

Lecture Notes on Data Structures

Matt Fredrikson

Carnegie Mellon University

Lecture 3

September 2, 2025

1 Introduction

Our study of logical contracts so far has focused on *control structures*: How do functions, assignments, and loops give rise to verification conditions that entail the correctness of the code? In the next example we will introduce two elementary data structures (lists and queues) and verify an implementation of queues using two lists. We then deepen our investigation into how to reason about data structures. A key concept that we need to capture is that of a *data structure invariant*. With executable contracts, such invariants are checked by functions. With logical contracts, they are expressed as logical properties of the data structures. We can classify data structures as *persistent* (also called *immutable*) and *ephemeral* (also called *mutable*). Persistent data structures are prevalent in functional programming, while ephemeral data structure are more common in imperative programming. Since WhyML covers a spectrum of functional and imperative programming, we will consider both and identify some commonalities and differences.

We will also take a look at testing, why it is still important, and briefly consider how to use the Why3 IDE during code development.

Learning goals. After this lecture, you should be able to:

- Write logical contracts for small functional programs over simple immutable data types such as lists
- Verify simple code over (inductive) data structure

2 Verifying Properties of Data Structures

For the moment, we'll stay in the functional world, although in WhyML we use data structures similarly also in imperative programs. There is a clever implementation of queues in a functional language using two stacks (usually directly represented by lists), sometimes called *functional queues*. If the queue is used in a single-threaded way (that is, we don't dequeue from a queue in the same state more than once) then amortized analysis shows that both enqueue and dequeue operations have constant amortized cost.

The basic algorithmic idea is as follows. The queue is represented by two lists, the front and the back. Initially both are empty. When we enqueue, we add elements to the back, and when we dequeue, we take them from the front. If the front happens to be empty when a dequeue request comes in we *reverse* the back to become the new front. For example, after enqueueing 1, 2, 3 in that order, the front is still empty and the back will be the list [3, 2, 1]. If we now dequeue, the front will become [1, 2, 3] (the reverse of the back) and we remove 1 from the front, leaving it [2, 3] with the back empty.

To represent the queue, we see a couple of new constructs. One of them is *polymorphism* because we would like queues to be generic in the types of the elements. In particular, the type `queue 'a` will be a queue with elements of type `'a` (usually pronounced *alpha*). We also need lists, which we can find in the [Why3 Standard Library](#). Among many other things, we find:

```

1  type list 'a = Nil | Cons 'a (list 'a)
2
3  let rec function (++) (l1 l2: list 'a) : list 'a =
4      match l1 with
5      | Nil          -> l2
6      | Cons x1 r1 -> Cons x1 (r1 ++ l2)
7
8  let rec function reverse (l: list 'a) : list 'a =
9      match l with
10     | Nil          -> Nil
11     | Cons x r -> reverse r ++ Cons x Nil
12  end

```

These functions are declared with `let rec function` which means they can be used in contracts as well as computationally. We do note that, computationally, we may want to use a different function that this particular `reverse` because its complexity is $O(n^2)$ for a list of length n . We'll note that, but we won't worry about it in today's lecture.

We see that lists have constructors `Nil` and `Cons` and that we discriminate between lists using the expression `match ... with ... end`. We use a *record* of two elements, the front and the back, as our representation of queues.

```

1  type queue 'a = { front : list 'a ; back : list 'a }

```

We would like to define the following functions

```

1  empty () : queue 'a
2  enq (x : 'a) (q : queue 'a) : queue 'a
3  deq (q : queue 'a) : option ('a , queue 'a)

```

A queue might be empty, so `deq` returns an optional pair consisting of the first element and the remainder of the queue. For this we need the option library:

```
1 type option 'a = None | Some 'a
```

Before we write code, we should decide on the specifications. The key idea is that we use a single list to represent the queue, with the first element in the queue at the front of the list. In other words, we use another data structure (a list) in the specification of the behavior of the queue. Of course, we do not want to use such a list as an *implementation* because the cost of an enqueue operation would be linear in the size of the queue (rather than have amortized constant cost). Therefore we define the function `sequence` that represents the state of a queue in the proper sequence. Recall that because we add the elements to the back, to represent the proper state of the queue we have to *reverse* the back.

```
1 function sequence (q : queue 'a) : list 'a = q.front ++ reverse q.back
```

Now let's write the postconditions for all of the functions. They have no precondition since any state of the queue is valid. The ones for empty and `enq` are fairly straightforward. For the enqueue operation we add the new element to the end of the sequence.

```
1 let empty () : queue 'a =
2   ensures { sequence result = Nil }
3   ...
4
5 let enq (x : 'a) (q : queue 'a) : queue 'a =
6   ensures { sequence result = sequence q ++ Cons x Nil }
7   ...
8
9 let deq (q : queue 'a) : option ('a , queue 'a) =
```

For the dequeue operation, we have to return `Nil` if the queue is empty and `Some (x, r)` if `x` is at the front of the queue and `r` is the remainder. Writing this out logically, we use an existential quantifier.

```
1 let deq (q : queue 'a) : option ('a , queue 'a) =
2   ensures { (result = None /\ sequence q = Nil)
3             \/ ( exists x:'a. exists r:queue 'a. result = Some(x,r)
4                 /\ sequence q = Cons x (sequence r)) }
```

At this point the code itself is not too difficult, just for the case of enqueue we nest two matches because if the front is empty the queue is empty only if the back is also empty. We show the code here; the live-coded version which may be different in minor details can be found in [queue.mlw](#).

```

1 module Queue
2
3   use int.Int
4   use list.List
5   use list.Append
6   use list.Reverse
7   use option.Option
8
9   (* type list 'a = Nil | Cons 'a (list 'a) *)
10  (* type option 'a = None | Some 'a *)
11
12  type queue 'a = { front : list 'a ; back : list 'a }
13
14  function sequence (q : queue 'a) : list 'a = q.front ++ reverse q.
    back
15
16  let empty () : queue 'a =
17  ensures { sequence result = Nil }
18  { front = Nil ; back = Nil }
19
20  let enq (x : 'a) (q : queue 'a) : queue 'a =
21  ensures { sequence result = sequence q ++ (Cons x Nil) }
22  { front = q.front ; back = Cons x q.back }
23
24  let deq (q : queue 'a) : option ('a , queue 'a) =
25  ensures { (sequence q = Nil /\ result = None)
26           \/ (exists x:'a, q':queue 'a. sequence q = Cons x (
27             sequence q') /\ result = Some (x,q')) }
28  match q.front with
29  | Nil -> match reverse q.back with
30          | Nil -> None
31          | Cons y ys -> Some (y, { front = ys ; back = Nil })
32          end
33  | Cons x xs -> Some (x, { front = xs; back = q.back })
34  end
35 end

```

While the code and specifications seem correct, we cannot be 100% confident that the prover will be able to verify it. In particular, the reasoning depends on the lemmas in the libraries pertaining to the properties of append ('++') and reverse. Fortunately, in this case it succeeds. Just for variety, we tried it with the CVC4 prover.

```

% why3 prove -P cvc4 queue.mlw
queue.mlw Queue empty'vc: Valid (0.04s, 6867 steps)
queue.mlw Queue enq'vc: Valid (0.05s, 7687 steps)
queue.mlw Queue deq'vc: Valid (0.08s, 13374 steps)
%

```

3 Data Structure Invariants

As a simple example of data structure invariants we reconsider our implementation of queues. We would like to extend the interface with a function `qsize` that returns the number of elements in a queue. The obvious way to compute this would be to compute the lengths of the front and back and add them, but the complexity would be linear in the size of the queue. To answer this query in constant time we add a `size` field to the queue and maintain it as we enqueue or dequeue elements. The data structure invariant here is that the `size` field always contains the sum of the lengths of the front and back.

```
1  type queue 'a = { front : list 'a ;
2                      back : list 'a ;
3                      size : int }
4  invariant { size = length front + length back }
```

In the logic, the verifier assumes the invariant when reasoning about queues. This could create a logical inconsistency (*everything is provable!*) if the invariant is unsatisfiable. For example, the trivial invariant *false* means any program in the scope of this declaration could now be verified. To avoid this, Why3 will try to prove that there exists an instance of the data structure for which the invariant is satisfied. For complex invariants this can be difficult, so WhyML gives you a way to specify an instance of the data structure satisfying the invariant with a `'by'` clause. In this particular case it would be unnecessary, but it is good practice to always supply it.

```
1  type queue 'a = { front : list 'a ;
2                      back : list 'a ;
3                      size : int }
4  invariant { size = length front + length back }
5  by { front = Nil ; back = Nil ; size = 0 }
```

How the witness helps. Without a witness, Why3 must prove an existentially quantified assertion that there is at least one value of the record fields that satisfies the invariant. With a witness, the existential disappears and the VC becomes a concrete equality on the provided values. Concretely, without a witness, Why3 produces the following VC for the type invariant:

```
goal queue'vc :
  exists front:list 'a, back:list 'a, size:int.
    size = (length front + length back)
```

With a witness, the VC becomes:

```
goal queue'vc : 0 = (length (Nil: list 'a) + length (Nil: list 'a))
```

In the first case, the prover is asked to show that there exist some values that make the invariant true, while the second simply checks that the concrete values given in the

witness do in fact make the invariant true. In simple cases, provers will often be able to prove such statements, but as the data structure and associated invariants becomes more complex, they will struggle to prove the existential goal.

Implementing `qsize`. In this example it is relatively easy to update the code to maintain the size field and Why3 will prove that it is always correct. In constructing the verification condition we may think of the invariant as being *assumed* at the beginning of a function (like a precondition) and *proved* at the end of a function (like a postcondition). In between, the invariant may be violated, although this possibility does not come into play here. It is necessary because you may build or modify an element of the data structure incrementally and only at the end does the invariant hold.

The `qsize` function just returns the size field. In addition, the postcondition certifies that it is indeed the length of the queue (when viewed as a single sequence of element, which represents the client's perspective).

```
1  let qsize (q : queue 'a) : int =  
2  ensures { result = length (sequence q) }  
3  q.size
```

4 Testing and Implicit Preconditions

In 1977 Donald Knuth famously wrote “*Beware of bugs in the above code; I have only proved it correct, not tried it.*” in a 5-page memo *Notes on the van Emde Boas construction of priority queues: An instructive use of recursion*. You might think that if he had proved it using Why3 then he wouldn't have to be worried. I think he still would have been, and he should have been! Here are some of the things that can still go wrong even if Why3 says “Verified!”.

- Your preconditions could be prohibitively strict, even to the point where *no client could possibly call your functions*.
- Your postconditions could be prohibitively lax, to the point where *the client obtains no information at all about the computation of your function*.
- Your definitions and axioms could be incorrect in the sense that they do not capture the property you were trying to prove. In the extreme case they could be *vacuous* (equivalent to true and therefore not saying anything) or *inconsistent* (equivalent to false and therefore implying everything whatsoever).
- There could be a bug in Why3 or one or more of the back-end provers, extracting an incorrect verification condition or proving one that isn't valid. In fact, it is almost certain that Why3 and all the back-end provers have bugs, so it is just a matter of probabilities whether you trip any of them.

You may think these are unlikely, but you should be prepared that almost certainly at least *some of these things will happen to you*.

When grading your homework, we combat these issues by combining manual inspection with replaying the Why3 sessions.

When you develop your code, you most likely will be in the *opposite* situation for a while: your code can not be verified. Then you have to look for the exact opposite of the points raised above, plus a very real first possibility:

- Your code is incorrect!
- Your preconditions could be prohibitively lax, to the point where they are too weak to imply loop invariants or preconditions for operations or function calls, or the postcondition at the end of the function. Of course, this may be the case even if your code is correct!
- Your postconditions could be prohibitively strict, to the point where they simply do not follow from what you know at the end of the function. Again, this may be the case even with correct code.
- Your definitions and axioms do not properly capture the property you wish to prove (and are convinced is true).
- The back-end provers in Why3 are not strong enough to prove the verification condition *even though it is true* and, on top of it, *your code is correct*.

To mitigate all these issues, it is sensible to combine testing with verification even during the development process. Among other things:

- It may help you to determine whether your code is correct and, if not, have some counterexamples. Unfortunately, today's technology is such that it is difficult to obtain counterexamples from failing provers. A reasonable set of successful test cases may point you towards other kinds of issues you might have.
- If you run Why3 on your program including the testing code it may help you uncover situations where the preconditions are too strict. That could mean your test function will fail to verify, even if Why3 assumes all the pre- and post-conditions in the rest of your program.
- It may help you to think about the code by writing out explicitly how it should behave on specific examples.

Even though our code has been verified, let's write a little test function. By convention, a test function should take unit as an argument, because the `why3 execute` command will pass the unit element to the specified function.

```
1 let test () : (int, int) =
2   let q0 = empty () in
3   let q1 = enq 1 q0 in
4   let q2 = enq 2 q1 in
5   match deq q2 with
6   | Some (x, q3) -> match deq q3 with
```

```

7 | Some (y, q4) -> match deq q4 with
8 | None -> (x,y)
9 end end end

```

We should be able to execute this with the command

```
% why3 execute queue.mlw --use=Queue 'test ()'
```

but unfortunately due to limitation (or bug) in the implementation of the WhyML interpreter related to polymorphism this fails. It is nevertheless interesting because in the first two match expressions above we only supply the branch for `Some`. When we dequeue from `q2` it is impossible for the queue to be empty, so the value returned cannot be `None`. When Why3 encounters a match expression that does not cover all the possible cases based only on the type information, it generates a verification condition to *prove* that the omitted cases are impossible. In this example, by tracking the postconditions, the verifier should know precisely that the sequence at the point where `deq q2` is called is 1, 2 and therefore the queue nonempty.

We can actually test this: let's add the postcondition that the answer *must be* (1,2) and run the alt-ergo theorem prover and the file with the additional, strict function test.

```

% why3 prove -P alt-ergo queue.mlw
queue.mlw Queue empty'vc: Valid (0.01s, 26 steps)
queue.mlw Queue enq'vc: Valid (0.02s, 123 steps)
queue.mlw Queue deq'vc: Valid (0.28s, 908 steps)
queue.mlw Queue qsize'vc: Valid (0.00s, 10 steps)
queue.mlw Queue test'vc: Timeout (5.00s)
%

```

Hmmm, it doesn't seem to be able to prove our test function. Suspecting that the code and contracts are correct and the theorem prover is too weak, let's try CVC4.

```

% why3 prove -P cvc4 queue.mlw
queue.mlw Queue queue'vc: Valid (0.02s, 4244 steps)
queue.mlw Queue empty'vc: Valid (0.04s, 6706 steps)
queue.mlw Queue enq'vc: Valid (0.06s, 9124 steps)
queue.mlw Queue deq'vc: Valid (0.19s, 25682 steps)
queue.mlw Queue qsize'vc: Timeout (5.00s, 2064194 steps)
queue.mlw Queue test'vc: Valid (0.43s, 57994 steps)
%

```

Although it is somewhat slow and take a lot of steps, CVC4 manages to verify that `test ()` must return (1,2)! However, for some strange reason it fails to verify the very simple `qsize` function which was no problem at all for alt-ergo.

At this point we can invoke

```
% why3 ide queue.mlw
```


and launch strategy Auto level 2 (or keyboard shortcut 2) which tries different provers on different subgoals and quickly verifies the code. If we save the session under the file menu (or keyboard shortcut Control-S) we can examine the session statistics.

```
% why3 session info --stats queue
== Number of root goals ==
    total: 6   proved: 6

== Number of sub goals ==
    total: 0   proved: 0

== Goals not proved ==

== Goals proved by only one prover ==
+-- file [../queue.mlw]
+-- theory Queue
+-- goal queue'vc: CVC4 1.7
+-- goal empty'vc: CVC4 1.7
+-- goal enq'vc: CVC4 1.7
+-- goal deq'vc: CVC4 1.7
+-- goal qsize'vc: Alt-Ergo 2.3.1
+-- goal test'vc: CVC4 1.7

== Statistics per prover: number of proofs, time (minimum/maximum/average) in seconds ==
Alt-Ergo 2.3.1      :   1  0.00  0.00  0.00
CVC4 1.7            :   5  0.02  0.46  0.16

%
```

One lesson from this example is that certain constructs have *implicit preconditions*. This includes pattern matches against data types (they should be exhaustive), calls to functions such as `div` or `mod` (the second argument should not be zero), array accesses (the index should be in bounds; not yet discussed), and functions passing data structures endowed with invariants (they should hold).