocols that control and organize the communication, synchronization and cooperation among the components that they interconnect. Primitive connectors, called *channels* in Reo, can be composed to build complex connectors. Reo has been successfully applied in different application domains, such as service-oriented computing and bioinformatics [7, 17]. In recent years, verifying the correctness of connectors is becoming a critical challenge, especially due to the advent of Cloud computing technologies. The rapid growth of size and complexity of the computing infrastructures has made it more difficult to model and verify connector properties, and thus leads to less confidence on the correctness of connectors.

Several works have been done for formal modeling and verifying connectors. An operational semantics for Reo using Constraint Automata (CA) was provided by Baier et al. [6], and later the symbolic model checker Vereofy [5] was developed, which can be used to check CTL-like properties. Besides, one attractive approach is to translate from Reo to other formal models such as Alloy [11], mCRL2 [13], UTP [2, 18], etc., which makes it possible to take advantage of existing verification tools.

In this paper, we aim to provide an approach to formally modeling and reasoning about connectors using Coq. The basic idea of our approach is to model the behavior of a connector by representing it as a logical predicate which describes the relation among the timed data streams on the input and output nodes, and to reason about connectors' properties, as well as the equivalence and refinement relations between connectors, by using proof principles and tactics in Coq. Compared with existing approaches for verifying connectors' properties [5,12,13], using Coq is especially helpful when we take infinite behavior into consideration. The coinductive proof principle makes it possible to prove connectors' properties easily while it is difficult (sometimes impossible) for other approaches (like model checking) because of the huge (or maybe infinite) number of states.

This is not a brand new idea, as we have already provided a solution for modeling Reo in Coq in [14], where connectors are represented in a constructive way, and verification is essentially based on simulations. We do believe that the approach in this paper is reasonably different from its predecessor where Coq seldom shows its real power. To be more specific, our new work has its certain advantanges comparing with [14] in the following aspects:

- **Modeling Method:** We use axioms to describe basic channels and their composition operations, which is more natural on a proof-assistant platform than the simulation-based approach in [14].
- **Expression Power:** Any valid Coq expression can be used to depict properties, which is obviously more powerful than just using LTL formulas in [14]. Furthermore, support for continuous time behavior is also possible in our approach in this paper.
- **Refinement and Equivalence Checking:** In our framework, equivalence and refinement relations can be proved among different connectors, while the previous one is not capable of either equivalence or refinement checking.

The paper is organized as follows: After this general introduction, we briefly summarize Reo and Coq in Section 2. Section 3 shows the notion of timed data streams and some pre-defined auxiliary functions and predicates. Section 4 presents the formal modeling of basic channels and operators, as well as complex connectors. Section 5 shows how to reason about connector properties and equivalence (or refinement) relations in our framework. In Section 6, we concludes with some further research directions. Full source codes can be found at [1] for further reference.

2 Preliminaries

In this section, we provide a brief introduction to the coordination language Reo and Coq.

2.1 The Coordination Model Reo

Reo is a channel-based exogenous coordination model wherein complex coordinators, called connectors, are compositionally built out of simpler ones [3]. Further details about Reo and its semantics can be found in [3, 4, 6]. The simplest connectors are channels with well-defined behavior such as synchronous channels, FIFO channels, etc. Each channel in Reo has exactly two directed ends, with their own identities. There are two types of channel ends: source ends and sink ends. A source channel end accepts data into the channel. A sink channel end dispenses data out of the channel.

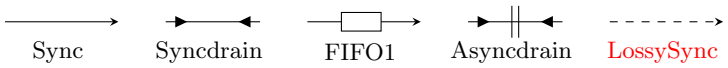


Fig. 1. Five types of basic channels.

The graphical notations of some basic channels are presented in Fig. 1, and their behavior can be interpreted as follows:

- **Sync:** a synchronous channel with one source end and one sink end. The pair of I/O operations on its two ends can succeed only simultaneously.
- **SyncDrain:** a synchronous channel which has two source ends. The pair of input operations on its two ends can succeed only simultaneously. All data items written to this channel are lost.
- **FIFOn:** an asynchronous channel with one source end and one sink end, and a bounded buffer with capacity n. It can accept data items from its source end. The accepted data items are kept in the internal buffer, and dispensed to the sink end in FIFO order. Especially, the FIFO1 channel is an instace of FIFOn where the buffer capacity is 1.
- **AsyncDrain:** an asynchronous channel which has two source ends. The channel guarantees that the operations on its two ends never succeed simultaneously. All data items written to this channel are lost.
- **LossySync:** a synchronous channel with one source end and one sink end. The source end always accepts all data items. If there is no matching I/O operation on the sink end of the channel at the time that a data item is accepted, then the data item is lost; otherwise, the channel transfers the data item exactly the same as a Sync channel, and the I/O operation at the sink end succeeds.

Complex connectors are constructed by composing simpler ones via the join and hiding operations. Channels are joined together in nodes. The set of channel

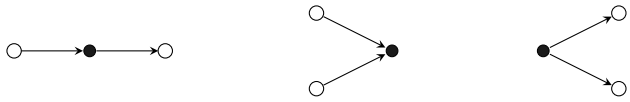


Fig. 2. Operations of channel composition.

ends coincident on a node is disjointly partitioned into the sets of source and sink channel ends that coincide on the node, respectively. Nodes are categorized into 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 the two. The hiding operation is used to hide the internal topology of a connector. The hidden nodes can no longer be accessed or observed from outside. There are three types of operations for channel composition: flow-through, merge and replicate. Fig. 2 provides the graphical representation of these operations.

2.2 The Proof Assistant Coq

Coq [9] is a widely-used proof assistant tool, where denotational formalizations (e.g. theorem and hypothesis) and operational formalizations (e.g. functions and algorithms) are naturally integrated. Moreover, it allows the interactive construction of formal proofs. The formal language used in Coq is called *Gallina*, which provides a convenient way to define both programming statements and mathematical propositions, for example:

```
(* a variable definition *)
Variables a b: nat.
(* a simple non-recursive function *)
Definition inc(a:nat) := a + 1.
(* axioms don't have to be proved *)
Axiom inc_ax: forall c:nat, inc(c) > c.
(* theorems rely on proving *)
Theorem inc_eq: forall c:nat, inc(c) = c + 1.
Proof.
  (* interactive proving based on tactics *)
  auto.
Qed.
```

As shown in this example, there are two rather different mode in Coq's interactive shell. When we start Coq, we can write declarations and definitions in a functional-programming mode. Then, when we start a *Theorem*, or *Lemma*, Coq jumps into the proving mode. We need to write different *tactics* to reduce the proving goal and finally finish the formal proof.

Furthermore, Coq is equipped with a set of well-written standard libraries. For example, as used in this paper, *List* describes the widely-used finite list structure, *Stream* provides a co-inductive definition of infinite lists, and *Reals* defines various operations and theorems on real numbers. Usually, quite a few

lemmas and theorems are pre-defined in such libraries, making it substantially easier to prove our goals.

3 Basic Definitions

In this section, we briefly introduce the notion of timed data streams and some pre-defined auxiliary functions and predicates in Coq, which are used in the following sections for modeling connectors.

The behavior of a connector can be formalized by means of data-flows at its sink and source nodes which are essentially infinite sequences. With the help of the stream library in Coq, such infinite data-flows can be defined as *timed data streams*:

```
Definition Time := R.
Definition Data := nat.
(*Inductive Data : Set :=
    |Natdata : nat-> Data
    |Empty : Data.*)
Definition TD := Time * Data.
Variable Input : Stream TD.
Variable Output : Stream TD.
```

In our framework, time is represented by real numbers. Benefit from the completeness of real number system, we can express and carry out the effective operation of a quantity at any precision request. The continuity of the set of real numbers is sufficiently enough for our modeling approach. Also the continuous time model is more appropriate since it is very expressive and closer to the nature of time in the real world. Thus, the time sequence consists of increasing and diverging time moments. For simplicity, here we take the natural numbers as the definition of data, which can be easily expanded according to different application domains. The Cartesian product of time and data defines a TD object. We use the stream module in Coq to produce streams of TD objects.

Some auxiliary functions and predicates are defined to facilitate the representation of axioms for basic channels in Reo. This part can be extended for further use in different problems.

The terms "PrL" and "PrR" take a pair of values (a, b) that has Cartesian product type A $\times$ B as the argument and return the first or second value of the pair, respectively.

The following functions provide some judgment of time, which can make the description of axioms and theorems for connectors more concise and clear. "Teq" means that time of two streams are equal and "Tneq" has the opposite meaning. "Tle" ("Tgt") represents that time of the first stream is strictly less (greater) than the second stream. Those can be defined as follows:

```
Definition Teq (s1 s2 : Stream TD) : Prop :=
forall n : nat, PrL(Str_nth n s1) = PrL(Str_nth n s2).
```

```
Definition Tneq(s1 s2 : Stream TD) : Prop := (...).
Definition Tle (s1 s2 : Stream TD) : Prop := (...).
Definition Tgt (s1 s2 : Stream TD) : Prop := (...).
```

The judgement about equality of data is analogous to the judgement of time.

```
Definition Deq(s1 s2 : Stream TD) : Prop :=
forall n:nat, PrR(Str_nth n s1) = PrR(Str_nth n s2).
```

4 Formal Modeling of Basic Channels and Operators

In this section, we show how primitive connectors, i.e., channels, and operators for connector composition are specified in Coq and used for modeling of complex connectors. Then we can apply the tactics provided in Coq to reason about connector properties. Basic channels, which can be regarded as axioms of the whole framework, are specified as logical predicates illustrating the relation between the timed data streams of input and output. When we need to construct a more complex connector, appropriate composition operators are applied depending on the topological structure of the connector.

4.1 Formal Modeling of Basic Channels

We use a pair of predicates to describe the constraints on time and data, respectively, and their intersection to provide the complete specification of basic channels. This model offers convenience for the analysis and proof of connector properties. In the following, we present a few examples of the formal model of basic channels.

The simplest form of a synchronous channel is denoted by the Sync channel type. For a channel of the Sync type, a read operation on its sink end succeeds only if there is a write operation pending on its source end. Thus, the time and data of a stream flowing into the channel are exactly the same as the stream that flows out of the channel [1]. The Sync channel can be defined as follows in the Coq system:

```
Definition Sync (Input Output:Stream TD) : Prop :=
Teq Input Output /\ Deq Input Output.
```

A LossySync channel behaves the same as a Sync channel, except that a write operation on its source always succeeds immediately. If a compatible read or take operation is already pending on the sink of a LossySync channel, the written data item is transferred to the pending operation and both succeed. Otherwise, the write operation succeeds and the data item is lost. The LossySync channel can be defined as follows:

[1] If we use α, β to denote the data streams that flow through the channel ends of a channel and a, b to denote the time stream corresponding to the data streams, i.e., the i-th element $a(i)$ in a denotes exactly the time moment of the occurrence of $\alpha(i)$, then we can easily obtain the specifications for different channels, as discussed in [16, 18]. For example, a synchronous channel can be expressed as $\alpha = \beta \wedge a = b$.

```
Parameter LossySync: Stream TD -> Stream TD -> Prop.
Axiom LossySync_coind:
  forall Input Output: Stream TD,
  LossySync Input Output ->
    (
     (hd Output = hd Input  /\ LossySync (tl Input)(tl Output))
     \/
     LossySync(tl Input) Output
    ).
```

The channel of type SyncDrain is a synchronous channel that allows pairs of write operations pending on its two ends to succeed simultaneously. All written data items are lost. Thus, the SyncDrain channel is used for synchronising two timed data streams on its two source ends. This channel type is an important basic synchronization building block for the construction of more complex connectors. The SyncDrain channel can be defined as follows:

```
Definition SyncDrain (Input Output:Stream TD) : Prop :=
Teq Input Output.
```

AsyncDrain is analogous to SyncDrain except that it guarantees that the pairs of write operations on the two channel ends never succeed simultaneously. Similarly it only has requirements on the time of the two streams on its opposite ends, but it requires that the times of the two streams are always different. The AsyncDrain channel can be defined as follows:

```
Parameter AsyncDrain: Stream TD -> Stream TD ->Prop.
Axiom AsyncDrain_coind:
  forall Input1 Input2: Stream TD,
  AsyncDrain Input1 Input2 ->
    (~ PrL(hd Input1)  =  PrL (hd Input2) )
    /\
    ( (
      (PrL(hd Input1) < PrL (hd Input2)) /\
      AsyncDrain (tl Input1) Input2
      ) \/
      (
      (PrL(hd Input1) > PrL (hd Input2))  /\
      AsyncDrain Input1 (tl Input2)
      )
    ).
```

The channel types FIFO and FIFOn where n is an integer greater than 0 represent the typical unbounded and bounded asynchronous FIFO channels. A write to a FIFO channel always succeeds, and a write to a FIFOn channel succeeds only if the number of data items in its buffer is less than its bounded capacity n. A read or take from a FIFO or FIFOn channel suspends until the first data item in the channel buffer can be obtained and then the operation

succeeds. For simplicity, we take the FIFO1 channel as an example. This channel type requires that the time when it consumes a data item through its source end is earlier than the time when the data item is delivered through its sink end. Besides, as the buffer has the capacity 1, time of the next data item that flows in should be later than the time when the data in the buffer is delivered. We use intersection of predicates in its definition as follows:

```
Definition FIFO1(Input Output:Stream TD) : Prop :=
Tle Input Output /\ Tle Output (tl Input)
/\ Deq Input Output.
```

For a FIFO1 channel whose buffer already contains a data element e, the communication can be initiated only if the data element e can be taken via the sink end. In this case, the data stream that flows out of the channel should get an extra element e settled at the beginning of the stream. And time of the stream that flows into the channel should be earlier than time of the tail of the stream that flows out. But as the buffer contains the data element e, new data can be written into the channel only after the element e has been taken. Therefore, time of the stream that flows out is earlier than time of the stream that flows in. The channel can be represented as the intersection of several predicates as follows:

```
Definition FIFO1e(Input Output:Stream TD)(e:Data) : Prop :=
Tgt Input Output /\ Tle Input (tl Output)
/\ PrR (hd Output) = e  /\ Deq Input (tl Output).
```

Defining basic channels by intersection of predicates provides the following benefits:

- Firstly, this makes the model intuitive and concise as each predicate describes a simple order relation on time or data.
- Secondly, we can easily split predicates for proofs of different properties which can make the proving process simpler.

4.2 Formal modeling of Operators

We have just described the way to define channel types, by means of definitions in Coq. Now we start defining the composition operators for connector construction. There are three types of composition operators for connector construction, which are *flow-through*, *replicate* and *merge*, respectively.

The flow-through operator simply allows data items to flow through the junction node, from one channel to the other. We need not to give the flow-through operator a specific definition in the Coq system. For example, while we illustrate two channels *Sync(A,B)* and *FIFO1(B,C)*, a flow-through operator that acts on node B for these two channels has been achieved implicitly.

The replicate operator puts the source ends of different channels together into one common node, and a write operation on this node succeeds only if all the channels are capable of consuming a copy of the written data. Similar to

the flow-through operator, it can be implicitly represented by the structure of connectors. For example, for two channels *Sync(A,B)* and *FIFO1(C,D)*, we can illustrate *Sync(A,B)* and *FIFO1(A,D)* in Coq instead of defining a function like *rep(Sync(A,B),FIFO1(C,D))* and the replicate operator is achieved directly by renaming C with A for the FIFO1 channel.

The merge operator is more complicated. We consider merging two channels AB and CD. When the merge operator acts on these two channels, it leads to a choice of taking from the common node that delivers a data item out of AB or CD. If both channels have data items available, the choice is non-deterministic. Similar to the definition of basic channels, we define merge as the intersection of two predicates and use recursive definition here. The merge operator is defined as follows:

```
Definition xor(a b: Prop) := (a \/ b) /\ ~(a /\ b).
Parameter merge:
Stream TD -> Stream TD ->Stream TD -> Prop.
Axiom merge_coind:
  forall s1 s2 s3:Stream TD,
  merge s1 s2 s3->
  (
    (
      (PrL(hd s1) < PrL(hd s2)) ->
      ((hd s3 = hd s1)  /\ merge (tl s1) s2 (tl s3))
    ) \/ (
      (PrL(hd s1) > PrL(hd s2)) ->
      ((hd s3 = hd s2)  /\ merge s1 (tl s2) (tl s3))
    ) \/ (
      (PrL(hd s1) = PrL(hd s2)) ->
      xor  ((hd s3 = hd s1)
      /\ merge (tl s1) s2 (tl s3)) ((hd s3 = hd s2)
      /\ merge s1 (tl s2) (tl s3))
    )
  ).
```

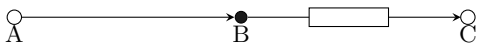


Fig. 3. A connector consisting of a Sync channel and a FIFO1 channel.

Based on the definition of basic channels and operators, more complex connectors can be constructed structurally. To show how a composite connector is constructed, we consider a simple example as shown in Fig. 3, where a FIFO1 channel is attached to the sink end of a Sync channel. Assume AB is of type Sync and BC is of type FIFO1, then we can construct the required connector by

illustrating *Sync(A,B)* and *FIFO1(B,C)*. The configuration and the functionality of the required connector can be specified using this concise method. Note that the composition operations can be easily generalized to the case of multiple nodes, where the modeling of connectors is similar. More examples can be found in Section 5.

5 Reasoning about Connectors

After modeling a connector in Coq, we can analyse and prove important properties of the connector. In this section, we give some examples to elucidate how to reason about connector properties and prove refinement/equivalence relations between different connectors, with the help of Coq.

5.1 Derivation of Connector Properties

The proof process of a property is as follows: the user states the proposition that needs to be proved, called a *goal*, then he/she applies commands called *tactics* to decompose this goal into simpler subgoals or solve it directly. This decomposition process ends when all subgoals are completely solved. In the following, we use some examples to illustrate our approach instead of giving all the complex technical details.

Example 1. We first consider the connector given in Fig. 3, which consists of two channels AB and BC with types Sync and FIFO1, respectively.

We use a and b to denote the time streams when the corresponding data streams flow into and out of the Sync channel AB, and c to denote the time stream for the data stream that flows out of the FIFO1 channel BC. Here we can see that a flow-through operation has acted on the mixed node B. The time when the stream flows into the FIFO1 channel BC is equal to the time when the stream flows out of the Sync channel AB. The following theorem states the property $a < c$ for this connector. The connector is based on the axioms Sync and FIFO1, which can be used as hypotheses for the proof of the theorem.

Theorem 1. $\forall A, B, C.\ Sync(A, B) \wedge FIFO1(B, C) \rightarrow Tle(A, C)$.

In Coq, the theorem can be proved as follows:

```
Theorem test1: forall A B C,
  Sync A B /\ FIFO1 B C -> Tle A C.
Proof.
    intros. destruct H. destruct H0.
    intro n. rewrite H. apply H0.
Qed.
```

First we give the Coq system a proposition `test1` which needs to be proved. The proposition is represented by a logical expression. Table 1 shows the detailed proving steps and the feedback that the Coq system provides during the proof.

Table 1. Steps and feedbacks for proving Theorem 1

Step	Feedback
Theorem test1: *forall A B C,* *Sync A B → FIFO1 B C* *→ Tle A C.*	1 subgoal: *forall A B C,* *Sync A B → FIFO1 B C → Tle A C*
intros	1 subgoal: *Tle A C* *H : Sync A B; H0 : FIFO1 B C*
destruct *H*	1 subgoal: *Tle A C* *H : Teq A B; H1 : Deq A B;* *H0 : FIFO1 B C*
destruct *H0*	1 subgoal: *Tle A C* *H : Teq A B; H1 : Deq A B;* *H0 : Tle B C; H2 : Tle C (tl B) ∧ Deq B C*
intro *n*	1 subgoal: *PrL (Str_nth n A) < PrL (Str_nth n C)* *H : Teq A B; H1 : Deq A B; H0 : Tle B C;* *H2 : Tle C (tl B)∧ Deq B C*
rewrite *H*	1 subgoal: *PrL (Str_nth n B) < PrL (Str_nth n C)* *H : Teq A B; H0 : Deq A B; H1 : Tle B C;* *H2 : Tle C (tl B)∧Deq B C; n : nat*
apply *H0*	No more subgoals

The advantages of using intersection of logical predicates to describe basic channels have emerged while proving this example. After constructing the new connector, we use "intros" to split conditions and conclusions. Then we can use "destruct" to obtain the conditions for time and data separately, and make the proving procedure much more convenient. Once the concrete conditions are obtained, using "intro" contributes to comparing each time point in a sequence element by element. Then by using "rewrite" *H*, we can make the proof a step forward with known conditions of the comparison of time a and b, and finally by "apply" *H0* we can prove the goal. This is the implementation for reasoning about the constructed connector. Note that proper selection of strategies and tactics is essential for the proof of connector properties.

Example 2. In this example, we show a more interesting connector named *alternator* which consists of three channels AB, AC and BC of type Syncdrain, FIFO1 and Sync, respectively. With the help of this connector, we can get data from node B and A alternatively at node C. By using the axioms for the basic channels and operators of composition, we can get the connector as shown in Figure 4(b). The two channels AC and BC are merged together at node C. Before the merge operation, the connector's structure is as shown in Figure 4(a), which is useful in the reasoning about the alternator.

We first introduce some lemmas to facilitate the proof.

```
Lemma transfer_eq : forall s1 s2 s3 : Stream TD,
((Teq s1 s2) /\ (Teq s2 s3)) -> (Teq s1 s3).
```

```
Lemma transfer_eqtl : forall s1 s2 : Stream TD,
(Teq s1 s2) -> (Teq tl s1)) (tl s2)).
Lemma transfer_leeq : forall s1 s2 s3 : Stream TD,
((Tle s1 s2) /\ (Teq s2 s3)) -> (Tle s1 s3).
Lemma transfer_hdle : forall s1 s2 : Stream TD,
(Tle s2 s1) -> (PrL (hd s1) > PrL (hd s2)).
```

(a) Before merge (b) After merge

Fig. 4. Alternator

Here the replicate operation has been applied twice for the alternator: node A becomes the common source node of *Syncdrain (A,B)* and *FIFO1(A,C1)*, and node B becomes the common source node of *Syncdrain(A,B)* and *Sync(B,C2)*. Let the time streams when the data streams flow into the two source nodes A and B be denoted by a and b, and the time streams when the data streams flow out of the channels *FIFO1(A,C1)* and *Sync(B,C2)* be denoted by *c1* and *c2*, respectively. Theorem 2 specifies the property $c2 < c1 \wedge c1 < tl(c2)$ of the connector in Fig. 4(a). The connector is based on the axioms Sync, Syncdrain and FIFO1. These three corresponding axioms are used as hypotheses for the proof of this theorem.

Theorem 2 (subtest). $\forall A, B, C1, C2.$

$$SyncDrain(A, B) \wedge FIFO1(A, C1) \wedge Sync(B, C2) \rightarrow$$
$$Tle(C2, C1) \wedge Tle(C1, tl(C2))$$

In Coq, the theorem can be proved as follows. Note that the formalizm is slightly different from the previous one. By the *section* environment, Coq is able to encapsulate hypothesises as assumptions of the theorem. So the two definitions are exactly equivalent.

```
Section Alt.
Hypothesis D1: SyncDrain A B.
Hypothesis D2: FIFO1 A C1.
Hypothesis D3: Sync B C2.
Theorem subtest:
   (Tle C2 C1)  /\  (Tle C1 (tl C2)).
```

After constructing the connector in Fig. 4(a), we use "destruct" to obtain the conditions for time and data, respectively. Since the goal we are going to prove is

an intersection of logical predicates, we use "split" to obtain the single subgoals represented by logical predicates. Besides, "intros" contributes to comparing each data in a sequence element by element. Then "rewrite" and "apply" are used similarly for multiple times until the goal is proved finally. Concrete proof steps and feedbacks are specified in Appendix A.

The proof of Theorem 2 for the connector in Fig. 4(a) can be used to simplify the proof for the following property of alternator.

An additional hypothesis is needed for the proof of alternator which merges *C1* and *C2* into a common node *C*. Based on the three hypotheses for channels and the additional hypothesis, the theorem of alternator is presented as the following proposition which needs to be proved:

```
Hypothesis D4: merge C1 C2 C.
Theorem test:
hd(C) = hd(C2) /\ merge C1 (tl C2) (tl C).
Proof.
  destruct subtest.  (* ... *)
```

Here we only present the first step which shows how a proven theorem can be applied in another proof and omit the full details because of the page limitation. And it greatly simplifies the process of proving the property of alternator.

5.2 Refinement and Equivalence

A refinement relation between connectors which allows us to systematically develop connectors in a step-wise fashion, may help to bridge the gap between requirements and the final implementations. The notion of refinement has been widely used in different system descriptions. For example, in data refinement [8], the 'concrete' model is required to have *enough redundancy* to represent all the elements of the 'abstract' one. This is captured by the definition of a surjection from the former into the latter (the *retrieve map*). If models are specified in terms of pre and post-conditions, the former are weakened and the latter strengthened under refinement [10]. In process algebra, refinement is usually discussed in terms of several 'observation' preorders, and most of them justify transformations entailing *reduction of nondeterminism* (see, for example, [15]). For connectors, the refinement relation can be defined as in [18], where a proper refinement order over connectors has been established based on the implication relation on predicates.

Here we adopt the definition of refinement in [18]. Two connectors are equivalent if each one of them is a refinement of the other. In the following, we show two examples of such connector refinement and equivalence relations.

Example 3 (Refinement). Taking the two connectors in Fig. 5 into consideration, connector Q is a refinement of connector P (denoted by $P \sqsubseteq Q$).

We have mentioned that newly constructed connectors can be specified as theorems. Given arbitrary input timed data stream at node A and output timed data streams at nodes C, D, essentially connector Q is a refinement of another

Fig. 5. Example of connector refinement

connector P only if the behavior property of P can be derived from *theorem* Q, i.e., the property of connector Q. Intuitively, connector P enables the data written to the source node A to be asynchronously taken out via the two sink nodes B and C, but it has no constraints on the relationship between the time of the two output events. On the other hand, connector Q refines this behavior by synchronizing the two sink nodes, which means that the two output events must happen simultaneously. To be more precise, we use b,c to denote the time streams of the two outputs and a to denote the time stream of the input. Connector P satisfies condition $a < b \wedge a < c$ and connector Q satisfies $a < b \wedge a < c \wedge b = c$.

The refinement relation can be formally defined in Coq as:

```
Theorem refinement : forall A C D,
(exists B, (FIFO1 A B) /\ (Sync B C) /\ (Sync B D)) ->
(exists E, (Sync A E) /\ (FIFO1 E C) /\ (FIFO1 E D)).
```

To prove this refinement relation, we first introduce a lemma which is frequently used in the proof.

Lemma 1 (Eq_self). $\forall A$*: Stream TD.* $A{=}A \Leftrightarrow$ *Sync (A,A).*

The lemma means that *Sync(A,A)* and $A{=}A$ can be derived from each other.

Lemma 2 (Eq). $\forall A,B$*: Stream TD. Sync (A,B)* $\rightarrow A{=}B$.

The lemma means that *Sync(A,B)* implies $A{=}B$, but if it is vice versa is not yet proved.

By using the axioms for the basic channels and the operators of composition, we can obtain the two connectors easily. In the process of constructing the connectors, the flow-through and replicate operations act once for each connector, respectively.

The concrete proof process using tactics and a flow chart describing the process are presented in Appendix B and C, respectively.

Example 4 (Equivalence). For the connector P in Example 3, we can add three more basic channels to build a new connector R which is equivalent to Q. R can be interpreted similarly based on basic channels and operators. We will omit the details for its construction here and prove the equivalence between the two connectors R and Q directly.

Equivalence relationship between the two connectors can be formalized as:

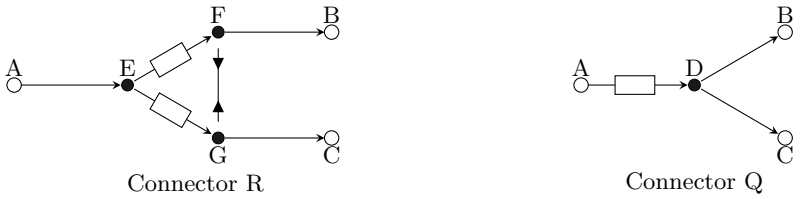


Fig. 6. Example of connector equivalence

```
Theorem equivalence: forall A B C,
  (exists E F G,
    (Sync A E) /\ (FIFO1 E F) /\ (Sync F B) /\
    (FIFO1 E G) /\ (Sync G C) /\ (SyncDrain F G)
  ) <->
  (exists D,
    (FIFO1 A D) /\ (Sync D B) /\ (Sync D C)
  ).
```

The proof of this theorem has two steps. Firstly, we prove that the new connector R is a refinement of connector Q. We hope to find an appropriate D that satisfies

$$FIFO1(A, D) \wedge Sync(D, B) \wedge Sync(D, C)$$

Similar to Example 3, we first assert $D = F$, which leads to

$$FIFO1(A, F) \wedge Sync(F, B) \wedge Sync(F, C)$$

From Lemma 2, we have $Sync(A, E)$, or $A = E$. Therefore, *FIFO1(E,F)* can be replaced by *FIFO1(A,F)*. By adopting *FIFO1(E,F)* and *FIFO1(E,G)*, we can prove that the data sequences at F and G are equal. Similarly, data sequences at C, G and F are also equal, wrt. $Sync(G, C)$.

Further according to *Sync(G,C)* and *Syncdrain(F,G)*, the time sequences at F and C are proved equal. With the combination of relations on time and data between F and C, we can draw the conclusion *Sync(F,C)*.

Up to now, we present a proof for *Sync(F,C)* and *FIFO1(A,F)* by the derivation. Besides, *Sync(F,B)* is already declared in the assumptions. Consequently, the refinement relation has been proved.

Secondly, we prove that connector Q is a refinement of connector R. We hope to find appropriate timed data streams at *E,F,G* which satisfy

$$Sync(A,E) \wedge Sync(G,C) \wedge FIFO1(E,G) \wedge Sync(F,B) \\ \wedge FIFO1(E,F) \wedge Syncdrain(F,G).$$

We can directly assume $E = A$, $F = D$ and $G = D$. Now we only need to prove *Sync(A,A)* $\wedge$ *Sync(D,C)* $\wedge$ *FIFO1(A,D)* $\wedge$ *Sync(D,B)* $\wedge$ *FIFO1(A,D)* $\wedge$ *Syncdrain(D,D)*, which can be easily derived from the assumptions.

6 Conclusion and Future Work

In this paper, we present a new approach to model and reason about connectors in the Coq system. The model naturally preserves the original structure of connectors. This also makes the connector description reasonably readable. We implement the proof of properties for connectors using identified techniques and tactics provided by Coq. Properties are defined in terms of predicates which provide an appropriate description of the relation among different timed data streams on the nodes of a connector. All the analysis and verification work are based on the logical framework where basic channels are viewed as axioms and composition operations are viewed as operators. As we can address the relation among different timed data streams, we can easily reason about temporal properties as well as equivalence and refinement relations for connectors.

As some of the benefits of this approach are inherited from Coq, our approach has also got some of its drawbacks as well. The main limitation is that the analysis needs much more tactics and techniques when the constructor becomes large. In the future work, we plan to enhance our framework by two different approaches. Firstly, we may try to encapsulate frequently-used proof patterns as new tactics, which may reduce lots of repetitive work. After that, automation methods may also help us to avoid tons of hand-written proof. For example, Coq provides several auto tactics to solve proof goals. With proper configuration, perhaps such tactics will work well in our framework. More attention is needed to precisely evaluate how expressive this way is for modeling temporal properties.

Acknowledgement

The work was partially supported by the National Natural Science Foundation of China under grant no. 61532019, 61202069 and 61272160.

Appendix

A. Steps and Feedbacks for Example 2

The following table shows the steps and feedbacks in the proof in Example 2.

B. Tactics for Refinement in Example 3

We now show the specific tactics used in the proof of refinement $P \sqsubseteq Q$ for connectors P and Q in Example 3. We need to find a timed data stream which specifies the data-flow through node E of connector P, i.e., we need to find an appropriate E that satisfies $Sync(A, E) \wedge FIFO1(E, C) \wedge FIFO1(E, D)$.

First we employ 'intros' to acquire a simpler subgoal $\exists E_0.Sync(A, E_0) \wedge FIFO1(E_0, C) \wedge FIFO1(E_0, D)$. Then we assert that $E = A$.After using 'split', we split the goal into two subgoals *Sync (A, E)* and *FIFO1 (E, C)* $\wedge$ *FIFO1 (E, D)*. And by 'rewrite' H_0 *(H_0: E = A)*, we replace the two subgoals with *Sync (A, A)* and *FIFO1 (E, C)* $\wedge$ *FIFO1 (E, D)*, respectively.

Table 2. Steps and feedback

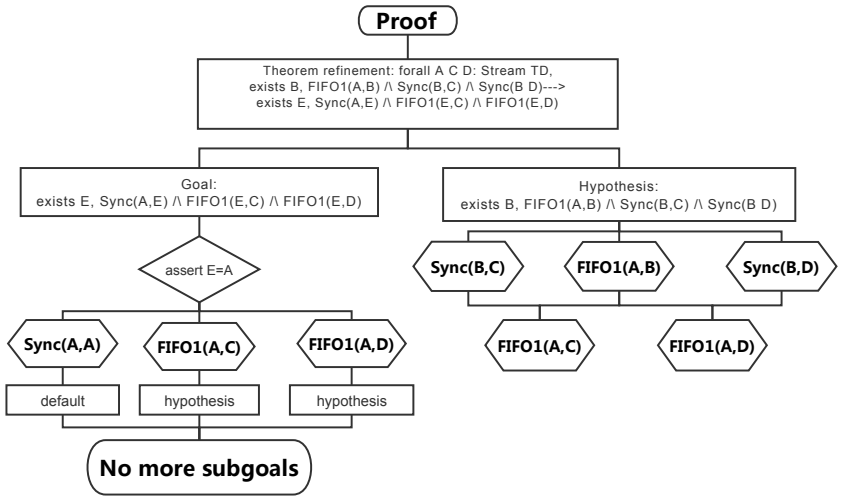
Step	Feedback
Theorem subtest	1 subgoal: *Tle C2 C1 ∧ Tle C1 (tl C2)*
destruct *D2*	1 subgoal: *Tle C2 C1 ∧ Tle C1 (tl C2)* *H : Tle A C1; H0 : Tle C1 (tl A) ∧ Deq A C1*
destruct *D3*	1 subgoal: *Tle C2 C1 ∧ Tle C1 (tl C2)* *H : Tle A C1; H0 : Tle C1 (tl A) ∧ Deq A C1;* *H1 : Teq B C2; H2 : Deq B C2*
destruct *H0*	1 subgoal: *Tle C2 C1 ∧ Tle C1 (tl C2)* *H : Tle A C1; H0 : Tle C1 (tl A);* *H3 : Deq A C1; H1 : Teq B C2; H2 : Deq B C2*
split	2 subgoals: *Tle C2 C1; Tle C1 (tl C2)* *H : Tle A C1; H0 : Tle C1 (tl A);* *H3 : Deq A C1; H1 : Teq B C2; H2 : Deq B C2*
intros *n*	2 subgoals: *PrL (Str_nth n C2) < PrL (Str_nth n C1); Tle C1 (tl C2)* *H : Tle A C1; H0 : Tle C1 (tl A); H3 : Deq A C1;* *H1 : Teq B C2; H2 : Deq B C2; n : nat*
rewrite ← *H1*	2 subgoals: *PrL (Str_nth n B) < PrL (Str_nth n C1); Tle C1 (tl C2)* *H : Tle A C1; H0 : Tle C1 (tl A); H3 : Deq A C1;* *H1 : Teq B C2; H2 : Deq B C2; n : nat*
rewrite ← *D1*	2 subgoals: *PrL (Str_nth n A) < PrL (Str_nth n C1); Tle C1 (tl C2)* *H : Tle A C1; H0 : Tle C1 (tl A); H3 : Deq A C1;* *H1 : Teq B C2; H2 : Deq B C2; n : nat*
apply *H*	1 subgoal: *Tle C1 (tl C2)* *H : Tle A C1; H0 : Tle C1 (tl A); H3 : Deq A C1;* *H1 : Teq B C2; H2 : Deq B C2*
intros *n*	1 subgoal: *Tle C1 (tl C2)* *H : Tle A C1; H0 : Tle C1 (tl A); H3 : Deq A C1;* *H1 : Teq B C2; H2 : Deq B C2*
rewrite ← *D4*	2 subgoals: *PrL (Str_nth n C1) < PrL (Str_nth n (tl B)); Teq B C2* *H : Tle A C1; H0 : Tle C1 (tl A); H3 : Deq A C1;* *H1 : Teq B C2; H2 : Deq B C2; n : nat*
rewrite ← *D5*	3 subgoals: *PrL (Str_nth n C1) < PrL (Str_nth n (tl A)); Teq A B; Teq B C2* *H : Tle A C1; H0 : Tle C1 (tl A); H3 : Deq A C1;* *H1 : Teq B C2; H2 : Deq B C2; n : nat*
apply *H0*	2 subgoals: *Teq A B; Teq B C2* *H : Tle A C1; H0 : Tle C1 (tl A); H3 : Deq A C1;* *H1 : Teq B C2; H2 : Deq B C2; n : nat*
apply *D1*	1 subgoal: *Teq B C2* *H : Tle A C1; H0 : Tle C1 (tl A); H3 : Deq A C1;* *H1 : Teq B C2; H2 : Deq B C2; n : nat*
apply *D3*	No more subgoals.

Through 'apply' Lemma 2 (*Eq*), we have $A = A$ in place of *Sync (A, A)*. Next the tactic *reflexivity* makes the subgoal $A = A$ proved directly. Up to now, the initial subgoal *Sync (A, E)* has been achieved.

Using 'split' again, the remaining unproven subgoal is split into two subgoals *FIFO1 (E, C)* and *FIFO1 (E, D)*. After destructing the precondition three times, we succeed in obtaining three hypotheses: *H: FIFO1 (A, x); H1: Sync (x, C); H2: Sync (x, D)*. Assume x = C and then using tactics *apply Eq* and *assumption*, *assertion* $x = C$ is proved easily. Meanwhile, we get hypothesis *H3:* $x = C$. Via *Rewrite* $\leftarrow$*H3*, we bring left in place of the right side of the equation *H3:* $x = C$ into *FIFO1 (E, C)* and have *FIFO1 (E, x)*. Similarly, rewrite *H0* and further we get the result *FIFO1 (A, x)* which is exactly hypothesis *H*. By using 'assumption', the second subgoal is proved already. Using substantially the same tactic steps, *FIFO1 (E, D)* can be proved. Finally, we have no more subgoals. Note that there is a new tactic 'reflexivity' used in the proof, which is actually synonymous with 'apply refl equal'. We can use it to prove that two statements are equal.

C. Flow Chart for Proving Connector Refinement

The following figure shows the flow chart for the proof steps of connector refinement in Example 3:



References

1. Package of source files. `http://www.math.pku.edu.cn/teachers/sunm/projects/reo2coq.zip`.

2. B. K. Aichernig, F. Arbab, L. Astefanoaei, F. S. de Boer, M. Sun, and J. Rutten. Fault-based test case generation for component connectors. In *Proceedings of TASE 2009*, pages 147–154. IEEE Computer Society, 2009.
3. F. Arbab. Reo: A Channel-based Coordination Model for Component Composition. *Mathematical Structures in Computer Science*, 14(3):329–366, 2004.
4. F. Arbab and J. Rutten. A coinductive calculus of component connectors. In *WADT 2002*, volume 2755 of *LNCS*, pages 34–55. Springer-Verlag, 2003.
5. C. Baier, T. Blechmann, J. Klein, S. Klüppelholz, and W. Leister. Design and verification of systems with exogenous coordination using vereofy. In *Proceedings of ISoLA 2010*, volume 6416 of *LNCS*, pages 97–111. Springer, 2010.
6. C. Baier, M. Sirjani, F. Arbab, and J. Rutten. Modeling component connectors in Reo by constraint automata. *Science of Computer Programming*, 61:75–113, 2006.
7. D. Clarke, D. Costa, and F. Arbab. Modelling coordination in biological systems. In *Proceedings of ISoLA'04*, volume 4313 of *LNCS*, pages 9–25. Springer, 2004.
8. W.-P. de Roever and K. Engelhardt. *Data Refinement: Model-Oriented Proof Methods and their Comparison*. Cambridge University Press, 1998.
9. G. Huet, G. Kahn, and C. Paulin-Mohring. The coq proof assistant a tutorial. *Rapport Technique*, 178, 1997.
10. C. B. Jones. *Systematic Software Development using VDM*. Prentice-Hall, 1990.
11. R. Khosravi, M. Sirjani, N. Asoudeh, S. Sahebi, and H. Iravanchi. Modeling and analysis of reo connectors using alloy. In *Proceedings of COORDINATION 2008*, volume 5052 of *LNCS*, pages 169–183. Springer, 2008.
12. S. Klüppelholz and C. Baier. Symbolic Model Checking for Channel-based Component Connectors. *Science of Computer Programming*, 74(9):688–701, 2009.
13. N. Kokash, C. Krause, and E. de Vink. Reo+mCRL2: A framework for model-checking dataflow in service compositions. *Formal Aspects of Computing*, 24:187–216, 2012.
14. Y. Li and M. Sun. Modeling and Verification of Component Connectors in Coq. *Science of Computer Programming*, 113(3):285–301, 2015.
15. A. W. Roscoe. *The Theory and Practice of Concurrency*. Prentice Hall, 1998.
16. M. Sun. Connectors as designs: The time dimension. In *Proceedings of TASE 2012*, pages 201–208. IEEE Computer Society, 2012.
17. M. Sun and F. Arbab. Web Services Choreography and Orchestration in Reo and Constraint Automata. In *Proceedings of SAC'07*, pages 346–353. ACM, 2007.
18. M. Sun, F. Arbab, B. K. Aichernig, L. Astefanoaei, F. S. de Boer, and J. Rutten. Connectors as Designs: Modeling, Refinement and Test Case Generation. *Science of Computer Programming*, 77(7-8):799–822, 2012.