Tasks in Modular Proofs of Concurrent Algorithms

Armando Castañeda, Aurélie Hurault, Philippe Quéinnec, Matthieu Roy

Tasks in Modular Proofs of Concurrent Algorithms

Armando Castañeda[1], Aurélie Hurault[2], Philippe Quéinnec[2(✉)], and Matthieu Roy[3,4]

[1] Instituto de Matemáticas, UNAM, Mexico City, Mexico
armando.castaneda@im.unam.mx
[2] IRIT – Université de Toulouse, Toulouse, France
{hurault,queinnec}@enseeiht.fr
[3] Laboratorio Solomon Lefschetz - UMI LaSoL, CNRS, CONACYT, UNAM, Cuernavaca, Mexico
[4] LAAS-CNRS, CNRS, Université de Toulouse, Toulouse, France
roy@laas.fr

Abstract. Proving correctness of distributed or concurrent algorithms is a mind-challenging and complex process. Slight errors in the reasoning are difficult to find, calling for computer-checked proof systems. In order to build computer-checked proofs with usual tools, such as Coq or TLA$^+$, having sequential specifications of all base objects that are used as building blocks in a given algorithm is a requisite to provide a modular proof built by composition. Alas, many concurrent objects do not have a sequential specification.

This article describes a systematic method to transform any *task*, a specification method that captures concurrent one-shot distributed problems, into a sequential specification involving two calls, `set` and `get`. This transformation allows system designers to *compose* proofs, thus providing a framework for modular computer-checked proofs of algorithms designed using tasks and sequential objects as building blocks. The Moir&Anderson implementation of *renaming* using *splitters* is an iconic example of such algorithms designed by composition.

Keywords: Formal methods · Verification · Concurrent algorithms · Renaming

1 Introduction

Fault-tolerant distributed and concurrent algorithms are extensively used in critical systems that require strict guarantees of correctness [23]; consequently, verifying such algorithms is becoming more important nowadays. Yet, proving distributed and concurrent algorithms is a difficult and error-prone task, due to the complex interleavings that may occur in an execution. Therefore, it is crucial to develop frameworks that help assessing the correctness of such systems.

A major breakthrough in the direction of systematic proofs of concurrent algorithms is the notion of *atomic* or *linearizable* objects [20]: a linearizable

https://doi.org/10.1007/978-3-030-34992-9_6

object behaves as if it is accessed sequentially, even in presence of concurrent invocations, the canonical example being the atomic register. Atomicity lets us model a concurrent algorithm as a transition system in which each transition corresponds to an atomic step performed by a process on a base object. Human beings naturally reason on sequences of events happening one after the other; concurrency and interleavings seem to be more difficult to deal with.

However, it is well understood now that several natural one-shot base objects used in concurrent algorithms cannot be expressed as sequential objects [9,16,33] providing a single operation.

An iconic example is the *splitter* abstraction [31], which is the basis of the classical Moir&Anderson renaming algorithm [31]. Intuitively, a splitter is a concurrent one-shot problem that splits calling processes as follows: whenever p processes access a splitter, at most one process obtains stop, at most $p-1$ obtain right and at most $p-1$ obtain down. Moir&Anderson renaming algorithm uses splitters arranged in a half grid to scatter processes and provide new names to processes. It is worth to mention that, since its introduction almost thirty years ago, the *renaming* problem [4] has become a paradigm for studying symmetry-breaking in concurrent systems (see, for example, [1,8]).

A second example is the *exchanger* object provided in Java, which has been used for implementing efficient linearizable elimination stacks [16,24,36]. Roughly speaking, an exchanger is a meeting point where pairs of processes can exchange values, with the constraint that an exchange can happen only if the two processes run concurrently.

Splitters and exchangers are instances of one-shot concurrent objects known in the literature as *tasks*. Tasks have played a fundamental role in understanding the computability power of several models, providing a topological view of concurrent and distributed computing [18]. Intuitively, a task is an object providing a single one-shot operation, formally specified through an input domain, an output domain and an input/output relation describing the valid output configurations when a set of processes run concurrently, starting from a given input configuration. Tasks can be equivalently specified by mappings between topological objects: an input simplicial complex (i.e., a discretization of a continuous topological space) modeling all possible input assignments, an output simplicial complex modeling all possible output assignments, and a carrier map relating inputs and outputs.

Contributions. Our main contribution is a generic transformation of any task T (with a single operation) into a sequential object S providing two operations, set and get. The behavior of S "mimics" the one of T by splitting each invocation of a process to T into two invocations to S, first set and then get. Intuitively, the set operation records the processes that are participating to the execution of the task. A process actually calls the task and obtains a return value by invoking get. Each of the operations is atomic; however, set and get invocations of a given process may be interleaved with similar invocations from other processes.

We show that these two operations are sufficient for any task, no matter how complicated it may be; since a task is a mapping between simplicial complexes,

it can specify very complex concurrent behaviors, sometimes with obscure associated operational semantics.

A main benefit of our transformation is that one can replace an object solving a task T by its associated sequential object S, and reason as if all steps happen sequentially. This allows us to obtain simpler models of concurrent algorithms using solutions to tasks and sequential objects as building blocks, leading to modular correctness proofs. Concretely, we can obtain a simple transition system of Moir&Anderson renaming algorithm, which helps to reason about it. In a companion paper [22], our model is used to derive a full and modular TLA$^+$ proof of the algorithm, the first available TLA$^+$ proof of it.

In Sect. 2, we explain the ideas in Moir&Anderson renaming algorithm that motivated our general transformation, which is presented in Sect. 3. Due to lack of space, some basic definitions, proofs and detailed constructions are omitted. They can be found in the extended version [7].

2 Verifying Moir&Anderson Renaming

We consider a concurrent system with n asynchronous processes, meaning that each process can experience arbitrarily long delays during an execution. Moreover, processes may crash at any time, i.e., permanently stopping taking steps. Each process is associated with a unique ID $\in \mathbb{N}$. The processes can access *base* objects like simple atomic read/write registers or more complex objects.

The original Moir&Anderson renaming algorithm [31] is designed and explained with splitters. Their seminal work first introduces the splitter algorithm based on atomic read/write registers and discusses its properties. Then, they describe a renaming algorithm that uses a grid of splitters. The actual implementation inlines splitters into the code of the renaming algorithm, and their proof is performed on the resulting program that uses solely read/write registers as base objects.

The Splitter Abstraction. A *splitter* [31] is a one-shot concurrent task in which each process starts with its unique ID $\in \mathbb{N}$ and has to return a value satisfying the following properties: (1) Validity. The returned value is right, downor stop. (2) Splitting. If $p \geq 1$ processes participate in an execution of the splitter, then at most $p-1$ processes obtain the value right, at most $p-1$ processes obtain the value down, at most one process obtains the value stop. (3) Termination. Every correct process (which doesn't crash) returns a value.

Notice that if a process runs solo, i.e., $p = 1$, it must obtain stop, since the splitting property holds for any $p \geq 1$.

Figure 1 contains the simple and elegant splitter implementation based on atomic read/write registers from [31] (register names have been changed for clarity). After carefully analyzing the code, the reader can convince herself that the algorithm described in Fig. 1 implements the splitter specification. The fact that the implementation is based on *atomic* registers allows us to obtain a transition system of it in which each transition corresponds to an atomic operation

```
initially CLOSED = false
operation splitter():
(01)  LAST ← my_ID;
(02)  if (CLOSED)
(03)     then return(right)
(04)     else  CLOSED ← true;
(05)           if (LAST = my_ID)
(06)           then return(stop)
(07)           else  return(down)
(08)           end if
(09)  end if.
```

Fig. 1. Implementation of a Splitter [31].

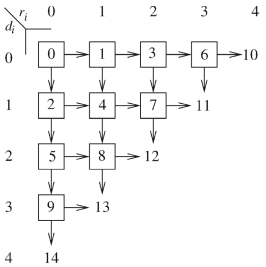

Fig. 2. Renaming using Splitters.

on an object. The benefit of this modelization is that every execution of the implementation is simply described as a sequence of steps, as concurrent and distributed systems are usually modeled (see, for example, [19,35])). Although the splitter implementation is very short and simple, its TLA$^+$ proof is long and rather complex —particularly when considering that it uses a boolean register and a plain register only—(see [22] for details).

The Renaming Problem. In the *M-renaming* task [4], each process starts with its unique ID $\in \mathbb{N}$, and processes are required to return an output name satisfying the following properties: (1) Validity. The output name of a process belongs to $[1, \ldots, M]$. (2) Uniqueness. No two processes obtain the same output name. (3) Termination. Every correct process returns an output name.

Let p be the number of processes that participate in a given renaming instance. A renaming implementation is *adaptive* if the size M of the new name space depends only on p, the number of participating processes. We have then $M = f(p)$ where $f(p)$ is a function on p such that $f(1) = 1$ and, for $2 \leq p \leq n$, $p - 1 \leq f(p-1) \leq f(p)$.

Moir&Anderson Splitter-Based Renaming Algorithm. Moir and Anderson propose in [31] a read/write renaming algorithm designed using the splitter abstraction. The algorithm is conceptually simple: for up to n processes, a set of $n(n+1)/2$ splitters are placed in a half-grid, each with a unique name, as shown in Fig. 2 for $n = 5$. Each process starts invoking the splitter at the top-left corner, following the directions obtained at each splitter. When a splitter invocation returns stop, the process returns the name associated with the splitter. We use here an adaptive version of their algorithm that allows p participating processes to rename in at most $p(p+1)/2$ names; the original solution in [31] is non-adaptive and the only difference is the labelling of the splitters in the grid.

Splitters as Sequential Objects? Although Moir&Anderson renaming algorithm is easily described in a modular way, the actual program is not modular as each splitter in the conceptual grid is replaced by an independent copy of the splitter implementation of Fig. 1. Thus, the correctness proof in [31] deals with the

State: Sets $Participants, Stop, Down, Right$
all sets are initialized to $\emptyset$

Function set(id)
  Pre-condition: id $\notin Participants$
  Post-condition: $Participants' \leftarrow Participants \cup \{\text{id}\}$
  Output: void
endFunction

Function get(id)
  Pre-condition: id $\in Participants \wedge$ id $\notin Stop, Down, Right$
  Post-condition:
    $D \leftarrow \emptyset$
    if $|Stop| = 0$ **then** $D \leftarrow D \cup \{\text{stop}\}$
    if $|Down| < |Participants| - 1$ **then** $D \leftarrow D \cup \{\text{down}\}$
    if $|Right| < |Participants| - 1$ **then** $D \leftarrow D \cup \{\text{right}\}$
    Let dec be any value in D
    if $dec =$ stop **then** $Stop \leftarrow Stop \cup \{\text{id}\}$
    if $dec =$ down **then** $Down \leftarrow Down \cup \{\text{id}\}$
    if $dec =$ right **then** $Right \leftarrow Right \cup \{\text{id}\}$
  Output: dec
endFunction

Fig. 3. An *ad hoc* specification of the Splitter.

possible interleavings that can occur, considering all read/write splitter implementations in the grid.

In the light of the simple splitter based conceptual description, we would like to have a transition system describing the algorithm based on splitters as building blocks, in which each step corresponds to a splitter invocation. Such a description would be very beneficial as it would allow us to obtain a modular correctness proof showing that the algorithm is correct as long as the building blocks are splitters, hence the correctness is independent of any particular splitter implementation.

As it is formally proved in Sect. 3, it is impossible to obtain such a transition system. The obstacle is that a splitter is inherently concurrent and cannot be specified as a sequential object with a single operation. The intuition of the impossibility is the following. By contradiction, suppose that there is a sequential object corresponding to a splitter. Since the object is sequential, in every execution, the object behaves as if it is accessed sequentially (even in presence of concurrent invocation). Then, there is always a process that invokes the splitter object first, which, as noted above, must obtain stop. The rest of the processes can obtain either down or right, without any restriction (the value obtained by the first process precludes that all obtain right or all down). However, such an object is allowing strictly fewer behaviors: in the original splitter definition it

is perfectly possible that all processes run concurrently and half of them obtain right and the other half obtain down, while none obtains stop.

The Splitter Task as a Sequential Object. One can circumvent the impossibility described above by splitting the single method provided by a splitter into two (atomic) operations of a sequential object. Figure 3 presents a sequential specification of a splitter with two operations, set and get, using a standard pre/post-condition specification style. Each process invoking the splitter, first invokes set and then get (always in that order). The idea is that the set operation first records in the state of the object the processes that are participating in the splitter, so far, and then the get operation nondeterministically produces an output to a process, considering the rules of the splitter. In Sect. 3, we formally prove that this sequential object indeed models the splitter defined above.

Proving Moir&Anderson Renaming with Splitters as Base Sequential Objects. Using the sequential specification of a splitter in Fig. 3, we can easily obtain a *generic* description of the original Moir&Anderson splitter-based algorithm: each renaming object is replaced with an equivalent sequential version of it, and every process accessing a renaming object asynchronously invokes first set and then get, which returns a direction to the process. The resulting algorithm does not rely on any particular splitter implementation, and uses only atomic objects, which allows us to obtain a transition system of it. This is the algorithm that is verified in TLA$^+$ in [22]. The equivalence between the concurrent renaming specification and the sequential set/get specification imply that the proof in [22] also proves for the original Moir&Anderson splitter-based algorithm.

3 Dealing with Tasks Without Sequential Specification

In this section, we show that the transformation in Sect. 2 of the splitter task into a sequential object with two operations, get and set, is not a trick but rather a general methodology to deal with tasks without a sequential specification. Our get/set solution proposed here is reminiscent to the *request-follow-up* transformation in [25] that allows to transform a *partial* method of a sequential object (e.g. a queue with a blocking dequeue method when the queue is empty) into two *total* methods: a total request method registering that a process wants to obtain an output, and a total follow-up method obtaining the output value, or *false* if the conditions for obtaining a value are not yet satisfied (the process invokes the follow-up method until it gets an output). We stress that the *request-follow-up* transformation [25] considers only objects with a sequential specification and is not shown to be general as it is only used for queues and stacks.

Model of Computation in Detail. We consider a standard concurrent system with n *asynchronous* processes, $p_1, \ldots, p_n$, which may *crash* at any time during an execution of the system, i.e., stopping taking steps (for more detail see for example [19,35]). Processes communicate with each other by invoking operations on shared, concurrent *base objects*. A base object can provide

Read/Write operations (also called *register*), more powerful operations, such as Test&Set, Fetch&Add, Swap or Compare&Swap, or solve a concurrent distributed problem, for example, Splitter, Renaming or Set_Agreement.

Each process follows a local state machines $A_1, \ldots, A_n$, where A_i specifies which operations on base objects p_i executes in order to return a response when it invokes a high-level operation (e.g. push or pop operations). Each of these base-objects operation invocations is a *step*. An *execution* is a possibly infinite sequence of steps and invocations and responses of high-level operations, with the following properties:

1. Each process first invokes a high-level operation, and only when it has a corresponding response, it can invoke another high-level operation, i.e., executions are *well-formed*.
2. For any invocation $inv(\langle \mathsf{opType}, p_i, input \rangle)$ of a process p_i, the steps of p_i between that invocation and its corresponding response (if there is one), are steps that are specified by A_i when p_i invokes the high-level operation $\langle \mathsf{opType}, p_i, input \rangle$.

A high-level operation in an execution is *complete* if both its invocation and response appear in the execution. An operation is *pending* if only its invocation appears in the execution. A process is *correct* in an execution if it takes infinitely many steps.

Sequential Specifications. A central paradigm for specifying distributed problems is that of a shared object X that processes may access concurrently [19,35], but the object is defined in terms of a *sequential specification*, i.e., an automaton describing the outputs the object produces when it is accessed sequentially. Alternatively, the specification can be described as (possibly infinite) prefix-closed set, $SSpec(X)$, with all sequential executions allowed by X.

Once we have a sequential specification, there are various ways of defining what it means for an execution to *satisfy* an object, namely, that it respects the sequential specification. *Linearizability* [20] is the standard notion used to identify correct executions of implementations of sequential objects. Intuitively, an execution is linearizable if its operations can be ordered sequentially, without reordering non-overlapping operations, so that their responses satisfy the specification of the implemented object. To formalize this notion we define a partial order on the completed operations of an execution E: $\mathsf{op} <_E \mathsf{op}'$ if and only if the response of op precedes the invocation of op' in E. Two operations are *concurrent* if they are incomparable by $<_E$. The execution is *sequential* if $<_E$ is a total order.

An execution E is *linearizable* with respect to X if there is a sequential execution S of X (i.e., $S \in SSpec(X)$) such that: (1) S contains every completed operation of E and might contain some pending operations. Inputs and outputs of invocations and responses in S agree with inputs and outputs in E, and (2) for every two completed operations op and op' in E, if $\mathsf{op} <_E \mathsf{op}'$, then op appears before op' in S.

Using the linearizability correctness criteria for sequential objects, we can define the set of *valid* executions for X, denoted $VE(X)$, as the set containing every execution E that consists of invocations and responses and is linearizable w.r.t. X. $VE(X)$ contains the behavior one might expect from any *building-block* implementation of X, e.g., any algorithm that implements X.

Tasks. A task is the basic distributed equivalent of a function in sequential computing, defined by a set of inputs to the processes and for each (distributed) input to the processes, a set of legal (distributed) outputs of the processes, e.g., [18].

In an algorithm designed to solve a task, each process starts with a private input value and has to eventually decide irrevocably on an output value. A process p_i is initially not aware of the inputs of other processes. Consider an execution where only a subset of $k \leq n$ processes participate; the others crash without taking any steps. A set of pairs $\sigma = \{(\mathrm{id}_1, x_1), \ldots, (\mathrm{id}_k, x_k)\}$ is used to denote the input values, or output values, in the execution, where x_i denotes the value of the process with identity id_i, either an input value or an output value. A set σ as above is called a *simplex*, and if the values are input values, it is an *input simplex*, if they are output values, it is an *output simplex*. The elements of σ are called *vertices*, and any subset of σ is a *face* of it. An *input vertex* $v = (\mathrm{id}_i, x_i)$ represents the initial state of process id_i, while an *output vertex* represents its decision. The *dimension* of a simplex σ, denoted $dim(\sigma)$, is $|\sigma| - 1$, and it is *full* if it contains n vertices, one for each process. A *complex* $\mathcal{K}$ is a set of simplexes (i.e. a set of sets) closed under containment. The dimension of $\mathcal{K}$ is the largest dimension of its simplexes, and $\mathcal{K}$ is *pure* of dimension k if each of its simplexes is a *face* of a k-dimensional simplex. In distributed computing, the simplexes (and complexes) are often *chromatic*: vertices of a simplex are labeled with a distinct process identities. The set of processes identities in an input or output simplex σ is denoted $\mathrm{ID}(\sigma)$.

A *task* T for n processes is a triple $(\mathcal{I}, \mathcal{O}, \Delta)$ where $\mathcal{I}$ and $\mathcal{O}$ are pure chromatic $(n-1)$-dimensional complexes, and Δ maps each simplex σ from $\mathcal{I}$ to a subcomplex $\Delta(\sigma)$ of $\mathcal{O}$, satisfying: (1) $\Delta(\sigma)$ is pure of dimension $dim(\sigma)$, (2) for every τ in $\Delta(\sigma)$ of dimension $dim(\sigma)$, $\mathrm{ID}(\tau) = \mathrm{ID}(\sigma)$, and (3) if σ, σ' are two simplexes in $\mathcal{I}$ with $\sigma' \subset \sigma$ then $\Delta(\sigma') \subset \Delta(\sigma)$. A task is a very compact way of specifying a distributed problem, and indeed typically it is hard to understand what exactly is the problem being specified. Intuitively, Δ specifies, for every simplex $\sigma \in \mathcal{I}$, the valid outputs $\Delta(\sigma)$ for the processes in $\mathrm{ID}(\sigma)$ assuming they run to completion, and the other processes crash initially, and do not take any steps.

As an example consider the *splitter* task [31]. Figure 4 shows a graphic description of the splitter task for three processes with IDs 1, 2 and 3. The input complex, shown at the left, consists of a triangle and all its faces. The output complex, at the right, contains all possible valid output simplexes (the triangle with all right outputs is not in the complex). The Δ function maps the input vertex with ID 1 to the output vertex $(1, \mathsf{stop})$, the input edge with IDs 1 and 2 to the complex with the bold edges in the output complex, and the

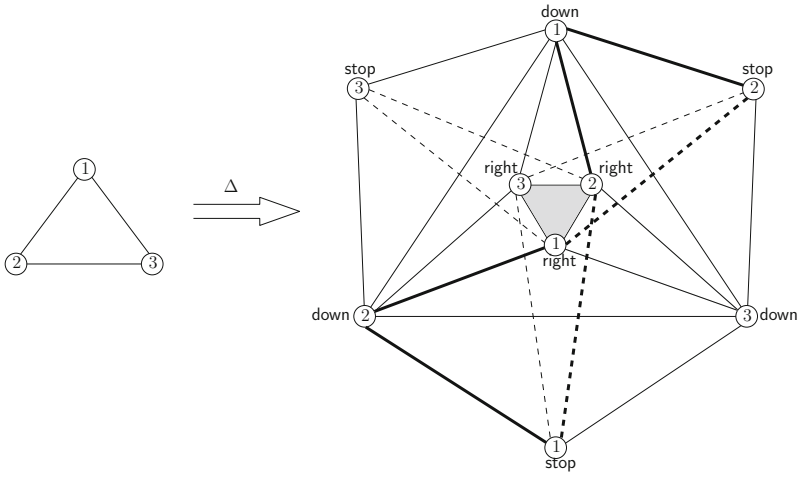

Fig. 4. The Splitter Task for Three Processes.

input triangle is mapped to the whole output complex. The rest of Δ is defined symmetrically.

Let E be an execution where each process invokes a task $\langle \mathcal{I}, \mathcal{O}, \Delta \rangle$ once. Then, σ_E is the input simplex defined as follows: (id_i, x_i) is in σ_E iff in E there is an invocation of $\mathsf{task}(x_i)$ by process id_i. The output simplex τ_E is defined similarly: (id_i, y_i) is in τ_E iff there is a response y_i to a process id_i in E. We say that E *satisfies* $(\mathcal{I}, \mathcal{O}, \Delta)$ if for every prefix E' of E, it holds that $\tau_{E'} \in \Delta(\sigma_{E'})$.

Using the satisfiability notion of tasks we can now consider the set of valid executions, $VE(T)$, for a given task $T = (\mathcal{I}, \mathcal{O}, \Delta)$: the set containing every execution E that has only invocations and responses and satisfies T. Arguably, the set $VE(T)$ contains the behavior one might expect from a *building-block* (e.g. an algorithm) that implements T.

Modeling Tasks as Sequential Objects. Intuitively, tasks and sequential specifications are inherently different paradigms for specifying distributed problems: while a task specifies what a set of processes might output when running concurrently, a sequential specification specifies the behavior of a concurrent object when accessed sequentially (and linearizability tells when a concurrent execution "behaves" like a sequential execution of the object). A natural question is if any task can be modeled as a sequential object with a single operation, namely, the object defines the same set of valid executions. A well-known example for which this is possible is the consensus distributed coordination problem that can be equivalently defined as a task or as a sequential object (see for example [19] where it is defined as an object[1] and [18] where it is defined as a task).

[1] Sometimes, for clarity or efficiency, the object is defined with two operations (in the style of the Theorem 1); however, consensus can be equivalently defined with one operation.

State: a pair (σ, τ) of input/output simplexes, initialized to $(\emptyset, \emptyset)$

Function $\mathsf{set}(\mathrm{id}_i, x_i)$
  Pre-condition: $\mathrm{id}_i \in \mathrm{ID} \wedge \mathrm{id}_i \notin \mathrm{ID}(\sigma)$
  Post-condition: $\sigma' \leftarrow \sigma \cup \{(\mathrm{id}_i, x_i)\}$
  Output: void
endFunction

Function $\mathsf{get}(\mathrm{id}_i)$
  Pre-condition: $\mathrm{id}_i \in \mathrm{ID} \wedge \mathrm{id}_i \notin \mathrm{ID}(\tau)$
  Post-condition: Let y_i be any output value such that $\tau \cup \{(\mathrm{id}_i, y_i)\} \in \Delta(\sigma)$.
    Then, $\tau' \leftarrow \tau \cup \{(\mathrm{id}_i, y_i)\}$
  Output: y_i
endFunction

Fig. 5. A Generic Sequential Specification of a Task $T = (\mathcal{I}, \mathcal{O}, \Delta)$.

Lemma 1. *Consider the splitter task* $T_{\mathsf{spl}} = (\mathcal{I}_{\mathsf{spl}}, \mathcal{O}_{\mathsf{spl}}, \Delta_{\mathsf{spl}})$. *There is no sequential object* X_{spl} *with a single operation satisfying* $VE(T_{\mathsf{spl}}) = VE(X_{\mathsf{spl}})$.

In a very similar way, one can prove that the following known tasks cannot be specified as sequential objects with a single operation: *exchanger* [17,36], *adaptive renaming* [4], *set agreement* [10], *immediate snapshot* [5], *adopt-commit* [6,13] and *conflict detection* [3].[2]

To circumvent the impossibility result in Lemma 1, we model any given task T through a sequential object S with two operations, set and get, that each process access in a specific way: it first invokes set with its input to the task T (receiving no output) and later invokes get in order to get an output value from T. Intuitively, decoupling the single operation of T into two (atomic) operations allows us to model concurrent behaviors that a single (atomic) operation cannot specify. In what follows, let $SSpec(S)$ be the set with all sequential executions of S in which each process invokes at most two operations, first set and then get, in that order.

Theorem 1. *For every task* $T = (\mathcal{I}, \mathcal{O}, \Delta)$ *there is a sequential object* S *with two operations,* $\mathsf{set}(id_i, x_i)$ *and* $\mathsf{get}(id_i) : y_i$, *such that there is a bijection* α *between* $VE(T)$ *and* $SSpec(S)$ *satisfying that: (1) each invocation or response of process* id_i *is mapped to an operation of process* id_i, *and (2) each invocation inv (response resp) with input (output)* x *is mapped to a completed* set *(*get*) operation with input (output)* x.

An implication of Theorem 1 is that if one is analyzing an algorithm that uses a building-block (subroutine, algorithm, etc.) B that solves a task T, one

[2] There are non-deterministic sequential specifications of these tasks with *unavoidable* and *pathological* executions in which some operations *guess* the inputs of future operations. See [9, Section 2] for a detailed discussion.

can safely replace B with the sequential object S related to T described in the theorem (each invocation to the operation of B is replaced with an (atomic) invocation to set and then an (atomic) invocation to get), and then analyze the algorithm considering the atomic operations of S. The advantage of this transformation is that (1) if all operations in an algorithm are atomic, we can think that each process takes a step at a time in an execution, hence obtaining a a transition system with atomic events, (2) at all times we have a concrete state of S in an execution (which is not clear in a task specification) and (3) given a state of S, an output for a get operation can be easily computed using the sequential object S (something that is typically complicated for B as it might be accessed concurrently).

The construction used (for simplicity) in the proof of Theorem 1 (in the full version of the paper) might be too coarse to be helpful for analyzing an algorithm. We would like to have a construction producing an equivalent sequential automaton modeling the task in a simpler way. Consider the simple sequential object in Fig. 5 obtained from any given task $T = (\mathcal{I}, \mathcal{O}, \Delta)$, which is described in a classic pre/post-condition form. Intuitively, the meaning of a state (σ, τ) is the following: σ contains the processes that have invoked the task so far (this represents the *participating set* of the current execution) while τ contains the outputs that have been produced so far. The main invariant of the specification is that $\tau \in \Delta(\sigma)$. It directly follows from the properties of the task: when a process invokes $\mathsf{set}(\mathrm{id}_i, x_i)$, we have that $\tau \in \Delta(\sigma \cup \{(\mathrm{id}_i, x_i)\})$ because $\Delta(\sigma) \subset \Delta(\sigma \cup \{(\mathrm{id}_i, x_i)\})$, and when a process invokes $\mathsf{get}(\mathrm{id}_i)$, it holds that $\tau \cup \{(\mathrm{id}_i, y_i)\} \in \Delta(\sigma)$ because $\Delta(\sigma)$ is a pure complex of dimension $dim(\sigma)$ and thus there must exist a simplex in $\Delta(\sigma)$ (properly) containing τ and with an output for id_i. One can formally prove that this sequential object and the one in the proof Theorem 1 define the same set of sequential executions.

Finally, one can obtain ad-hoc and equivalent specifications for specific tasks, like the one for splitters in Fig. 3 in Sect. 2.

4 Related Work

Linearizability Criteria. Neiger observed for the first time that some fundamental tasks, like *set agreement* [10] and *immediate snapshot* [5], cannot be modeled as sequential objects [33] (with a single operation). Motivated by the need of a unified framework for tasks and objects, he proposed *set-linearizability* [33]. Roughly speaking, a set sequential object is generalization of a sequential object in which transitions between states involve more than one operation (formally, a set of operations), meaning that these operations are allowed to occur concurrently, and their results can be *concurrency-dependent*. Set linearizability is the consistency condition for set-sequential objects, where one needs to find linearizability points (same as in linearizability) and several operations can be linearized at the same point (different from linearizability).

Later on, it was again observed that for some concurrent objects it is impossible to provide a sequential specification, and *concurrency-aware* linearizability

was defined [16]. Set linearizability and concurrency-aware linearizability are very closely related, both based on the same principle: sets of operations can occur concurrently. Also, a non-automatic verification technique for reasoning about concurrency-aware objects is presented in [16].

Recently it was observed in [9] that some natural tasks specify concurrency dependencies that are beyond the set-linearizability and concurrency-aware formalisms, hence that paper proposed *interval linearizability*. In an interval-sequential object not only sets of operations can occur concurrently but some of these operations might be pending and then overlap operations in the next transition; thus each operation corresponds to an interval instead of a single point. Interval linearizability is the related consistency condition in which, for each operation, one needs to find an interval in which the operation happens. It is shown in [9] that interval-linearizability is *complete* for tasks in the sense that it is possible to specify *any* task as an interval-sequential object (with a single operation).

Although interval-sequential specifications can model any task, this approach does not seem to be useful when one is searching for machine-checked proofs of concurrent algorithms. The main reason is that by replacing a task with its equivalent interval-sequential object, we obtain a transition system in which one still needs to think in concurrent behaviors, which is usually hard to deal with. In contrast, our proposed get-set transformation allows to "decouple" the inherent concurrency in tasks in a way that in the resulting transition system all events are atomic, namely, they happen one after the other.

Mechanized Verification of Distributed Algorithms. Mechanized (or machine-assisted) verification of distributed and concurrent algorithms is usually done with model checking or theorem proving or a combination of both. Enumerative model-checking is the oldest fully automatic method with tools like Spin [21] or TLC, the TLA$^+$ model checker [27]. To avoid the well-known problem of state explosion, various optimisations such as symmetry or reduction have been introduced, and recent work is on going on parameterized model checking, for instance with MCMT (Model Checking Modulo Theory) [14], Cubicle [11] or ByMC [26]. Nevertheless, automatic verification of a distributed/concurrent algorithm is still restricted to small finite instances of the algorithm or imposes significant constraints on its description, due to the limited expressiveness of the specification language.

Fully automatic theorem proving is based on a proof decision procedure. For useful logics, it is often semi-decidable at best and heavily depends on heuristics to achieve good performance. Recent work on SMT has made a substantial leap forward checking complex formulae combining first-order reasoning with decision procedures for theory such as arithmetic, equality, arrays. Nonetheless, the overall proof of a distributed algorithm is still largely manual and, when seeking confidence in this proof, an interactive proof assistant is the current approach. Several examples of verification of complex distributed algorithms exist: Chord

with Alloy [38], Pastry with TLA$^+$ [29,30], Paxos also with TLA$^+$ [28], snapshot algorithms in Event-B [2], just to cite a few.

Several wait-free implementations of tasks have been mechanically proven (e.g. [12,34,37]). However, to the best of our knowledge, no non-trivial algorithm built upon concurrent tasks have been mechanically proved. Our intuition for this situation is that proofs cannot be made modular and compositional when using bricks which are inherently concurrent if their internal structure must be visible to take into account this concurrency. Several complex and original algorithms can be found in the literature such as Moir and Anderson renaming algorithm [31] that we have considered in this paper, stacks implemented with elimination trees [36], lock-free queues with elimination [32]. In these papers, the correctness proofs are intricate as they must consider the algorithm as a whole, including the tricky part involving wait-free objects, and they have not been mechanically checked. Our approach which exposes a more simple and sequential specification (instead of a complex concurrent implementation) seeks to alleviate this limitation.

5 Final Remarks and Future Work

In this paper, we showed a technique to circumvent the known impossibility of specifying a task as a sequential object. Our technique consists in modeling the single operation of the task with two atomic operations, set and get. This transformation leads to a framework for developing transitional models of concurrent algorithms using tasks and sequential objects as building blocks. As a proof of concept, we developed in a companion paper [22] a full and modular TLA$^+$ proof of the Moir&Anderson renaming algorithm [31].

A natural extension of our work is to apply the framework to other concurrent algorithms. Another direction is to extend our techniques to the case of *refined tasks* and *interval-sequential* objects, recently defined in [9]. These two formalisms are generalization of the task and sequential object formalism with strictly more expressiveness; particularly, contrary to the task formalism, refined task are *multi-shot*, namely, each process may perform several invocations, possibly infinitely many.

A third direction is to study if the duality between the epistemic logic approach and the topological approach shown in [15] might be useful in verifying concurrent algorithms. Generally speaking, it is shown in [15] that a task can be represented as a *Kripke model* with an *action model*, specifying the knowledge obtained by processes when solving the task. It could be interesting to explore how this knowledge could be reflected in our set/get construction and if it could be useful in proving correctness.

Acknowledgements. Armando Castañeda was supported by PAPIIT project IA102417.

References

1. Alistarh, D.: The renaming problem: recent developments and open questions. Bull. EATCS **3**, 117 (2015)
2. Andriamiarina, M.B., Méry, D., Singh, N.K.: Revisiting snapshot algorithms by refinement-based techniques. Comput. Sci. Inf. Syst. **11**(1), 251–270 (2014)
3. Aspnes, J., Ellen, F.: Tight bounds for adopt-commit objects. Theory Comput. Syst. **55**(3), 451–474 (2014)
4. Attiya, H., Bar-Noy, A., Dolev, D., Peleg, D., Reischuk, R.: Renaming in an asynchronous environment. J. ACM **37**(3), 524–548 (1990)
5. Borowsky, E., Gafni, E.: Generalized FLP impossibility result for t-resilient asynchronous computations. In: STOC 1993: Proceedings of the ACM Symposium on Theory of computing, pp. 91–100. ACM, New York (1993)
6. Borowsky, E., Gafni, E., Lynch, N.A., Rajsbaum, S.: The BG distributed simulation algorithm. Distrib. Comput. **14**(3), 127–146 (2001)
7. Castañeda, A., Hurault, A., Quéinnec, P., Roy, M.: Tasks in modular proofs of concurrent algorithms. CoRR, arXiv:1909.05537 [cs.DC] (2019)
8. Castañeda, A., Rajsbaum, S., Raynal, M.: The renaming problem in shared memory systems: an introduction. Comput. Sci. Rev. **5**(3), 229–251 (2011)
9. Castañeda, A., Rajsbaum, S., Raynal, M.: Unifying concurrent objects and distributed tasks: interval-linearizability. J. ACM **65**(6), 45 (2018)
10. Chaudhuri, S.: More choices allow more faults: set consensus problems in totally asynchronous systems. Inf. Comput. **105**(1), 132–158 (1993)
11. Conchon, S., Goel, A., Krstić, S., Mebsout, A., Zaïdi, F.: Cubicle: a parallel SMT-based model checker for parameterized systems. In: Madhusudan, P., Seshia, S.A. (eds.) CAV 2012. LNCS, vol. 7358, pp. 718–724. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-31424-7_55
12. Drăgoi, C., Gupta, A., Henzinger, T.A.: Automatic linearizability proofs of concurrent objects with cooperating updates. In: Sharygina, N., Veith, H. (eds.) CAV 2013. LNCS, vol. 8044, pp. 174–190. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39799-8_11
13. Gafni, E.: Round-by-round fault detectors: unifying synchrony and asynchrony (extended abstract). In: Proceedings of the Seventeenth Annual ACM Symposium on Principles of Distributed Computing, PODC 1998, pp. 143–152 (1998)
14. Ghilardi, S., Ranise, S.: MCMT: a model checker modulo theories. In: Giesl, J., Hähnle, R. (eds.) IJCAR 2010. LNCS (LNAI), vol. 6173, pp. 22–29. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14203-1_3
15. Goubault, É., Ledent, J., Rajsbaum, S.: A simplicial complex model for dynamic epistemic logic to study distributed task computability. In: Ninth International Symposium on Games, Automata, Logics, and Formal Verification, GandALF 2018, pp. 73–87 (2018)
16. Hemed, N., Rinetzky, N., Vafeiadis, V.: Modular verification of concurrency-aware linearizability. In: Moses, Y. (ed.) DISC 2015. LNCS, vol. 9363, pp. 371–387. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-48653-5_25
17. Hendler, D., Shavit, N., Yerushalmi, L.: A scalable lock-free stack algorithm. J. Parallel Distrib. Comput. **70**(1), 1–12 (2010)
18. Herlihy, M., Kozlov, D.N., Rajsbaum, S.: Distributed Computing Through Combinatorial Topology. Morgan Kaufmann, Burlington (2013)
19. Herlihy, M., Shavit, N.: The Art of Multiprocessor Programming. Morgan Kaufmann, Burlington (2008)

20. Herlihy, M., Wing, J.M.: Linearizability: a correctness condition for concurrent objects. ACM Trans. Program. Lang. Syst. **12**(3), 463–492 (1990)
21. Holzmann, G.J.: The SPIN Model Checker - Primer and Reference Manual. Addison-Wesley, Boston (2004)
22. Hurault, A., Quéinnec, P.: Proving a non-blocking algorithm for process renaming with TLA$^+$. In: Beyer, D., Keller, C. (eds.) TAP 2019. LNCS, vol. 11823, pp. 147–166. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-31157-5_10
23. IEC: IEC-61508: Functional safety. https://www.iec.ch/functionalsafety/
24. Scherer III, W.N., Lea, D., Scott, M.L.: Scalable synchronous queues. Commun. ACM **52**(5), 100–111 (2009)
25. Scherer, W.N., Scott, M.L.: Nonblocking concurrent data structures with condition synchronization. In: Guerraoui, R. (ed.) DISC 2004. LNCS, vol. 3274, pp. 174–187. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-30186-8_13
26. John, A., Konnov, I., Schmid, U., Veith, H., Widder, J.: Parameterized model checking of fault-tolerant distributed algorithms by abstraction. In: Formal Methods in Computer-Aided Design, FMCAD 2013, pp. 201–209. IEEE, October 2013
27. Lamport, L.: Specifying Systems. Addison Wesley, Boston (2002)
28. Lamport, L.: Byzantizing paxos by refinement. In: Peleg, D. (ed.) DISC 2011. LNCS, vol. 6950, pp. 211–224. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-24100-0_22
29. Lu, T.: Formal verification of the pastry protocol. Ph.D. thesis, Université de Lorraine - Universität des Saarlandes, July 2013
30. Lu, T., Merz, S., Weidenbach, C.: Towards verification of the pastry protocol using TLA$^+$. In: Bruni, R., Dingel, J. (eds.) FMOODS/FORTE -2011. LNCS, vol. 6722, pp. 244–258. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21461-5_16
31. Moir, M., Anderson, J.H.: Wait-free algorithms for fast, long-lived renaming. Sci. Comput. Program. **25**(1), 1–39 (1995)
32. Moir, M., Nussbaum, D., Shalev, O., Shavit, N.: Using elimination to implement scalable and lock-free FIFO queues. In: 17th ACM Symposium on Parallelism in Algorithms and Architectures, SPAA 2005, pp. 253–262. ACM (2005)
33. Neiger, G.: Set-linearizability. In: Proceedings of the Thirteenth Annual ACM Symposium on Principles of Distributed Computing, Los Angeles, California, USA, 14–17 August 1994, p. 396 (1994)
34. O'Hearn, P.W., Rinetzky, N., Vechev, M.T., Yahav, E., Yorsh, G.: Verifying linearizability with hindsight. In: 29th Annual ACM Symposium on Principles of Distributed Computing, PODC 2010, pp. 85–94. ACM (2010)
35. Raynal, M.: Concurrent Programming - Algorithms, Principles, and Foundations. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-32027-9
36. Shavit, N., Touitou, D.: Elimination trees and the construction of pools and stacks. Theory Comput. Syst. **30**(6), 645–670 (1997)
37. Tofan, B., Schellhorn, G., Reif, W.: A compositional proof method for linearizability applied to a wait-free multiset. In: Albert, E., Sekerinski, E. (eds.) IFM 2014. LNCS, vol. 8739, pp. 357–372. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10181-1_22
38. Zave, P.: Using lightweight modeling to understand Chord. SIGCOMM Comput. Commun. Rev. **42**(2), 49–57 (2012)