Robust and Compositional Verification of Object Capability Patterns
(Long Version)

DAVID SWASEY, MPI-SWS*
DEEPAK GARG, MPI-SWS*
DEREK DREYER, MPI-SWS*

In scenarios such as web programming, where code is linked together from multiple sources, object capability patterns (OCPs) provide an essential safeguard, enabling programmers to protect the private state of their objects from corruption by unknown and untrusted code. However, the benefits of OCPs in terms of program verification have never been properly formalized. In this paper, building on the recently developed Iris framework for concurrent separation logic, we develop OCPL, the first program logic for compositionally specifying and verifying OCPs in a language with closures, mutable state, and concurrency. The key idea of OCPL is to account for the interface between verified and untrusted code by adopting a well-known idea from the literature on security protocol verification, namely robust safety. Programs that export only properly wrapped values to their environment can be proven robustly safe, meaning that their untrusted environment cannot violate their internal invariants. We use OCPL to give the first general, compositional, and machine-checked specs for several commonly-used OCPs—including the dynamic sealing, membrane, and caretaker patterns—which we then use to verify robust safety for representative client code. All our results are fully mechanized in the Coq proof assistant.

1 INTRODUCTION

Suppose you have a mutable reference $\ell$ whose contents you care about, meaning that you want to impose some invariant on it (e.g., $\ell$ always points to an even number). Suppose further that you want to share access to $\ell$ with code you did not write and that you do not trust to preserve the invariant on $\ell$. To ensure the invariant on $\ell$ is maintained, you therefore do not want to pass the untrusted code the reference $\ell$ directly. Instead, you might construct a read-only wrapper $w$ as follows:

$$\text{readonly} \triangleq \lambda r. \lambda _- \! r \quad w \triangleq \text{readonly} \ell$$

Here, $\text{readonly}$ transforms a reference $r$ into a thunk that, when applied, returns the current contents of $r$. The expression $w$ applies $\text{readonly}$ to our reference of interest $\ell$, constructing a function for reading $\ell$’s contents. You can now pass $w$ to untrusted code without worrying about it corrupting your invariant on $\ell$.

Wrappers like $w$ are often called object capabilities, and the wrapper construction function $\text{readonly}$ is a very simple example of an object capability pattern (OCP) (Miller et al. 2000). Although OCPs date back at least to the 1970s (Morris 1973), they have gained increased currency in recent years, both in new languages centered around object capabilities (Miller et al. 2000; Mettler et al. 2010; Spiessens and Roy 2004; Stiegler and Miller 2006) and in the context of web programming, where interfacing with untrusted code is commonplace. Web sandboxing systems like Yahoo!’s ADsafe (Crockford 2008; Politz et al. 2014) and Google’s Caja (Miller et al. 2008), for example, elaborate untrusted JavaScript source code into an ostensibly safe subset of JavaScript while introducing wrappers that attenuate access to JavaScript libraries. Caja relies on the so-called membrane pattern, which automatically wraps all objects crossing to untrusted, sandboxed code so that access to the underlying objects is appropriately restricted.
OCPs are believed to provide crucial security guarantees because, in a language like JavaScript, objects provide essentially no data abstraction on their own—the only way to enforce data abstraction is to hide private state in the environment of a closure (e.g., as $\ell$ was hidden in the environment of `readonly $\ell$` in the example above). So OCPs, which use closures to mediate access to private state in a systematic way, are one of the few effective mechanisms available in a language like JavaScript for enforcing data abstraction in the presence of possibly malicious code.

Unfortunately, despite the ubiquity of OCPs in modern web programming, remarkably little attention has been paid to the question of what exactly are the security guarantees that such OCPs are supposed to provide, and how might we prove that they actually provide them. Even in the case of the extremely basic `readonly` pattern shown above, it is not at all obvious what is the "right" formal specification for `readonly`. What, in particular, are the formal conditions on $\ell$ that are needed to guarantee that `readonly $\ell$` can be "safely" shared with untrusted code? If $\ell$ merely points to an integer, no conditions may be necessary, but what if $\ell$ points to a closure or some other higher-order object? How do we know that giving readonly access to $\ell$ will not indirectly give untrusted code a way of gaining full access and violating important invariants maintained by the user of this OCP?

The most recent, state-of-the-art attempt to grapple with these types of questions is due to Devriese et al. (2016). They build a Kripke logical relations model for reasoning about object capabilities in a language with higher-order state, and they use their model to verify several concrete examples of capability-wrapped user code. In each case, they demonstrate that invariants on the private state of the user code are preserved even when the user code is shared with unknown attacker code. However, their model is limited in that it provides no way to compositionally specify what an OCP does, and it provides no clear specification of the general property that a piece of user code must satisfy in order to be safely shareable with untrusted code. Furthermore, they only consider very simple capability patterns, none as complex as, say, the aforementioned membrane pattern.

In this paper, we present OCPL (a Logic for OCPs), the first formal system for compositionally specifying and verifying the security guarantees provided by OCPs, in the context of a simple but representative programming language with higher-order functions, state, and concurrency. In contrast to prior work, OCPL enables one to reason modularly about both OCP implementations and user code that depends on them, and to specify a general property on user code that ensures such code can be safely shared with untrusted code without having its internal invariants violated. We use OCPL to reason about several commonly-used OCPs, including the dynamic sealing (sealer-unsealer), membrane, and caretaker patterns, and in so doing, provide the first formal explanation of what these OCPs achieve.

OCPL is a program logic derived from Iris, a recently developed framework for higher-order concurrent separation logic (Jung et al. 2015, 2016; Krebbers et al. 2017a). Iris was originally proposed as a very general logic with a few simple primitives, using which one can derive advanced proof principles from a variety of modern separation logics as needed. Moreover, Krebbers et al. (2017b) have recently developed a powerful new proof mode for Iris in the Coq proof assistant. The Iris proof mode enables one to carry out interactive, tactical, machine-checked proofs about programs in the Iris logic (embedded in Coq) in much the same way as one normally carries out interactive tactical proofs when working in the meta-logic of Coq itself. By virtue of building OCPL on top of Iris, we inherit the flexibility of the existing Iris framework, as well as its support for mechanizing proofs about programs in Coq. All our results and examples are, in fact, fully verified in Coq.

---

1Despite the name, object capabilities are not fundamentally limited to object-oriented programming languages.

2The Coq formalization is available online (OCPL 2017).
The key idea in OCPL is how it characterizes the interface between verified user code and untrusted code, via the concept of a "low-integrity value" (or low value for short) adapted from the literature on verification of security protocols (Abadi 1999). Roughly speaking, a low value is a value that can be safely shared with untrusted code, such as the closure wrappers returned by OCPs. More precisely, a low value is a value from which no code can possibly extract a direct reference to private state. To formalize this notion, we employ a logical relation, which is easy to define using Iris's built-in support for guarded recursive predicates. Then, with the idea of low values in hand, it becomes clear how to specify when a piece of user code can be safely linked with untrusted code—namely, when the only values passed back and forth between them are low values.

To make things concrete, let us return to our motivating example. Using the notion of low values, we can prove the following specification for the readonly pattern:

$$\forall \ell. \left( \mathbb{T} \Rightarrow \ell \Rightarrow \left( x. \text{lowval } x \right) \right) \Rightarrow \left( \mathbb{T} \Rightarrow \text{readonly } \ell \left( f. \text{lowval } f \right) \right)$$

The premise of this spec says that, in order to apply readonly to a location \( \ell \), we must first prove that dereferencing \( \ell \) always produces a low value. (This makes sense because once we pass the read-only wrapper to untrusted code, it may invoke the wrapper and obtain the contents of \( \ell \) at any time.) The payoff is in the conclusion: the closure returned by readonly \( \ell \) is itself a low value and may therefore be safely shared with untrusted code.

Given this specification for readonly, let us now consider a client of readonly that wishes to maintain an invariant on its private (i.e., high-integrity) state:

$$\text{usetwo} \triangleq \text{let } r = \text{ref } 2 \text{ in}$$
$$\text{let } w = \text{readonly } r \text{ in}$$
$$\text{let use} = \lambda \_ . \text{assert } ((!r) = 2) \text{ in}$$
$$\left( \text{use}, w \right)$$

Here, the expression usetwo allocates a reference cell \( r \) containing the value 2, and returns a pair of functions, one of which simply asserts that \( r \) continues to contain 2, and the other of which is the result of readonly \( r \). The reference cell \( r \) is, however, kept private (i.e., hidden in the environments of the closures use and \( w \)).

Our aim is to verify that usetwo can be safely linked with arbitrary untrusted code. To do this, we first prove that it satisfies the spec \( \mathbb{T} \Rightarrow \text{usetwo } \left( x. \text{lowval } x \right) \) which means that we can always run usetwo and, if we do, it returns a low value. It is easy to see intuitively why this spec holds: reference cell \( r \) always contains 2, which is trivially low, so together with the spec for readonly, this is sufficient to ensure that the closures returned by usetwo are low as well.

Next, we appeal to a general meta-theorem—and one of the main technical results of this paper—called robust safety. Robust safety states that, if some user code satisfies a spec like the one given above for usetwo—i.e., a spec whose postcondition stipulates that the resulting value is low-integrity—then we can run that verified user code under an arbitrary adversarial context \( C \) (i.e., any context \( C \) that does not itself contain any assert \( t \) statements), and we will be assured that the execution of the resulting program will never result in a violation of any of the user code's internal assertions. Thus, in the particular case of usetwo, we know that the assertion that \( r \) equals 2 will never fail.

Robust safety is a well-known meta-theorem in the security literature (Bengtson et al. 2011; Gordon and Jeffrey 2001), but it has not heretofore been employed in the context of object capability programming. One of the central contributions of this paper is the observation that robust safety is exactly the property a language must satisfy in order to support OCPs. Moreover, as we will
Head reduction (selected rules)

\[
\begin{align*}
\text{Lit} & \ni \text{lit} \ ::= n \mid b \\
\text{Exp} & \ni e \ ::= x \mid \text{rec } f \ x \ e \mid e \ e \mid \text{lit} \mid \text{unop } e \mid \text{binop } e \ e \mid \text{if } e \ \text{then } e \ \text{else } e \mid \\
& \quad \text{fork } e \mid \ell \mid \text{ref } e \mid ! e \mid e \leftarrow e \mid \text{CAS } e \ e \ e \ e \mid \text{assert } e \\
\text{Val} & \ni v \ ::= \text{rec } f \ x \ e \mid \text{lit} \mid () \mid (v, v) \mid \text{inl } v \mid \text{inr } v \mid \ell \\
\text{Ectx} & \ni K ::= \text{(standard left-to-right evaluation contexts)}
\end{align*}
\]

\[
\begin{align*}
& h \in \text{Heap} \triangleq \text{Loc} \xrightarrow{\text{fin}} \text{Val} \\
& \sigma \in \text{State} \triangleq \text{Heap} \times \text{GoodnessBit} \\
& T \in \text{Tpool} \triangleq \text{Exp}^* \\
& e; \sigma \rightarrow r; e'; \sigma' ; e_f \\
& (\text{rec } f \ x \ e) \ \text{closed} \\
& (\text{rec } f \ x \ e) \ v; \sigma \rightarrow r; e[v/x, \text{rec } f \ x \ e/f]; \sigma; \epsilon \\
& \text{assert true}; \sigma \rightarrow r; () ; \sigma; \epsilon \\
& \text{fork } e; \sigma \rightarrow r; (); \sigma; e \\
& \text{ref } v; (h, g) \rightarrow r; (h \cup [\ell \mapsto v], g); \epsilon \\
& \sigma. h(\ell) = v \\
& ! \ell; \sigma \rightarrow r; v; \sigma; \epsilon \\
& h(\ell) = v \\
& \ell \leftarrow w; (h, g) \rightarrow r; (); (h[\ell \mapsto w], g); \epsilon \\
& \text{CAS } \ell \ v_1 \ v_2; \sigma \rightarrow r; \text{false}; \sigma; \epsilon \\
& \sigma. h(\ell) \neq v_1 \\
& \text{CAS } \ell \ v_1 \ v_2; (h \cup [\ell \mapsto v_1], g) \rightarrow r; (h \cup [\ell \mapsto v_2], g); \epsilon \\
& T; \sigma \rightarrow T'; \sigma' \\
& e; \sigma \rightarrow r; e'; \sigma' ; e_f \\
& (T_1, K[e], T_2); \sigma \rightarrow (T_1, K[e'], T_2, e_f); \sigma'
\end{align*}
\]

Fig. 1. Heap language with assertions

demonstrate via a range of interesting examples, the notion of low-integrity values is essential to compositionally specifying an OCP’s contribution toward the robust safety of programs that use it.

The remainder of the paper is structured as follows. First, in §2, we introduce some basics of OCPL, along with our formalization of low values, explain how we verify our motivating readonly example, and state the key metatheorems of adequacy and robust safety. In §§3–5, we present specifications for several more complex and realistic OCPs, along with examples of how to use those specifications to verify representative clients. In §6, we discuss related work. Finally, in §7, we conclude with a discussion of future work.

2 Robust Safety and OCPL

In this section, we define the semantics of the programming language we will be reasoning about (§2.1), present the main ideas of OCPL and show how to use it to reason formally about our motivating example (§2.2), and state key metatheorems for OCPL (§2.3).
2.1 A higher-order concurrent heap language with assertions

The programming language we consider in this paper, which we call HLA, is a fairly standard higher-order concurrent imperative language, essentially the same as the one studied by Krebbers et al. (2017a), but with one important extension. We define the language in Fig. 1. The core of the language is a call-by-value λ-calculus with recursive functions, products, sums, references, and fork-based concurrency. The definitions of free and bound variables and capture-avoiding substitution are standard and omitted. We endow the language with an operational semantics in the style of Felleisen and Hieb (1992). The head reduction judgment \( e; \sigma \rightarrow e'; \sigma' \) describes the single steps of computation that may be taken by individual threads. It means that, starting in state \( \sigma \), a thread running expression \( e \) may step to expression \( e' \) while changing the state to \( \sigma' \) and forking off a list of threads running expressions \( \overrightarrow{\sigma'} \). Threadpool reduction \( T; \sigma \rightarrow T'; \sigma' \) lifts head reduction to lists of concurrently running threads. We leave implicit the choice of operators \( unop \) and \( binop \), but note that they reduce according to (partial) evaluation functions and are subject to a semantic constraint; for example, the reduction rule for unary operators

\[
unop \ v; \sigma \rightarrow; \ v'; \sigma; e \quad \text{if eval } unop \ v = v'
\]

employs a partial function eval : Unop → Val → Val (which, for the robust safety theorem presented below, is assumed to send low-integrity values to low-integrity values). As is standard in semantics for languages with mutable references, locations \( \ell \) are not part of the "surface" syntax of the language; they may only arise dynamically during execution. Using compare-and-swap (CAS), one can implement locks (and other concurrency primitives). We use the symbol ‘\( \_\)’ in place of a bound variable to indicate any variable not occurring free in the scope of the binder. In addition to standard derived forms (e.g., n-ary products and let expressions), we write

\[
\lambda x. \ e \triangleq \ \text{rec } \_x. \ e
\]

\[
\lambda x_1 \cdots x_n. \ e \triangleq \ \lambda x_1. \ \cdots. \ \lambda x_n. \ e
\]

and we use pattern-matching notation to deconstruct products, rather than a series of fst and snd projections.

The one new feature in HLA, not present in Krebbers et al.’s calculus, is assertion expressions, \( \text{assert } e \), which serve to specify safety properties that should always hold dynamically. Assertions in HLA operate a little differently from assertions in languages like C or Java, where failed assertions abort execution or raise an exception. Here, \( e \) is a boolean expression. If \( e \) evaluates to \( \text{true} \), then \( \text{assert } e \) has no effect. But if \( e \) evaluates to \( \text{false} \), then \( \text{assert } e \) has the effect of irreversibly setting a “goodness bit” \( g \), which is maintained as part of the machine state, to \( \text{FAIL} \). (The bit is initially set to \( \text{OK} \).) This goodness bit simply checks whether any assertion expression has dynamically failed during execution. Ultimately, our robust safety theorem will use this goodness bit to ensure that, for properly verified code, the goodness bit will remain \( \text{OK} \) throughout execution (and hence all dynamic \( \text{assert } \) checks must succeed).

\[3\]OCPs are typically studied in the context of an event-driven programming model. We work in the more general setting of fork-based concurrency because Iris supports it natively and it does not seem to significantly affect reasoning about OCPs.

\[4\]The expression \( CAS \ f \ v_1 \ v_2 \) atomically compares the contents of location \( \ell \) to value \( v_2 \) and, if equal, writes \( v_2 \) to \( \ell \) and returns \( \text{true} \); otherwise, it returns \( \text{false} \). For simplicity, we do not restrict \( CAS \) to comparing first-order, "word-sized" values. This is unrealistic, but we use \( CAS \) in only realistic ways. (Spinlocks, in particular, are compatible with such a restriction and suffice for our purposes.)

\[5\]We define C-style assertions—which we call assumption expressions—in §3.
The reader may wonder why we employ a goodness bit instead of just saying that `assert false` gets stuck or aborts the program. The reason is simple: we wish to use OCPL to reason not only about fully verified code but also about the behavior of verified code when linked with untrusted code (that does not itself contain assertions). So it is important that we have a way of verifying Hoare triples even for code that may very well get stuck (i.e., fail to make progress) thanks to dynamic type errors introduced by the untrusted code. We nevertheless want to say that `assert` expressions in such code always succeed, so we must differentiate `assert false` from other stuck states.\footnote{Instrumenting the state with a goodness bit is not strictly necessary; for example, one could alternatively use a special machine configuration to indicate failed assertions.}

### 2.2 A program logic for reasoning about OCPs

We now present some of the essential features of OCPL, our logic for reasoning about OCPs.

**Progressive vs. non-progressive triples.** OCPL is a Hoare-style program logic derived from Iris, a modern separation logic for higher-order concurrent imperative programs.\footnote{Readers unfamiliar with separation logic can, to a first approximation, read separating conjunction \( \mathcal{P} \land \mathcal{Q} \) as conjunction \( \mathcal{P} \land \mathcal{Q} \) and magic wand \( \mathcal{P} \rightarrow \mathcal{Q} \) as implication \( \mathcal{P} \Rightarrow \mathcal{Q} \).} As such, one of the main assertions in OCPL is the *Hoare triple*, which is used to specify the behavior of expressions in terms of preconditions and postconditions. The Hoare triple \( \{ P \} e [ x . Q ] \) asserts that, assuming \( e \) is executed starting in a state satisfying the precondition \( P \), then it will execute without any dynamically failing `assert` expressions, and if it terminates with value \( v \), then the final state will satisfy the postcondition \( Q \) (which may mention the bound variable \( x \)). The *progress bit* \( p \in \{ \text{progress}, \text{noprog} \} \) indicates whether the triple is *progressive* (i.e., ensures that \( e \) does not get stuck) or *non-progressive*.\footnote{Note: this is a point of difference from the original Iris, which only supported progressive triples.} As shorthand, we will write \( \{ P \} e \{ x . Q \} \) and \( \{ P \} e \{ x . Q \} \), for progressive and non-progressive triples, respectively; the former assertion implies the latter. We often write the postcondition of a Hoare triple as \( \{ x_1 \cdots x_n . \text{ret } pat . Q \} \), which binds the \( x_i \) and specifies that any returned value will be precisely `pat`. This notation is (roughly) syntactic sugar for the post-condition \( \{ x . \exists x_1 , \ldots , x_n . x = \text{pat} * Q \} \).

Hoare triples in Iris (and thus OCPL) enjoy the standard *structural rules* for higher-order concurrent separation logic (e.g., the rule of consequence, the frame rule, and monadic bind and return) as well as rules for defining and enforcing logical *protocols* on physical and ghost state. Readers interested in these aspects of OCPL should consult Jung et al. (2015) for a high-level overview and Jung et al. (2017) for a recent account of Iris’s foundations.

In addition, OCPL enjoys a full set of *basic rules* for HLA, comprising progressive triples for reasoning about head reductions and non-progressive triples for reasoning about stuck expressions. The following three rules, for example, concern application.

\[
\begin{align*}
\text{App} & \quad \frac{(\text{rec } f \; x . \; e) \text{ closed}}{\{ P \} \ e[v/x, \text{rec } f \; x . \; e/f] [x . Q]} \\
\text{AppNRec} & \quad \frac{\neg(\exists f, x, e) \; v = \text{rec } f \; x . \; e}{\{ \top \} \; v \; v' [x . Q]} \\
\text{AppOpen} & \quad \frac{\neg(\text{rec } f \; x . \; e) \text{ closed}}{\{ \top \} \; (\text{rec } f \; x . \; e) \; v [x . Q]}
\end{align*}
\]

Rule \text{App} says that, to prove a triple for a closed function application, it suffices to show an analogous triple after $\beta$-reduction. Rules \text{AppNRec} and \text{AppOpen} give non-progressive triples to applications that, according to the operational semantics of HLA, always get stuck: as such applications never produce a value, they can be given an arbitrary postcondition \( Q \).
Robust and Compositional Verification of Object Capability Patterns

We omit analogous rules for other expressions that ignore the ambient machine state, turning now to the more interesting rules for assertions and locations.

Assertion expressions. OCPL supports the following rule for assertion expressions.

\[
\text{Assert} \quad \{P\} \ e \ \{x. \ x = \text{true} \}^p \vdash \{P\} \ \text{assert} \ e \ \{\text{ret ()}. \ Q\}^p
\]

\text{Assert} enables reasoning about successful assertions. It says that, to prove a Hoare triple for \text{assert} \ e with postcondition \( Q \) and return value \( () \), it suffices to prove that \( e \) evaluates to \text{true} with postcondition \( Q \).

High vs. low locations. OCPL divides memory locations (i.e., mutable references) into two types: high-integrity and low-integrity (or just high and low for short). High locations are locations that are private to user code, on which it may place invariants of its choosing, and to which untrusted code should not be given direct, unfettered access. Low locations are locations that may be freely shared with—and may in fact have been allocated by—untrusted code. Note that there is no distinction between high and low locations in the operational semantics of HLA; rather, this distinction is merely something we track in OCPL in order to formally specify the interface between user code and untrusted code.

Reasoning about high locations. Since high locations are locations controlled privately by verified user code, OCPL supports reasoning about them in essentially the same way as one reasons about pointers in traditional separation logic—via the classic "points-to" assertion \( \ell \leftarrow v \). This points-to assertion denotes the knowledge that location \( \ell \) currently points to value \( v \) as well as ownership of \( \ell \). What this means is that if we own \( \ell \), then no other party (e.g., neither another thread nor the current thread’s context) can be simultaneously making any assertion about \( \ell \), so we should have permission to freely read from and write to \( \ell \) without worrying about violating any other party’s reasoning. This reasoning is formalized with standard rules about high locations displayed in Fig. 2a.

Reasoning about low locations. We represent the knowledge that a location \( \ell \) is low via the assertion lowloc \( \ell \). In contrast to the points-to assertion for high locations, the assertion lowloc \( \ell \) does not denote any ownership of \( \ell \) because as soon as a location is considered safe to be shared with untrusted code, there is no way of knowing what that code will do with it. Rather, lowloc \( \ell \) merely tells us that \( \ell \) is low now and will remain low forever. Moreover, if \( \ell \) is low—and hence...
safely shareable with untrusted code—then we know that the value \( v \) it points to must also be safely shareable with untrusted code—i.e., \( v \) is a low value.

Of course, this begs the question: what is a low value? Intuitively, a low value is a value from which there is no way to extract a high location. We will make this more precise in a moment. For now, assume we have a predicate \( \text{lowval} \), which says that \( v \) is a low value.

The description of low locations given above is formalized in the proof rules of Fig. 2b. Rule \( \text{HighNotLow} \) says that a location cannot be high and low at the same time. Rules \( \text{AllocLow}, \text{LoadLow}, \) and \( \text{StoreLow} \) mimic the corresponding rules for allocation, reading, and writing of high locations, except that for low locations, we do not track the precise contents of the location \( \ell \)—we merely insist that \( \ell \) always points to a low value.

**Lifting low locations to low values.** We return now to the question of what it means for a value \( v \) to be low. Intuitively, a low value is one from which the language constructs of HLA provide no way to get direct access to any high location. This is fundamentally an extensional property, i.e., a property about the observations that a program can make (i.e., the information it can extract) when passed the value \( v \). As such, for those readers familiar with classic techniques from program semantics, it will not come as a surprise that a natural way of formally accounting for this property is via a logical relation.

Fig. 3 shows the definition of this logical relation, presented in a somewhat more general form, \( \text{lift } \Psi \ v \). Essentially, \( \text{lift} \) takes as input \( \Psi \), a predicate on locations, and lifts it to a predicate on values, with the property that \( \text{lift } \Psi \ v \) is true if \( \Psi \) holds for any location that can be extracted from \( v \). Given this definition, \( \text{lowval} \ v \) can be defined simply as \( \text{lift } \text{lowloc} \ v \)—i.e., a value is low if any location that can be extracted from it is low.

On literals and unit, \( \text{lift } \Psi \) is trivially true since no location can be extracted from them; on locations, \( \text{lift } \Psi \) is simply \( \Psi \); and on products and sums, \( \text{lift } \Psi \) is defined in the obvious recursive manner. The only interesting case is the one for functions. For a function value \( f \), we want to say that \( \text{lift } \Psi \ f \) if, whenever we apply \( f \), the resulting term only ever produces values that satisfy \( \text{lift } \Psi \). (This property on applications of \( f \) is expressed via a non-progressive Hoare triple since \( f \) may be applied to arguments constructed by untrusted code.) But what arguments should we consider when applying \( f \)? This is where the logical relation comes in: we need only consider argument values that themselves satisfy \( \text{lift } \Psi \), because ultimately we will make sure that all values passed to untrusted code satisfy \( \text{lift } \Psi \).

Note that the definition of \( \text{lift } \Psi \) in the function case is rather circular, since it quantifies over values that satisfy \( \text{lift } \Psi \). To ensure that this definition is well-founded, we rely on OCPL’s support

\[
\begin{align*}
\text{lift } \Psi \ (\text{rec } f \ x \ e) & \triangleq \forall v. \{ \text{lift } \Psi \ v \} e[\text{rec } f \ x \ e/f] \{ y. \text{lift } \Psi \ y \} \text{.} & \text{(LiftRec)} \\
\text{lift } \Psi \ \ell & \triangleq \Psi \ell & \text{(LiftLoc)} \\
\text{lift } \Psi \ \text{lit} & \triangleq \top & \text{(LiftLit)} \\
\text{lift } \Psi \ \text{()} & \triangleq \top & \text{(LiftUnit)} \\
\text{lift } \Psi \ (v_1, v_2) & \triangleq \text{lift } \Psi \ v_1 \star \text{lift } \Psi \ v_2 & \text{(LiftPair)} \\
\text{lift } \Psi \ (\text{inl } v) & \triangleq \text{lift } \Psi \ v & \text{(LiftInl)} \\
\text{lift } \Psi \ (\text{inr } v) & \triangleq \text{lift } \Psi \ v & \text{(LiftInr)} \\
\end{align*}
\]

Fig. 3. Lifting location predicates to value predicates
for guarded recursive definitions (inherited from Iris). In OCPL, a definition may be arbitrarily recursive so long as any recursive references are guarded by an occurrence of the \( \triangleright \) ("later") modality.

Under the hood, all assertions in OCPL are implicitly indexed by steps of computation—i.e., the model of OCPL is a kind of step-indexed model (Birkedal et al. 2011)—and \( \triangleright P \) means that \( P \) holds, roughly speaking, "after one step of computation". Since the fine points of the "later" modality are not particularly relevant to our main contribution, and our use of "later" is standard, we refer the reader to prior work for a more detailed presentation (Appel et al. 2007).

Using the aforementioned structural and basic rules, we can readily derive rules for working with lifted values. For example, the derived rule

\[
\text{LiftApp}
\{ \text{lift } \Psi \ u \ast \text{lift } \Psi \ u' \} \ u \ u' \ \{ x . \ \text{lift } \Psi \ x \}\]

says that one can always apply a lifted value to a lifted argument, and such applications either get stuck (the rule uses a non-progressive triple) or produce a lifted result. These derived rules help, for instance, in verifying clients of the public membrane (§5), which we specify using lifted.

To make concrete this rather abstract tour of OCPL, we return to our motivating example. To verify that \text{readonly}satisfies its specification, we have to show that the function it returns, \( \lambda \_ . \ ! \ell \), is a low value given a triple \( \{ \top \} \ ! \ell \ {x.\ \text{low}val \ x} \) for dereferencing \( \ell \). This follows easily by \text{LiftRec}.

To verify that the expression use two returns a low value, we proceed as follows. From the allocation \text{ref} 2 we obtain \( \ell \leftrightarrow 2 \) for some location \( \ell \) (by \text{Alloc}). Since we own \( \ell \), we are free to transfer ownership of \( \ell \) into a shared invariant, which in Iris (and OCPL) is written \( [\ell \leftrightarrow 2] \), stipulating that \( \ell \) will always contain 2. (Invariants like \([\ell \leftrightarrow 2] \), which are established "dynamically" in the course of a program proof and hold persistently thereafter, are a feature inherited directly from Iris (Jung et al. 2015). We omit rules for working with them here since they are orthogonal to the focus of this paper.) By \text{Load}, we easily obtain a lemma

\[
[\ell \leftrightarrow 2] \vdash \{ \top \} \ ! \ell \ {\text{ret} \ 2.\ \top}\]

stating that, under our invariant, dereferencing \( \ell \) always returns 2. To prove that the value \( w \) returned by \text{readonly} \( \ell \) is low-integrity, we apply \text{readonly}'s specification and have to show that dereferencing \( \ell \) returns a low value (which is trivial, by (†) and \text{LiftLit}). To prove that the function \text{use} \( \triangleq \lambda \_ . \ \text{assert} \ ((\! \ell) = 2) \) is low-integrity, we apply \text{LiftRec}. In this subproof, we have to prove that, under our invariant, the assertion returns a low value (trivial by \text{Assert}, (†), \text{LiftUnit}). It remains to show that the pair \((w,\ \text{use})\) is low-integrity given knowledge that both \( w \) and \text{use} are low-integrity, which is trivial by \text{LiftPair}.

### 2.3 Metatheory

OCPL inherits from Iris the basic metatheorem of adequacy for progressive triples (omitted here), which asserts that expressions verified with progressive triples make progress (i.e., do not get stuck) and that, if such expressions terminate, the postconditions of their triples actually hold. OCPL also enjoys an analogous metatheorem for non-progressive triples, which simply drops the progress guarantee. More interestingly, OCPL enjoys the following metatheorems concerning dynamically checked safety properties. The entire metatheorem has been verified formally in Coq (OCPL 2017).

**Theorem AdequacySafety.** If an expression is verified and run in a good state, we can observe that every reachable state is good (i.e., no assertions fail):

\[
\{ \top \} e \ {x.\ \top} \rho (e); (h,\ \text{OK}) \rightarrow^* T'; (h',\ g')
\]

\[
g' = \text{OK}
\]
The functionality of dynamic sealing. Morris (1973) introduced dynamic sealing to enforce data abstraction while interoperaing with untrusted, potentially ill-typed code. He stipulated a function \textit{makeseal} for generating pairs of functions (\textit{seal}, \textit{unseal}), such that (i) for every value \( v \), \textit{seal} \( v \) returns a value \( v' \) serving as an \textit{opaque, low-integrity proxy} for \( v \); and (ii) for every value \( v' \), \textit{unseal} \( v' \) returns \( v \), if \( v' \) was produced by \textit{seal} \( v \), and otherwise gets stuck. The key point is that this \textit{seal-unseal} pair supports data abstraction: the client of these functions can freely pass sealed values to untrusted code since they are low-integrity, while at the same time imposing whatever internal invariant it wants on the underlying values that they represent.
To see how this is useful, consider a simple client of dynamic sealing, namely Morris’s example of a library for integer intervals (Fig. 4). Let \([n_1, n_2]\) denote the set \([n_1, n_1 + 1, \ldots, n_2 - 1, n_2]\). The function \textit{intervals} returns several interval-manipulating routines: \textit{makeint} \(n_1 n_2\) constructs \([\min n_1 n_2, \max n_1 n_2]\), \textit{imin} \([n_1, n_2]\) returns \(n_1\), \textit{imax} \([n_1, n_2]\) returns \(n_2\), and \textit{isum} \([n_1, n_2]\) \([n_1', n_2']\) returns \([n_1 + n_1', n_2 + n_2']\).

These routines use dynamic sealing to enforce the internal data representation of intervals, namely that the interval \([n_1, n_2]\) is represented by the pair \((n_1, n_2)\), which is critical to ensuring correctness of the library. In particular, notice that \textit{seal} and \textit{unseal} are kept private to the \textit{intervals} implementation, which means it can enforce that the only values sealed with \textit{seal} are pairs \((n_1, n_2)\) representing intervals (i.e., where \(n_1 \leq n_2\)). Consequently, the \textit{imin} (resp. \textit{imax}) function can simply return the first (resp. second) component of its argument after unsealing it, because it knows that, even if the argument comes from untrusted code, \textit{so long as the unsealing succeeds}, the resulting value will be a pair where the first (resp. second) component represents the lower (resp. upper) bound of the input interval. Furthermore, a client of \textit{intervals} knows that if it applies both \textit{imin} and \textit{imax} to an arbitrary untrusted value \(v\), and both operations succeed, producing values \(v_1\) and \(v_2\), then \(v\) must indeed represent a proper interval \([n_1, n_2]\) (with \(n_1 \leq n_2\)), and the resulting values of \textit{imin} and \textit{imax} are precisely \(v\)’s lower and upper bounds (i.e., with \(v_1 = n_1\) and \(v_2 = n_2\)). In this way, dynamic sealing affords programmers the essential functionality of data abstraction, even when interfacing with untrusted code, at the cost of some simple dynamic checks at the boundaries of the abstraction.

\textit{Implementation of dynamic sealing.} We implement dynamic sealing in HLA as follows.\(^9\)

\[
\text{makeseal} \triangleq \lambda_. \ \text{let tbl} = \text{ref mapempty} \ \text{in} \\
\text{let sync} = \text{makesync()} \ \text{in} \\
\text{let seal} = \lambda x. \ \text{let} \ k = \text{ref}() \ \text{in} \ \text{sync}(\lambda_. \ \text{tbl} \leftarrow \text{mapinsertnew}(!\text{tbl})\text{ }x); k \ \text{in} \\
\text{let unseal} = \lambda k. \ \text{assume} (\text{isloc} k); \ \text{sync}(\lambda_. \ \text{maplookup}(!\text{tbl})\text{ }k) \ \text{in} \\
(\text{seal}, \text{unseal})
\]

The function \textit{makeseal} allocates a fresh lookup table \textit{tbl}, which is used to store the mapping between sealed values and their proxies, and access to which is shared by the \textit{seal} and \textit{unseal} functions that \textit{makeseal} returns. To ensure proper synchronization between table accesses during calls to \textit{seal} and \textit{unseal}, \textit{makeseal} also creates a lock via the call to \textit{makesync}. (The latter returns a higher-order function \textit{sync}, which ensures mutual exclusion among any expressions \(e_i\) run under \(\text{sync}(\lambda_. \ e_i)\). For more details on \textit{makesync}, see Appendix F.) An entry \([k \mapsto v]\) in \textit{tbl} signifies that location \(k\) is a low-integrity proxy for sealed value \(v\). To seal a value \(v\), we allocate a fresh, low location \(k\), extend the table, and return \(k\). To unseal a value \(v'\), we require that \(v'\) be a location \(k\) and then look up \(k\) in the table.

To test whether the argument to \textit{unseal} is a location, we use a primitive unary operator \textit{isloc} (with boolean return type), as well as an \textit{assumption expression}. The expression \textit{assume} \(e\) resembles a C- or Java-style assertion: it returns \textit{unit} if \(e\) evaluates to \textit{true}; otherwise, it gets stuck. It is easy to define assumption expressions using a stuck term \textit{abort}:

\[
\text{abort} \triangleq 0 0 \quad \text{assume} \ e \triangleq \text{if } e \ \text{then} () \ \text{else} \ \text{abort}
\]

\(^9\)We discuss the straightforward functions \textit{makesync}, \textit{mapempty}, \textit{mapinsertnew}, and \textit{maplookup} in Appendix F. Note that \textit{maplookup} \(f \ k\) gets stuck if location \(k\) is not in the domain of the partial function represented by value \(f\).
Data abstraction.

\[
\text{MakeSealSpec}
\quad \{ \top \} \quad \text{makeseal} \; (v_1, v_2, \text{ret} \; (v_1, v_2)) \quad \text{is seal} \; y \; v_1 \; \phi \; \ast \; \text{is un seal} \; y \; v_2 \; \phi
\]

\[
\text{SealSpec}
\quad \{ \text{is seal} \; y \; s \; \phi \; \ast \; v \} \quad s \; v \; \{ x' \} \quad \text{is sealed} \; y \; v \; x' \; \phi
\]

\[
\text{UnsealSpec}
\quad \{ \text{is un seal} \; y \; u \; \phi \; \ast \; \text{is sealed} \; y \; v \; u' \; \phi \} \quad u \; u' \; \{ \text{ret} \; v \; \top \} \quad \{ \text{is seal} \; y \; u \; \phi \} \quad u \; u' \; \{ x \} \quad \text{is sealed} \; y \; x \; u' \; \phi
\]

\[
\text{SealedInv}
\quad \text{is sealed} \; y \; u' \; \phi \; \ast \; \phi \; v
\]

\[
\text{SealedAgree}
\quad \text{is sealed} \; y \; v_1 \; \phi \; \ast \; \text{is sealed} \; y \; v_2 \; u' \; \phi \; \ast \; v_1 = v_2
\]

\[
\text{Low values.}
\quad \text{SealLow}
\quad (\forall v) \; \text{low val} \; v \; \ast \; \phi \; v
\]

\[
\text{UnsealLow}
\quad (\forall v) \; \phi \; v \; \ast \; \text{low val} \; u
\]

\[
\text{SealedLow}
\quad \text{is sealed} \; y \; v \; u' \; \phi \; \ast \; \text{low val} \; u'
\]

\[
\text{UnsealedLow}
\quad \text{is sealed} \; y \; s \; \phi \; \ast \; \text{low val} \; s
\]

\[
\text{UnsealedAnySpec}
\quad \{ \text{is sealed} \; y \; u \; \phi \; \ast \; \text{is sealed} \; y \; u' \; \phi \; \ast \; v_1 = v_2
\]

Fig. 5. Sealing interface (presupposing \( \phi \) persistent)

These satisfy the following specifications.

\[
\begin{array}{ll}
\text{Abort} & \{ \top \} \quad \text{abort} \; \{ x \; Q \} \\
\text{Assume} & \{ P \} \quad e \; \{ x \; x = \text{true} \; \rightarrow \; Q \} \\
\end{array}
\]

\text{Abort} follows trivially from \text{AppNrec} and gives \text{abort} a non-progressive triple because in fact \text{abort} always gets stuck and never produces a value; correspondingly, it can also be given an arbitrary postcondition \( Q \). Compared to \text{Assert} in §2.2, \text{Assume} (i) uses non-progressive triples since \text{assume} \( e \) may get stuck, and (ii) only requires the postcondition \( Q \) to be shown to hold under the assumption that \( e \) returns true.

Specifying dynamic sealing. The motivation for dynamic sealing given above was completely informal. Let us now see how OCPL lets us formalize it.

We specify dynamic sealing in Fig. 5. Rule \text{MakeSealSpec} for allocating a sealer-unsealer pair ties the two functions it returns to an abstract (logical) name \( y \) and representation invariant \( \phi \). In particular, the assertions\(^{10}\) \text{is seal} \; y \; v_1 \; \phi \; and \text{is un seal} \; y \; v_2 \; \phi \; represent knowledge that \( v_1 \) and \( v_2 \) are the seal and unseal functions associated with the sealer-unsealer pair named \( y \) with representation invariant \( \phi \). Similarly, the assertion \text{is sealed} \; y \; v \; u' \; \phi \; represents knowledge that \( u' \) is a low-integrity proxy for value \( v \) (for the sealer-unsealer pair named \( y \) with representation invariant \( \phi \)). This relation is functional in \( v' \) —i.e., unsealing the same \( v' \) twice will produce the same \( v \) (\text{SealedAgree})—and implies \( \phi \; v' \) —i.e., that the unsealed value \( v \) must satisfy the representation invariant (\text{SealedInv}). Conversely, to seal a value \( v \), one is required to prove \( \phi \; v \) (\text{SealSpec}). The progressive triple \text{UnsealSpec} is suitable for use when one already knows that \( v \) and \( u' \) are related; it returns \( v \). The non-progressive triple \text{UnsealedAnySpec} returns the value \( v \) related to \( u' \), or gets stuck.

Finally, the three rules at the bottom of Fig. 5 concern the lowval relation. Rule \text{SealedLow} says that all sealed values are low values—indeed, the whole point of using sealing is to produce low-integrity proxy values—while rules \text{SealLow} and \text{UnsealLow} say when the seal and unseal functions themselves

\(^{10}\)In our specifications, predicates whose names start with “is” produce, when fully applied, persistent propositions (and are themselves called persistent predicates). Persistent propositions may be regarded as intuitionistic (technically, they may be freely duplicated). A sealer-unsealer pair’s representation invariant \( \phi \) must be persistent.
Properties.

\[\text{MinSpec}\]
\[
\{\text{isinterval } y \ n_1 \ n_2 \ \text{i} \ \text{imin} \ i \ | \ \text{ret} \ n_1. \ \top\}
\]

\[\text{IntervalInv}\]
\[
isinterval y \ n_1 \ n_2 \ i \ + \ n_1 \leq n_2
\]

\[\text{IntervalAgree}\]
\[
isinterval y \ n_1 \ n_2 \ i \ + \ \text{isinterval } y \ n_1' \ n_2' \ i \ + \ n_1 = n_1' \ n_2 = n_2'
\]

Non-progressive triples.

\[\text{MinAnySpec}\]
\[
\{\text{ismin } y \ \text{imin} \ v \ n_1 \ n_2, \ \text{ret} \ n_1. \ \text{isinterval } y \ n_1 \ n_2 \ v\}
\]

\[\text{MaxAnySpec}\]
\[
\{\text{ismax } y \ \text{imax} \ v \ n_1 \ n_2, \ \text{ret} \ n_2. \ \text{isinterval } y \ n_1 \ n_2 \ v\}
\]

Low values.

\[\text{IntervalLow}\]
\[
isinterval y \ n_1 \ n_2 \ i \ + \ \text{lowval} \ i
\]

\[\text{MinLow}\]
\[
ismin y \ \text{imin} \ + \ \text{lowval} \ \text{imin}
\]

\[\text{MaxLow}\]
\[
ismax y \ \text{imax} \ + \ \text{lowval} \ \text{imax}
\]

Fig. 6. Intervals interface (selected rules—see Appendix C)

can be considered low values. Intuitively, seal and unseal can be considered low values whenever doing so does not either (i) violate the representation invariant \( \phi \) or (ii) result in high-integrity values flowing to a low context. Thus, seal is low so long as any value that is passed to it (i.e., any low value) satisfies \( \phi \), and unseal is low so long as any value returned by it (i.e., any value that satisfies \( \phi \)) is low. (We show how to gainfully employ low seal and unseal functions in Appendix B.)

The specification for dynamic sealing from Fig. 5 can be straightforwardly used to derive an analogous specification for the interval-manipulation routines generated by \textit{intervals}, as we described above. In particular, by choosing the representation invariant \( \phi \ v \ \triangleq \ \exists n_1, n_2. \ v = (n_1, n_2) \ * \ n_1 \leq n_2 \), we can easily verify the rules shown in Fig. 6 (the full specification for \textit{intervals} is given in Appendix C). The assertion isinterval \( y \ n_1 \ n_2 \ i \) represents knowledge that value \( i \) represents interval \([n_1, n_2]\) in the instance of the intervals library with abstract name \( y \). A value may represent at most one interval (\textit{IntervalAgree}). A set of progressive triples (e.g., \textit{MinSpec}) account for our informal specification and are suited to reasoning about values known to represent intervals. A set of non-progressive triples are suited to reasoning about values that \textit{might} represent intervals. Rule \textit{MinAnySpec}, for example, says that for any value \( v \), \( \text{imin} \ v \) might get stuck, but if it terminates with value \( \hat{v} \), then \( v \) represents some interval with minimum \( n_1 \) and \( \hat{v} = n_1 \). Lastly, there are rules (e.g., \textit{IntervalLow} and \textit{MinLow}) expressing the fact that values representing intervals, and the interval routines themselves, may be safely shared with untrusted code. Crucial to the soundness of these rules is the fact that the seal function used internally by \textit{intervals} is \textit{not} shared with untrusted code (i.e., not low): according to rule \textit{SealLow}, in order for seal to be treated as low, the representation invariant \( \phi \) would have to be satisfied by all low values, which it clearly is not.

With the spec for \textit{intervals} in hand, we can readily prove that the following simple but illustrative client of \textit{intervals} is robustly safe.

\[
\text{client} \ \triangleq \ \text{let} \ cap = \text{intervals()} \ \text{in}
\]
\[
\text{let} \ (\text{makeint}, \ \text{imin}, \ \text{imax}, \ \text{isum}) = \text{cap} \ \text{in}
\]
\[
\text{let} \ \text{check} = \lambda j. \ \text{assert} \ (\text{imin} \ j \ \leq \ \text{imax} \ j) \ \text{in}
\]
\[(\text{check}, \ \text{cap})
\]
The expression constructs an instance \textit{cap} of the intervals library, along with a function \textit{check} which takes an arbitrary value \(j\) as its argument and asserts that \(imin j\) is no greater than \(imax j\). Intuitively, even if \textit{client} is shared with untrusted code, this assertion must always succeed, because if the applications of \(imin\) and \(imax\) do not get stuck, it means that \(j\) is a proper interval value, whose lower bound is \(\leq\) its upper bound.

Formally, this is guaranteed as follows. First, using the spec for \textit{intervals}, we can prove \[\{\top\} \textit{client} \{x. \text{lowval } x\}\] The key step in the proof involves showing that the assertion in \textit{check} succeeds for arbitrary \(j\), which follows directly from \texttt{MinMaxSpec}, \texttt{MaxAnySpec}, \texttt{IntervalAgree}, and \texttt{IntervalInv}. Second, we appeal to Theorem RobustSafety from §2.3, which implies that the assertion in \textit{check} will not fail, even when \textit{client} is linked with untrusted code.

4 CARETAKER

Next, we consider another well-known OCP, the caretaker pattern (Miller and Shapiro 2003; Miller 2006). This OCP allows verified (trustworthy) code to grant untrusted code access to a high-integrity resource (a high-integrity location or an API that modifies high-integrity locations), and subsequently disable or enable the access at any time. When the caretaker is enabled, untrusted code can access the resource; when the caretaker is disabled, untrusted code cannot access the resource and the verified code has full control over the resource. The caretaker pattern is useful when the verified code wants to ensure that the untrusted code can access the resource only while some invariant holds. Disabling the caretaker allows the verified code to temporarily break the invariant, secure in the knowledge that untrusted code won’t be able to access the resource until the caretaker is re-enabled.

In this section, we first implement a caretaker for APIs in HLA, and specify and verify it in OCPL. Then, we use the API caretaker to implement a second caretaker for locations and, again, specify and verify it. Finally, we present a simple client for the location caretaker and establish that it is robustly safe.

API caretaker. The API caretaker allows verified code to share revokable access to any set of functions, which may have side-effects.

\[
\begin{align*}
\text{makecaretaker} & \triangleq \lambda_. \text{let } enabled = \text{ref false in let } sync = \text{makesync()} \text{ in (sync, enabled)} \\
\text{wrap} & \triangleq \lambda (sync, enabled) \ f \ x. \ \text{sync}(\lambda_. \ \text{assume} (! enabled); f \ x) \\
\text{enable} & \triangleq \lambda (sync, enabled). \ \text{sync}(\lambda_. \ \text{enabled} \leftarrow \text{true}) \\
\text{disable} & \triangleq \lambda (sync, enabled). \ \text{sync}(\lambda_. \ \text{enabled} \leftarrow \text{false})
\end{align*}
\]

The function \textit{makecaretaker} returns a new caretaker \(ct\), which comprises a fresh lock \(sync\), whose purpose is described soon, and a fresh boolean reference \(enabled\). The caretaker is disabled (enabled) when \(enabled\) is \texttt{false} (true). The functions \textit{enable ct} and \textit{disable ct} enable and disable the caretaker \(ct\) by setting \(enabled\) appropriately. The function \textit{wrap ct f} wraps the function \(f\) in the caretaker \(ct\), returning a function that behaves exactly like \(f\) when \(ct\) is enabled, and gets stuck when \(ct\) is disabled. Additionally, accesses to all functions wrapped in \(ct\) are serialized using the lock \(sync\).

To use this interface, verified code creates a caretaker \(ct\) and holds it privately. It can then wrap any number of API functions using \textit{wrap} and disclose the wrapped functions to untrusted code. The untrusted code’s access to all those functions can be simultaneously disabled and enabled by calling \textit{disable ct} and \textit{enable ct}, respectively.
**EnabledExclusive**

\[
\text{enabled } \gamma \ b_1 \ast \text{ enabled } \gamma \ b_2 \vdash \bot
\]

**MakeCaretakerSpec**

\[
\{ \top \} \ \text{makecaretaker}() \ \text{ct } \ \text{iscaretaker } \gamma \ \text{ct } \ R \ast \text{ enabled } \gamma \ \text{false}
\]

**CanWrap**

\[
\text{canwrap } f \ R \ \triangleq \ \forall \ u. \ \{ \text{lowval } v \ast R \} \ f \ \{ x'. \ \text{lowval } x' \ast R \},
\]

**WrapSpec**

\[
\{ \text{iscaretaker } \gamma \ \text{ct } \ R \ast \text{canwrap } f \ R \} \ \text{wrap } f \ \{ f'. \ \text{lowval } f' \}
\]

**EnableSpec**

\[
\{ \text{iscaretaker } \gamma \ \text{ct } \ R \ast \text{ enabled } \gamma \ \text{false } \ast R \} \ \text{enable } \text{ct} \ \{ \text{ret }(). \ \text{enabled } \gamma \ \text{true} \}
\]

**DisableSpec**

\[
\{ \text{iscaretaker } \gamma \ \text{ct } \ R \ast \text{ enabled } \gamma \ \text{true } \ast R \} \ \text{disable } \text{ct} \ \{ \text{ret }(). \ \text{enabled } \gamma \ \text{false } \ast R \}
\]

Fig. 7. API caretaker interface

**API caretaker specification.** The goal of using the API caretaker is to ensure that untrusted code can access wrapped functions only while some logical resource or invariant R holds on the heap. Our specification of the API caretaker in Fig. 7 formalizes this intuition. The assertion \text{iscaretaker } \gamma \ \text{ct } \ R \ represents knowledge that value \text{ct} is a caretaker with (logical) name \gamma whose wrappers allow (untrusted code) access to wrapped functions only when \text{R} holds. The assertion \text{enabled } \gamma \ b \ represents exclusive ownership of the (verified code’s) right to enable and disable all wrappers created using the caretaker named \gamma as well as knowledge that the caretaker is currently enabled (if \ b = \text{true}) or disabled (otherwise).

When applying **MakeCaretakerSpec**, one may pick an arbitrary invariant \text{R}. The spec returns a disabled caretaker tied to \text{R} and a fresh name \gamma. To enable \text{ct}, the (verified) code calling **enable** must establish that \text{R} holds (rule **EnableSpec**). Dually, rule **DisableSpec** says that disabling \text{ct} provides ownership of \text{R} to the caller (verified code).

Rule **WrapSpec** is at the heart of our specification. It says that to wrap a function \text{f} using \text{ct}, one has to prove that \text{f}, when applied to any low value \text{v} (possibly provided by the untrusted code), preserves the invariant \text{R} and returns a low value. We capture this client responsibility in the precondition (\text{canwrap } f \ R) on \text{wrap}. So, any wrapped function \text{f} can assume \text{R} in its precondition but it must re-establish \text{R} if it terminates (note that the triple for \text{f } v \text{ in rule **CanWrap**} is non-progressive). This ensures that while \text{ct} is enabled, calls to the wrapped functions by the untrusted code preserve \text{R}. While \text{ct} is disabled, verified code can change the state to break \text{R} but, prior to re-enabling \text{ct}, it must re-establish \text{R}.

The implementation of the API caretaker verifies against this specification. We note that the specification and verification of the API caretaker closely resemble similar efforts for specifying and verifying locks in concurrent programs (Dinsdale-Young et al. 2010). Based on this observation, we offer a shorter implementation of the API caretaker in Appendix D. This implementation uses lock release and acquire instead of the boolean reference **enabled** to enable and disable the caretaker.

The API caretaker is a very general OCP. It can be used to build caretakers for other kinds of resources. As an instance, we show next how to build a caretaker for locations using the API caretaker.
enabled γ b₁ = enabled γ b₂ ⊢ ⊥

\[ \text{isRmon} \quad \text{isrmon } f \Psi \triangleq \forall v. \{ \Psi v \} \ f \nu \{ x' \cdot \Psi v \ast \text{lowval } x' \} \]

\[ \text{isWmon} \quad \text{iswmon } f \Psi \triangleq \forall v. \{ \text{lowval } v \} \ f \nu \{ x' \cdot \Psi x' \} \]

\[ \text{makelocct } f, f_w, \ell \]

\[ \{ \text{ct } \gamma v, \text{ret } (\text{ct }, v) \}. \text{islocct } \gamma \ ct \ \ell \Psi \ast \text{enabled } \gamma \ \text{false } \ast \text{lowval } v \}

\[ \text{LocCtEnableSpec} \quad \{ \text{islocct } \gamma \ ct \ \ell \Psi \ast \text{enabled } \gamma \ \text{false } \ast \ell \leftarrow v \ast \Psi v \} \ \text{enable } \text{ct } \{ \text{ret } (). \ \text{enabled } \gamma \ \text{true} \}

\[ \text{LocCtDisableSpec} \quad \{ \text{islocct } \gamma \ ct \ \ell \Psi \ast \text{enabled } \gamma \ \text{true} \} \ \text{disable } \text{ct } \{ v, \text{ret } (). \ \text{enabled } \gamma \ \text{false } \ast \ell \leftarrow v \ast \Psi v \}

\[ \text{Fig. 8. Location caretaker interface} \]

**Location caretaker.** The location caretaker, defined by the function `makelocct` below, facilitates *revokable and mediated* read and write access to a location. Specifically, it ensures that the location is accessed by untrusted code only while some stipulated predicate \( \Psi \) holds on the location’s contents.

\[ \text{makelocct } \triangleq \lambda \text{rmon } w\text{mon } r. \ \text{let } \text{ct } = \text{makecaretaker } () \ \text{in} \]

\[ \text{let read } = \text{wrap } \text{ct } (\lambda x. \ \text{rmon } (! r)) \ \text{in} \]

\[ \text{let write } = \text{wrap } \text{ct } (\lambda x. \ r \leftarrow (\text{wmon } x) \ \text{in} \]

\[ (\text{ct }, (\text{read }, \text{write})) \]

The function `makelocct`, when applied to a function \( f_r \), a function \( f_w \), and a location \( \ell \), returns a pair \((\text{ct }, (\text{read }, \text{write}))\). The API caretaker \( \text{ct } \) controls whether or not the functions \{read, write\} are enabled. When \( \text{ct } \) is enabled, the function \( \text{read } \) reads \( \ell \) and filters the read value through \( f_r \). Similarly, \( \text{write } \) writes \( \ell \) after filtering the value to be written through \( f_w \). Thus, \( f_r \) and \( f_w \) mediate reads from and writes to \( \ell \). Consider, for example, an application \( \text{write } v \), representing a request to write value \( v \) to location \( \ell \). If \( \text{ct } \) is enabled, the application reduces to \( \ell \leftarrow (f_w v) \) and \( f_w \) gets to decide what is written.

To use the location caretaker on a location \( \ell \), verified code invokes `makelocct f_r f_w \ell` with appropriate \( f_r \) and \( f_w \). It holds \( \ell \) and the returned API caretaker \( \text{ct private} \), but passes \( \text{read} \) and \( \text{write} \) to untrusted code. It can then enable and disable access to the protected location using the calls `enable ct` and `disable ct`, respectively. While \( \text{ct } \) is enabled, all reads and writes to \( \ell \) are mediated by \( f_r \) and \( f_w \), respectively.

**Location caretaker specification.** Our location caretaker specification (Fig. 8) ensures that the value in the protected location \( \ell \) satisfies some stipulated predicate \( \Psi \) whenever the location is accessible to the untrusted code. The proposition `islocct γ ct ℓ Ψ` means that value \( \text{ct } \) is an API caretaker for location \( ℓ \) with logical name \( γ \) and predicate \( Ψ \) governing the contents of location \( ℓ \) when the caretaker is enabled. Logical names and propositions of the form enabled \( γ b \) serve the same purpose as in the API caretaker interface.

The triple `makelocctSpec` specifies the function `makelocct` and imposes conditions on the function’s first two arguments, \( f_r \) and \( f_w \). Since \( f_r \) receives a value read from the protected location, \( f_r \)'s input can be assumed to satisfy \( Ψ \). Since the output of \( f_r \) will be returned to untrusted code, that output
must be low. Hence, \( f_r \) must transform values satisfying \( \Psi \) to low values. Dually, since \( f_w \) receives a value from untrusted code and its output is written to the protected location, \( f_w \) must transform low values to values satisfying \( \Psi \). The postcondition for \texttt{makelocct} says that the returned functions \texttt{read}, \texttt{write} are low, so they can be passed to untrusted code safely. The precondition of \texttt{enable} insists that \( \ell \) point to a value \( v \) satisfying \( \Psi \) before the caretaker is enabled (rule \texttt{LocCtEnableSpec}). Dually, the postcondition of \texttt{disable} reveals the fact that \( \ell \) points to a value \( v \) satisfying \( \Psi \) (rule \texttt{LocCtDisableSpec}).

The location caretaker’s code satisfies this specification, assuming the specification of the API caretaker (Fig. 7). To illustrate how programs may use a location caretaker, and how such programs may be verified, we consider the following very simple program, \texttt{client}.

\[
\text{asserteven} \triangleq \lambda n. \text{assert (even } n) \}; n
\]
\[
\text{assumeeven} \triangleq \lambda n. \text{assume (even } n) \}; n
\]
\[
\text{client} \triangleq \text{let } r = \text{ref } 0 \text{ in}
\]
\[
\text{let } (ct, w) = \text{makelocct asserteven assumeeven } r \text{ in}
\]
\[
\text{enable } ct;
\]
\[
\text{let } sync = \text{makesync } () \text{ in}
\]
\[
\text{let } use = \text{sync } (\lambda _. \text{disable } ct; \text{assert even } (!r);)
\]
\[
\quad r \leftarrow 1; r \leftarrow 0; \text{enable } ct
\]
\[
\text{in } (use, w)
\]

The expression \texttt{client} constructs a reference cell \( r \) (initially containing 0) and constructs and immediately enables a location caretaker \( ct \) serving access to \( r \). The goal is to maintain the invariant that untrusted code only sees even numbers in \( r \). To ensure this invariant, the location caretaker \( ct \) is created with the write filter \( f_w = \text{assumeeven} \), which gets stuck unless the value being written is an even number. The read monitor, \( f_r = \text{asserteven} \), checks that the value being read from \( r \) is indeed even. Additionally, \texttt{client} exposes to its (untrusted) context a function \texttt{use} that (within the scope of an exclusive lock, \texttt{sync}) locally disables the caretaker and temporarily breaks the invariant by writing 1 to \( r \).

Using the specification of the location caretaker, we prove that \texttt{client} satisfies the triple

\[
\{ \top \} \text{client} \{ \}\{ x. \text{lowval } x\}
\]

By \textbf{Theorem RobustSafety}, \texttt{client} is robustly safe. This means that irrespective of how the (untrusted) context calls \texttt{use} and the wrapped read/write interface \( w \), it can never observe a non-even number in \( r \) (else, the assertion in \texttt{asserteven} would fail).

\section{5 Membrane}

When verified (trustworthy) and untrusted code interoperate, a general concern is that values passed from verified to untrusted code may accidentally reveal high-integrity locations. The membrane pattern (Miller 2006; Miller et al. 2008; Google, Inc. 2015) was designed to offset this possibility. A membrane is a bidirectional transformation on values passing from verified to untrusted code and vice versa. In the verified-to-untrusted direction, it sanitizes values to make the extraction of high-integrity locations impossible. A trivial way to implement the membrane is to use the \texttt{seal} function of \S 3, which replaces all values with proxies. However, these proxies cannot be \texttt{used} by untrusted code, other than by passing them back to the verified code. To address this shortcoming, the membrane pattern uses a \textit{deep inspection} of values to selectively hide high-integrity locations.
inside values, while preserving their overall structure (pairs map to pairs, functions map to functions, etc.). This allows the untrusted code to use the transformed values, while still ensuring that it won’t get access to high-integrity locations.

While the specifics of the transformation depend on the language and security objectives of the application, a common core of all membranes is a function that lifts a use case-specific bidirectional transformation on locations to a bidirectional transformation on values. We first define this function, \textit{membrane}, in HLA below, show its OCPL specification and verify the function. Then, we use this function to implement a specific membrane that we call the \textit{public membrane}. This membrane replaces each high-integrity location nested in a value with a fresh \textit{shadow} location and provides the verified code special functions to inspect those shadow locations. This is similar to how the membrane pattern is implemented in Google’s Caja, a JavaScript library for securing communication between mutually distrusting domains in web applications (Miller et al. 2008; Google, Inc. 2015).

\textbf{Membrane code.} The function \textit{membrane} that forms the common core of the membrane pattern is defined below. This is a higher-order function that takes as its first two arguments two other functions, \textit{locout} and \textit{locin}, both of type \textit{Loc} \rightarrow \textit{Val}. The function \textit{locout} defines how locations crossing from verified to untrusted code are transformed by the membrane. Dually, the function \textit{locin} defines how locations crossing in the other direction, from untrusted to verified code, are transformed. Given these two arguments, \textit{membrane locout locin} is a function of type \textit{Val} \rightarrow \textit{Val} that transforms values passing from verified to untrusted code. The function for transforming values passing in the other direction—from untrusted to verified code—is obtained by reversing the arguments to \textit{membrane}, \textit{i.e.}, it is \textit{membrane locin locout}.

\begin{verbatim}
membrane ≜ rec memb locout locin x.
    let wrap = memb locout locin in
    if isfun x then let unwrap = memb locin locout in
       \lambda y. wrap(x(unwrap y))
    else if isloc x then locout x
    else if islit x then x
    else if x = () then()
    else if ispair x then (wrap(fst x), wrap(snd x))
    else ifinl x as inl x' ⇒ inl(wrap x')
    else ifinr x as inr x' ⇒ inr(wrap x')
    else assert false // i.e., not reached
\end{verbatim}

Internally, \textit{membrane} recurses on the structure of the value \textit{x} being transformed. When the value is a location, the result is simply \textit{locout x}. When the value is a literal, it is returned immediately. When the value is a pair, \textit{membrane} recurses on the two components of the pair. The interesting case arises when the value \textit{x} is a function. In this case, \textit{membrane} returns a function, which when applied (by the untrusted code), first recursively applies the membrane to the untrusted argument \textit{y}, then applies the given function \textit{x} to the transformed argument, and then re-applies the membrane recursively to the result. Importantly, the membrane is applied in the untrusted-to-verified direction to the function argument \textit{y} and in the verified-to-untrusted direction to the result of the function. Technically, the function \textit{x} is transformed to \textit{wrap} \circ \textit{x} \circ \textit{unwrap}, where \textit{wrap} \triangleq \textit{memb locout locin} and

\footnote{Google’s Caja partly automates the invocation of the membrane transformation by rewriting untrusted code. The automation is orthogonal to the specification and verification of the membrane pattern itself, so we do not model it here.}
ismon \( p \cdot v \cdot \Psi_1 \cdot \Psi_2 \triangleq \forall a. \{ \Psi_1 a \cdot v \cdot a' \cdot \Psi_2 a' \}_p \\
MembraneSpec \\
\{ \text{ismon} \; p \; \text{locout} \; \Psi_1 \; \Psi_2 \; * \; \text{ismon} \; p' \; \text{locin} \; \Psi_2 \; \Psi_1 \} \\
\text{membrane} \; \text{locout} \; \text{locin} \\
\{ \text{w. ismon} \; p \; w \; \text{(lift} \; \Psi_1 \; \text{)(lift} \; \Psi_2 \}) \}

Fig. 9. Membrane interface (presupposing \( \Psi_1, \Psi_2 \) persistent)

unwrap \( \triangleq \; \text{memb locin locout} \) are recursive instantiations of \( \text{membrane} \) in the verified-to-untrusted and untrusted-to-verified directions, respectively.

Membrane specification. Our specification of \( \text{membrane} \) formalizes the intuition that \( \text{membrane} \) lifts transformations on locations to transformations on values. The specification is shown in Fig. 9. The defined predicate \( \text{ismon} \; p \; v \; \Psi_1 \; \Psi_2 \) means that \( v \) is a function that transforms values satisfying the predicate \( \Psi_1 \) to values satisfying the predicate \( \Psi_2 \) (the progress bit \( p \) is needed for technical reasons, that readers may ignore). The specification of \( \text{membrane} \) says that if \( \text{locout} \) transforms locations satisfying \( \Psi_1 \) to values satisfying \( \Psi_2 \) and \( \text{locin} \) does the reverse, then \( \text{membrane} \; \text{locout} \; \text{locin} \) transforms values satisfying lift \( \Psi_1 \) to values satisfying lift \( \Psi_2 \), where lift is the predicate transformer defined in Fig. 3. Hence, \( \text{membrane} \) really “lifts” the transformation on locations to a transformation on values in a precise technical sense.

This specification of \( \text{membrane} \) is very general, since it holds for any predicates \( \Psi_1 \) and \( \Psi_2 \). In any use of \( \text{membrane} \), these predicates can be instantiated to match what the arguments \( \text{locout} \) and \( \text{locin} \) do. (Later, we show a specific instantiation with \( \Psi_2 = \text{lowloc} \).)

Despite the specification’s generality, \( \text{membrane} \) is easily verified against it. In the recursive cases, the proof obligations match the inductive hypotheses precisely because of the concordance between the pattern matches in the definitions of lift (Fig. 3) and \( \text{membrane} \).

Public membrane. Next, we describe how the function \( \text{membrane} \) can be used to construct a specific membrane, which we call the \( \text{public membrane} \). This membrane is similar to the membrane used in Google’s Caja library (Miller et al. 2008; Google, Inc. 2015). The public membrane maintains a unique low-integrity shadow location for every high-integrity location that the verified code declares as important. When values cross from verified to untrusted code, all nested high-integrity locations in them are effectively “replaced” with the corresponding low-integrity shadows (this transformation is implemented using \( \text{membrane} \)). Thus, the untrusted code only sees shadow locations, not the high-integrity locations. Additionally, the public membrane provides the verified code special functions, \( \text{shadowread} \) and \( \text{shadowwrite} \), to read and write the contents of the shadow locations.

To understand why this is useful, consider a library that allocates an integer reference \( \ell \), and shares it with (untrusted) clients as an I/O buffer. Clients are expected to write only positive integers to \( \ell \), although the library does not strictly require this and the library’s algorithms can execute safely even if the integer is not positive. Over time, many clients of this library have been written. Now, suppose that the library is updated to use different algorithms that \emph{really require} \( \ell \) to \emph{always be positive} (else they crash). The obvious way to do this would be to rewrite the library to hold \( \ell \) private, and to export two closures that read and write \( \ell \), the latter only after checking that the value being written is positive. However, this change \emph{breaks compatibility with all existing clients}, since they must now be rewritten to invoke the new closures to access \( \ell \). The public membrane offers a general solution to this problem. Rather than export closures, the library can deploy a
\[
\text{makepub} \triangleq \lambda_. \begin{array}{l}
\text{let tbl = ref bijempty in} \\
\text{let sync = makesync () in} \\
\text{(sync, tbl)}
\end{array}
\]

\[
\text{pubunwrap} \triangleq \lambda m. \begin{array}{l}
\text{membrane (pubout m) (pubout m)}
\end{array}
\]

\[
\text{pubref} \triangleq \lambda m x_1. \begin{array}{l}
\text{let r_1 = ref x_1 in} \\
\text{let r_2 = ref (pubwrap m x_1) in} \\
\text{let (sync, tbl) = m in} \\
\text{sync(\__, bijlookup (.tbl r_1))} \\
\text{sync(\__, bijlookup (bijinvert (!tbl) r_2))}
\end{array}
\]

\[
\text{pubout} \triangleq \lambda (\text{sync, tbl}) r_1. \begin{array}{l}
\text{sync(\__, bijlookup (!tbl) r_1)}
\end{array}
\]

\[
\text{pubin} \triangleq \lambda (\text{sync, tbl}) r_2. \begin{array}{l}
\text{sync(\__, bijlookup (bijinvert (!tbl)) r_2)}
\end{array}
\]

\[
\text{shadowread} \triangleq \lambda m. \text{pubunwrap m (! (pubout m))}
\]

\[
\text{shadowwrite} \triangleq \lambda m x. \text{pubout m r} \leftarrow \text{pubwrap m x}
\]

Fig. 10. Public membrane implementation

public membrane and declare \( \ell \) as high-integrity. The membrane consistently replaces \( \ell \) with a low-integrity shadow, say \( \ell' \), for the library’s clients. \textit{Importantly, the library’s clients don’t have to change.} After a client updates \( \ell' \) (believing that it updated \( \ell \)), the library can access \( \ell' \) using \texttt{shadowread} and copy it to \( \ell \) if the updated value is a positive integer. Additionally, whenever the library updates \( \ell \) internally, it can also copy the update to \( \ell' \) using \texttt{shadowwrite}. This way, the library can maintain its new invariant and retain complete compatibility with existing clients.

The functions comprising the public membrane are listed in Fig. 10.\footnote{We discuss the straightforward functions \texttt{makesync}, \texttt{bijempty}, \texttt{bijinsertnew}, \texttt{bijlookup}, and \texttt{bijinvert} in Appendix F.} The expression \texttt{makepub()} creates a public membrane (consistently denoted \( m \)), which comprises a pair \((\text{sync, tbl})\) where the reference cell \( tbl \) contains a finite partial bijection on locations protected by (the lock buried in) \texttt{sync}. When a public membrane’s table sends location \( \ell_1 \) to \( \ell_2 \), we say that location \( \ell_1 \) is a (high-integrity) private location and location \( \ell_2 \) its (low-integrity) shadow location. The functions \texttt{pubout} and \texttt{pubin} transform between private and shadow locations by consulting the table. Both functions get stuck on locations not in the table. The functions \texttt{pubwrap} and \texttt{pubunwrap} lift these to transformations between values by applying the function \texttt{membrane}. Effectively, \texttt{pubwrap} replaces all private locations in the value passed to it with their shadow locations and \texttt{pubunwrap} does the opposite.

The expression \texttt{pubref m v_1} constructs a new private location initially containing value \( v_1 \). In addition, it allocates a shadow location (called \( r_2 \) in the code) that initially contains the low-integrity counterpart of \( v_1 \). It stores an association between the two locations in \( m \)’s table. The shadow location \( r_2 \) can be retrieved from the table by applying either \texttt{pubout} or \texttt{pubwrap} to the private location returned by \texttt{pubref}.

The function \texttt{shadowread} inspects a private location’s shadow location, converting its low-integrity contents into a high-integrity value using \texttt{pubunwrap}, whereas \texttt{shadowwrite} updates a private location’s shadow location after converting the given value to a low-integrity value using \texttt{pubwrap}.

To use this public membrane interface, the verified code first creates a new public membrane using \texttt{makepub}. Subsequently, it allocates private locations that \textit{might} flow to untrusted code using the function \texttt{pubref} instead of the language construct \texttt{ref}. Before sending any value to untrusted code, it invokes \texttt{pubwrap} to replace all nested private locations with their shadows. Dually, after receiving any value from untrusted code, it applies \texttt{pubunwrap} to replace nested shadow locations with corresponding private locations. At any point, the verified code can access shadow locations.
The assertion \( \text{isprivval} \gamma \nu \triangleq \text{lift} (\text{isprivloc} \gamma) \nu \) and \( \text{isprivloc} \gamma \nu \triangleq \text{makepub} () \{ m \gamma, \text{ret} m. \text{ismembrane} \gamma m \}

\[
\text{PubAllocSpec} \quad \{ \text{ismembrane} \gamma m \ast \text{isprivval} \gamma \nu \} \text{pubref} m \nu \{ \ell, \text{ret} \ell. \text{isprivloc} \gamma \ell \ast \ell \twoheadrightarrow \nu \}
\]

\[
\text{PubWrapSpec} \quad \{ \text{ismembrane} \gamma m \ast \text{isprivval} \gamma \nu \} \text{pubwrap} m \nu \{ x_2. \text{lowval} x_2 \}
\]

\[
\text{PubUnwrapSpec} \quad \{ \text{ismembrane} \gamma m \ast \text{lowval} v_2 \} \text{pubunwrap} m v_2 \{ x_1. \text{isprivval} \gamma x_1 \}
\]

\[
\text{ShadowReadSpec} \quad \{ \text{ismembrane} \gamma m \ast \text{isprivloc} \gamma \ell \} \text{shadowread} m \ell \{ x. \text{isprivval} \gamma x \}
\]

\[
\text{ShadowWriteSpec} \quad \{ \text{ismembrane} \gamma m \ast \text{isprivloc} \gamma \ell \ast \text{isprivval} \gamma \nu \} \text{shadowwrite} m \ell \nu \{ \text{ret} () \}. \top
\]

Fig. 11. Public membrane interface

using \text{shadowread} and \text{shadowwrite}. Appendix E shows an example of how the public membrane is used.

**Public membrane specification.** The OCPL specification of the public membrane is shown in Fig. 11. The assertion \( \text{ismembrane} \gamma m \) associates a public membrane \( m \) to a logical name \( \gamma \). The abstract predicate \( \text{isprivloc} \gamma \ell \) represents knowledge that location \( \ell \) is a private location for the membrane named \( \gamma \), i.e., \( \ell \) was previously allocated using \( \text{pubref} \) on the public membrane named \( \gamma \). The assertion \( \text{isprivval} \gamma \nu \) is defined as \( \text{lift} (\text{isprivloc} \gamma) \nu \) by rule \text{IsPrivval}. In particular, the assertion means that any location “extractable” from \( \nu \) is in the domain of \( \gamma \)’s table. We call such values \( \nu \) private values.\(^\text{13}\)

The specification rules are straightforward, so we explain only some salient points here. The triple for \( \text{pubref} \) (rule \text{PubAllocSpec}) says that the application \( \text{pubref} m \nu \) returns a location \( \ell \twoheadrightarrow \nu \) together with the knowledge that \( \ell \) is a private location, represented by \( \text{isprivloc} \gamma \ell \). It subsequently becomes the responsibility of the public membrane’s \text{client}, who owns \( \ell \twoheadrightarrow \nu \), to protect any high-integrity state reachable from \( \ell \) —for example, by keeping \( \ell \) hidden as we do in Appendix E. The triples \text{PubWrapSpec} and \text{PubUnwrapSpec} say that the function \( \text{pubwrap} m \) converts private values to low values, while \( \text{pubunwrap} m \) does the opposite. In particular, the output of \( \text{pubwrap} \), being low, can be freely shared with untrusted code. Also, \( \text{pubunwrap} \) has a non-progressive triple because unwrapping an arbitrary low value \( v_2 \) may apply \( \text{pubin} m \ell_2 \) with some low-integrity location \( \ell_2 \) unknown to \( m \), and our implementation of \( \text{pubin} \) gets stuck in such cases.\(^\text{14}\)

The implementation of the public membrane satisfies this specification. As an example, we briefly describe how we verify the \text{pubwrap} specification, \text{PubWrapSpec}. First, we prove the following two lemmas showing that for a valid membrane \( m \), \text{pubout} \( m \) and \( \text{pubin} m \) produce functions that satisfy

\(^{13}\)Reasoning with private values is fairly straightforward using the definition in Fig. 3 and derived rules like \text{LiftApp} (§2.2).

\(^{14}\)For simplicity, we do not specify the functionality of \text{pubwrap} and \text{pubunwrap} beyond “returns a low, resp., private value”. More useful specifications would allow one to prove that, say, after wrapping private location \( \ell_1 \) twice and obtaining values \( v_2 \) and \( v'_2 \), there exists a low location \( \ell_2 \) such that \( v_2 = \ell_2 = v'_2 \).
the predicate ismon.

\[
\{ \text{ismembrane } \gamma m \} \ \text{pubout } m \ \{ f. \ \text{ismon progress } f (\text{isprivloc } \gamma) \ \text{lowloc} \}
\]

\[
\{ \text{ismembrane } \gamma m \} \ \text{pubin } m \ \{ f. \ \text{ismon noprogress } f \ \text{lowloc} (\text{isprivloc } \gamma) \}
\]

Next, we instantiate MembraneSpec (Fig. 9) with \( \Psi_1 = \text{isprivloc } \gamma \), \( \Psi_2 = \text{lowloc} \) and the two lemmas above to obtain

\[
\text{ismon progress pubwrap (isprivval } \gamma) \ \text{lowval}
\]

The triple PubWrapSpec follows by unfolding the definition of ismon (Fig. 9).

6 RELATED WORK

This paper focuses on the formal specification and verification of OCPs. While we know of only rather preliminary prior work in this space, a major point of our paper is to observe that the tools needed for effective verification of OCPs are already to a large extent available—the pieces just need to be assembled properly. In particular, our logic, OCPL, gets significant mileage out of existing verification techniques, notably robust safety and concurrent separation logic. In the following, we discuss related work on robust safety, the (limited amount of) existing work on specification of OCPs, and some other relevant work on object capabilities and ownership types.

Robust safety. The concept of robust safety arose in the context of verifying security protocols that interact with adversaries. Early work used typing to prove “correspondence properties” for cryptographic protocols modeled in the spi calculus (Gordon and Jeffrey 2001). In their work on the refinement type-checker F7, Bengtson et al. (2011) generalized robust safety to a richer class of integrity properties for a process calculus, RCF, with higher-order state. We inherit from their work the basic idea of using a notion of low-integrity values and proving robust safety, but the approaches differ greatly in detail. First, we show how to apply this idea to OCPs, a completely different domain. Second, we show how low-integrity values are directly encodable in modern separation logics, using a simple logical relation.

Object capabilities. There has been only very preliminary work on specifying and verifying functional properties of object capabilities and OCPs. In his seminal paper on dynamic sealing, Morris (1973) proposed informal reasoning principles for programmers using dynamic sealing, but did not prove anything formally. Drossopoulou et al. (2015a) proposed predicates modeling trust and risk and used those predicates to specify a capability-based escrow exchange example (Miller et al. 2013). They focused on this one example, whereas we develop general specifications for several OCPs. Further, they focus on syntactic specifications and do not define the semantics of their predicates. A subsequent manuscript (Drossopoulou et al. 2015b) bridges this gap to the semantics, but their Hoare logic seems inadequate for the examples we consider, e.g., it lacks a rule for dynamic allocation.

Perhaps the most closely related work to our own is that of Devriese et al. (2016). As discussed in the introduction, they use a Kripke logical relation and a meta-property called effect parametricity to verify integrity properties for several examples of capability-wrapped user code in a language with higher-order state. There are several points of difference between our work and theirs. First, we work in a concurrent separation logic rather than directly in a low-level logical relation. As a result, in addition to being able to conduct our proofs at a much higher level of abstraction, we can give compositional specifications for higher-order object capability patterns (i.e., libraries), whereas they only verify specific programs that use object capabilities. We also exploit the notion of robust safety in verifying integrity properties of code that uses OCPs, whereas corresponding arguments only appear implicitly in their proofs. On the other hand, they develop semantic variants of the
Robust and Compositional Verification of Object Capability Patterns

so-called “reference graph properties” from the literature on object capabilities. These properties are important but they are also orthogonal to verification and, hence, we do not examine them. Last but not least, our proofs are machine-checked in Coq.

While specifying OCPs has received little attention so far, there has been a lot of other work on object capabilities and OCPs. Responding to renewed interest in dynamic sealing, Sumii and Pierce (2004) propose a bisimulation for proving contextual equivalences in a language with a dynamic sealing primitive. Roughly, the examples they consider are pairs of expressions—“modules” implementing the same “interface” using sealing—and the question they study is whether those expressions are indistinguishable, even if their internal representations differ. Bengtson et al. (2011) use dynamic sealing to define ideal implementations of cryptographic operations in RCF. They offer no general specifications for dynamic sealing analogous to the specification in §3, but derive instances of such a specification as needed. Van Cutsem and Miller (2013, §§4.3–4.4) describe a variant of our public membrane, which they use to implement proxy objects supporting so-called language invariants (such as the reasoning enabled, in JavaScript, by the ability to “freeze” object properties). Their focus is on the use of such membranes as an implementation technique, rather than on verification. Spiessens and Van Roy (2005); Spiessens (2007) and Murray (2010) use, respectively, model- and refinement-checking tools to establish certain safety and liveness properties of abstract models of object capability systems, including some OCPs. To take one small example, Murray shows that a specific model of an unsealing operation does not reveal a sealed value unless a capability to that sealed value was passed (possibly indirectly) to the unsealing operation by the context. Although useful, such properties cannot be directly used to verify clients of the OCPs. In contrast, our goal is very different: we write compositional specifications for concrete implementations of OCPs and our specifications can be directly used to verify clients.

Ownership types. There are interesting similarities between the core mechanisms of OCPL and prior work on ownership types (Clarke et al. 2013). In their seminal paper, Clarke et al. (1998) proposed a type system based on what came to be called owners-as-dominators. Roughly, the idea is to impose an ownership relation on objects, separating a public owner object from its hidden representation objects and ensuring that an owner mediates all outside access to its representation. Clarke et al. then proved that well-typed code preserves an invariant classifying data and enforcing isolation according to owners-as-dominators. In a similar vein, OCPL classifies locations as high-integrity (private) vs. low-integrity (public), and enforces that high-integrity locations cannot be accessed from low-integrity locations. However, OCPL is not a static type system, but rather a logic for verifying the enforcement of integrity properties in a dynamically-typed setting. Moreover, although OCPL requires that low locations contain low values, it does not stipulate any particular ownership discipline (like owners-as-dominators) for how the user should achieve this.

Other researchers have proved deeper semantic properties of owners-as-dominators. Banerjee and Naumann (2005a,b) prove relational parametricity theorems, using logical relations as a proof technique for showing that one representation may be safely replaced by another. Patrignani et al. (2011) show that an extension of owners-as-dominators can be used to prove relational secrecy properties for the join calculus, based on a distinction between high- and low-integrity values. In OCPL, we define low-integrity values using a simple (implicitly step-indexed) logical relation, and (like Patrignani et al.) our key meta-theorem concerns the interaction of user code with untyped, untrusted code. However, despite these similarities to prior work, our concrete goals are very different: rather than reasoning about relational properties, we focus on verifying functional specifications of OCPs and the robust safety of capability-wrapped code.
7 CONCLUSION AND FUTURE WORK

Object capability patterns (OCPs) enable programmers to enforce invariants on the private state of their objects, even when those objects are passed to untrusted code. In principle, this should make it easier to write secure and correct programs, but in practice, programmers may use OCPs incorrectly, resulting in subtle security flaws. In this paper, we develop OCPL, the first logic for compositionally specifying and verifying OCPs. We deploy it in reasoning about both implementations and representative clients of several well-known OCPs, in the context of a simple but expressive programming language with higher-order state.

We believe that using robust safety and separation logic to reason about OCPs scales to much richer settings. The Firefox web browser, for example, uses an automatic and significantly more sophisticated membrane pattern to enforce the so-called same-origin policy (Mozilla 2016; Barth 2011). In ongoing work, we are exploring an extension of OCPL to reason about this system. Furthermore, while the Iris proof mode enables relatively high-level Coq proofs, such manual proofs are nonetheless often tedious and routine. Additional research is needed to scale such proofs up to realistic languages and improve automation.

ACKNOWLEDGMENTS

This research was supported in part by a Microsoft Research PhD Scholarship, and in part by a European Research Council (ERC) Consolidator Grant for the project “RustBelt”, funded under the European Union’s Horizon 2020 Framework Programme (grant agreement no. 683289). We wish to thank Andrea Fischer and the OOPSLA reviewers for extremely helpful and detailed feedback.

REFERENCES


OCPL 2017. Long version of this paper (with appendices) and Coq development. (July 2017). Available at the Iris project website at http://plv.mpi-sws.org/iris/.


A SEALING WITH LOW FUNCTIONS

In this appendix, we point out that, in the dynamic sealing OCP, representing proxy values as low-integrity functions renders unsealing non-deterministic, forcing somewhat surprising changes to otherwise “obviously safe” code.

Our implementation of sealing differs from Morris’ in one important respect. Ignoring synchronization, Morris sealed a value $v$ by returning a function $(\lambda k. tbl \leftarrow \text{mapinsertnew}(!tbl) k v)$ for extending the table.\footnote{We are not so much interested in Morris’ implementation but in the fact that it returns a function. Other dynamic sealing implementations return functions and otherwise have attractive properties (e.g., dispensing with the table). Ignoring synchronization, the table-free implementation of sealing (Miller 2008; Taly et al. 2011) allows for the followingFootnote(\mathrm{\footnote{We are not so much interested in Morris’ implementation but in the fact that it returns a function. Other dynamic sealing implementations return functions and otherwise have attractive properties (e.g., dispensing with the table). Ignoring synchronization, the table-free implementation of sealing (Miller 2008; Taly et al. 2011)})

\begin{verbatim}
let r = ref None in let seal = \lambda x. r \leftarrow Some x in let unseal = \lambda f. (r \leftarrow None; f (!r)) in (seal, unseal)
\end{verbatim}

shares this drawback with Morris’ implementation.

\begin{verbatim}
UnsealAnySpec
\{isunseal \gamma u \phi\} u v' \{x. issealed \gamma x v' \phi\},
\end{verbatim}

This switch to non-deterministic unsealing of low values comes at some cost. We must weaken the triple, assuming that $\phi$ is low so that, if $u$ is a function $f$, we know by assumption that we can apply $f k$ (if we can mark $k$ low—we can). Second, the triple returns a value $v$ such that $\text{issealed}\ \gamma\ v\ v'\ \phi$ which means that subsequent applications $u\ v'$ must also return $v$. This is absurd! An adversary with values $f_1, f_2$ obtained by sealing values $v_1, v_2$ can, for example, construct

\begin{verbatim}
let x = new in let y = new in
let snap = \lambda j. seal (unseal x) in
\lambda j. unseal (\text{mapinsertnew}(!snap) j x)
\end{verbatim}

that, when unsealed, oscillates between $v_1$ and $v_2$. An easy fix is to weaken the triple further, giving up on deterministic unsealing.

We verified Morris’ implementation of sealing against a weaker specification, obtained by replacing rule $\text{UnsealAnySpec}$ in Fig. 5 with

\begin{verbatim}
UnsealLowSpec
\{isunseal \gamma u \phi \ast \text{lowval}\ v'\} u v' \{x. \phi\ x\},
\end{verbatim}

This switch to non-deterministic unsealing of low values comes at some cost. We must weaken the non-progressive triples in our intervals specification; for example, replacing $\text{MinAnySpec}$ by

\begin{verbatim}
MinLowSpec
\{isimin \gamma imin \ast \text{lowval}\ v\} imin v\ {n, ret n. \top}\}
\end{verbatim}

With these weaker rules, we can no longer verify the function $\text{check} = \lambda j. \text{assert } (imin j \leq imax j)$ in our intervals client. Applying $\text{check}$ to a monster like the oscillating function $f$ can cause the assertion to fail. All is not lost, of course. We can rewrite $\text{check}$ to read $\text{check} = \lambda j. \text{let } j = \text{snap } j \text{ in assert } (imin j \leq imax j)$ after extending our intervals library with the trivial snapshot function $\text{snap} = \lambda x. \text{seal(unseal x)}$. The point of $\text{snap}$ is to replace an unpredictable function by a predictable function.

\begin{verbatim}
let r = ref None in let seal = \lambda x. r \leftarrow Some x in let unseal = \lambda f. (r \leftarrow None; f (!r)) in (seal, unseal)
\end{verbatim}
Asymmetric signatures.

\[
\text{MakeSignSpec} \quad \phi \text{ persistent } \quad (\forall v) \, \phi \, v \vdash \text{lowval} \, v
\]

\[
\{ \top \} \quad \text{makeseal} ()
\]

\[
\{ v_1 \, v_2 \, \gamma \, , \, \text{ret} \,(v_1, \, v_2) \, . \, \text{issign} \, \gamma \, v_1 \, \phi \, * \, \text{isverify} \, \gamma \, v_2 \, \phi \} 
\]

\[
\text{SignSpec} \quad \{ \text{issign} \, \gamma \, \text{sign} \, \phi \, * \, \phi \, v \} \quad \text{sign} \, \{ x' . \, \text{issigned} \, \gamma \, v \, x' \, \phi \}
\]

\[
\text{VerifySpec} \quad \{ \text{isverify} \, \gamma \, \text{verify} \, \phi \, * \, \text{issigned} \, \gamma \, v \, v' \, \phi \} \quad \text{verify} \, v' \, \{ \text{ret} \, v \, . \, \top \}
\]

\[
\text{VerifyAnySpec} \quad \{ \text{isverify} \, \gamma \, \text{verify} \, \phi \} \quad \text{verify} \, v' \, \{ x . \, \text{issigned} \, \gamma \, x \, x' \, \phi \}, \quad \text{isverify} \, \gamma \, \text{verify} \, \phi \, * \, \text{lowval} \, \text{verify}
\]

\[
\text{SignedLow} \quad \text{issigned} \, \gamma \, v \, v' \, \phi \, * \, \text{lowval} \, v' \quad \text{SignedInv} \quad \text{issigned} \, \gamma \, v \, v' \, \phi \, * \, \phi \, v
\]

\[
\text{SignedAgree} \quad \text{issigned} \, \gamma \, v_1 \, v' \, \phi \, * \, \text{issigned} \, \gamma \, v_2 \, v' \, \phi \, + \, v_1 \, = \, v_2
\]

Asymmetric encryption.

\[
\text{MakeEncryptSpec} \quad \{ \top \} \quad \text{makeseal} ()
\]

\[
\{ v_1 \, v_2 \, \gamma \, , \, \text{ret} \,(v_1, \, v_2) \, . \, \text{isencrypt} \, \gamma \, v_1 \, * \, \text{isdecrypt} \, \gamma \, v_2 \} 
\]

\[
\text{EncryptSpec} \quad \{ \text{isencrypt} \, \gamma \, \text{enc} \} \quad \text{enc} \, v \, \{ x' . \, \text{isctext} \, \gamma \, v \, x' \}
\]

\[
\text{DecryptSpec} \quad \{ \text{isdecrypt} \, \gamma \, \text{dec} \} \quad \text{dec} \, v' \, \{ x . \, \text{isctext} \, \gamma \, x \, v' \}
\]

\[
\text{DecryptAnySpec} \quad \{ \text{isdecrypt} \, \gamma \, \text{dec} \} \quad \text{dec} \, v' \, \{ x . \, \text{isctext} \, \gamma \, x \, v' \} \quad \text{isencrypt} \, \gamma \, \text{enc} \, * \, \text{lowval} \, \text{enc}
\]

\[
\text{CtextLow} \quad \text{isctext} \, \gamma \, v \, v' \, * \, \text{lowval} \, v' \quad \text{CtextAgree} \quad \text{isctext} \, \gamma \, v_1 \, v' \, * \, \text{isctext} \, \gamma \, v_2 \, v' \, + \, v_1 \, = \, v_2
\]

Fig. 12. Public-key interfaces (derived from Fig. 5)

one:

\[
\text{SnapSpec} \quad \{ \text{issnap} \, \gamma \, \text{snap} \, * \, \text{isinterval} \, \gamma \, n_1 \, n_2 \, v \} \quad \text{snap} \, v \, \{ i . \, \text{isinterval} \, \gamma \, n_1 \, n_2 \, i \}
\]

\[
\text{SnapLowSpec} \quad \{ \text{issnap} \, \gamma \, \text{snap} \, * \, \text{lowval} \, v \} \quad \text{snap} \, v \, \{ i \, n_1 \, n_2 \, , \, \text{ret} \, i . \, \text{isinterval} \, \gamma \, n_1 \, n_2 \, i \}
\]

\[
\text{SnapLow} \quad \text{issnap} \, \gamma \, \text{snap} \, + \, \text{lowval} \, \text{snap}
\]

With these changes, we can prove the assertion expression in our modified intervals client succeeds and the client is robustly safe.
B PUBLIC-KEY INTERFACES FOR SEALING

In this appendix, we show that dynamic sealing can be used to implement ideal cryptographic signing/verification as well as encryption/decryption primitives, and that the specification of dynamic sealing (Fig. 5) can be specialized to derive very general specifications for these primitives. Such ideal primitives are sometimes used in the verification of cryptographic protocols (Bengtsson et al. 2011). While the observation that dynamic sealing can implement these cryptographic primitives is not new—Morris (1973) notes this even though he did not use the terms signing and encryption back then—the observation that specifications for the cryptographic primitives can be derived from those for dynamic sealing is new.

Signature/verification. A signature/verification primitive provides a function sign that creates a low proxy for its argument. This proxy is usually called a signature. The proxy (signature) can be verified using a dual function verify, which when applied to the value returned by sign v, returns v and if the argument was not the output of sign, gets stuck. The important point is how sign and verify are used in practice. In typical use, sign is held private by a piece of code that has been verified to apply sign only to values that satisfy a representation invariant ϕ. The verify function is publicly available to everyone and anyone can use it to verify signatures created by the code that holds sign. In this mode of use, signing/verification transfers knowledge of the representation invariant from the signer to the verifier: since the signer only signs values with representation invariant ϕ, and verify only returns previously signed values, the verifier always knows that any value returned by verify must satisfy ϕ.

It should be clear that sign and verify are quite similar to seal and unseal returned by makeseal (). The top half of Fig. 12 shows that, in fact, seal and unseal are an implementation of sign and verify. The specification shown in this figure is a specific instance of the general sealing specification from Figure 5, derived by defining the abstract predicates issign, isverify, and issigned to be isseal, isunseal, and issealed, respectively; using UnsealLow to prove VerifyLow; and dropping SealLow. The derived specification says that makeseal () returns a pair of functions that behave like sign and verify if the representation invariant ϕ implies lowval. This condition is necessary since verify returns previously signed values to untrusted code, so signed values should always be low—in cryptographic terms, signatures provide no secrecy. Additionally, the specification says that the value returned by verify always satisfies ϕ (rule VerifyAnySpec), that verify itself is low (rule VerifyLow) so verify can be shared safely with untrusted code, and that the output of sign is low (rule SignedLow) so signatures can also be shared safely with untrusted code. On the other hand, sign itself is not low, so it should not be shared with untrusted code.

Encryption/decryption. An encryption/decryption primitive provides a function enc that creates a low proxy for its argument. A dual function dec converts the proxy back to the original argument. Typically, encryption/decryption is used by holding dec private in some piece of code, and making enc public. This allows anyone to encrypt a secret using enc and share it freely, with the guarantee that only the code holding dec can ever access the secret.

Again, enc and dec look similar to seal and unseal returned by makeseal () and the bottom half of Fig. 12 formalizes this intuition by specializing the specification of Fig. 5 to match the intuitive description of encryption/decryption. The abstract predicates isencrypt, isdecrypt, and isctext are defined to be isseal, isunseal, and issealed, respectively. Importantly, the specification says that makeseal () returns functions that behave like enc and dec (rule MakeEncryptSpec), that decrypting a previously encrypted value v returns v (rules EncryptSpec and DecryptSpec), and that the encryption function and ciphertexts are low (rules EncryptLow and CtextLow) so they can be shared with...
The following rules constitute the full specification of the intervals library in §3.

**Progressive triples.**

\[
\text{INTERVALS SPEC} \\
\{ \top \} \\
\text{intervals (}
\{ v_1, v_2, v_3, v_4, \text{ ret } (v_1, v_2, v_3, v_4). \text{ ismakeint } v_1 * \text{ ismin } v_2 * \text{ ismax } v_3 * \text{ issum } v_4 \}
\]

\[
\text{MAKEINT' SPEC} \\
\{ \text{ismakeint' } y \ n_1 \ f \} \\
\]

\[
\text{MAKEINT SPEC} \\
\{ \text{ismakeint } y \ m_k \} \ mk \ n_1 \ \{ f. \ \text{ismakeint' } y \ n_1 \ f \} \\
\]

\[
\text{MIN SPEC} \\
\{ \text{ismin } y \ \text{imin } * \text{ isinterval } y \ n_1 \ n_2 \ i \} \ \text{imin } i \ \{ \text{ret } n_1. \ \top \}
\]

\[
\text{MAX SPEC} \\
\{ \text{ismax } y \ \text{imax } * \text{ isinterval } y \ n_1 \ n_2 \ i \} \ \text{imax } i \ \{ \text{ret } n_2. \ \top \}
\]

\[
\text{SUM SPEC} \\
\{ \text{isum } y \ \text{isum } * \text{ isinterval } y \ n_1 \ n_2 \ i \} \ \text{isum } i \ \{ f. \ \text{isum' } y \ n_1 \ n_2 \ f \}
\]

\[
\text{SUM' SPEC} \\
\{ \text{isum' } y \ n_1 \ n_2 \ f * \text{ isinterval } y \ n_1' n_2' i' \} \\
\]

\[
\text{INTERVAL INV} \\
\{ j. \ \text{isinterval } y \ (n_1 + n_1') \ (n_2 + n_2') \} \ \text{isinterval } y \ n_1 \ n_2 \ i + n_1 \leq n_2
\]

\[
\text{INTERVAL AGREE} \\
\text{isinterval } y \ n_1 \ n_2 \ i * \text{ isinterval } y \ n_1' n_2' i' \\
\implies n_1 = n_1' * n_2 = n_2'
\]

**Non-progressive triples.**

\[
\text{MIN ANY SPEC} \\
\{ \text{ismin } y \ \text{imin } v \ \{ n_1 \ n_2, \text{ ret } n_1. \ \text{isinterval } y \ n_1 \ n_2 \ v \}\}
\]

\[
\text{MAX ANY SPEC} \\
\{ \text{ismax } y \ \text{imax } v \ \{ n_1 \ n_2, \text{ ret } n_2. \ \text{isinterval } y \ n_1 \ n_2 \ v \}\}
\]

\[
\text{SUM ANY SPEC} \\
\{ \text{isum } y \ \text{isum } \} \ \text{isum } v_1 \ \{ f \ n_1 \ n_2, \text{ ret } f. \ \text{isinterval } y \ n_1 \ n_2 \ v_1 * \text{ issum' } y \ n_1 \ n_2 \ f \}\}
\]

\[
\text{SUM' ANY SPEC} \\
\{ \text{isum' } y \ n_1 \ n_2 \ f \} \ f \ v_2 \ \{ i \ n_1' n_2', \text{ret } i. \ \text{isinterval } y \ n_1' n_2' v_2 * \text{ isinterval } y \ (n_1 + n_1') \ (n_2 + n_2') \ i \}
\]

**Low values.**

\[
\text{INTERVAL LOW} \\
\text{isinterval } y \ n_1 \ n_2 \ i + \text{ lowval } i \\
\text{MAKEINT' LOW} \\
\text{ismakeint' } y \ n_1 \ f + \text{ lowval } f \\
\text{MAKEINT LOW} \\
\text{ismakeint } y \ m_k + \text{ lowval } m_k
\]
D BLOCKING CARETAKERS

In this appendix, we briefly present a simpler implementation of API caretakers.\(^{16}\)

\[
\text{makecaretaker} \triangleq \text{makelock}' \\
\text{wrap} \triangleq \lambda ct x. \text{syncwith ct} (\lambda_. f x) \\
\text{disable} \triangleq \text{acquire}
\]

The idea is that, if we are happy to let wrappers block until a caretaker is enabled, we can dispense with the enabled flag used in §4. In this implementation, a caretaker is a lock. The function makelock' allocates a fresh lock, initially locked. The function syncwith is analogous to sync, except that it takes the lock to use as an initial argument. To disable and enable wrappers, we simply acquire and release the lock they need. This implementation also satisfies the API caretaker specification (Fig. 7).

E PUBLIC MEMBRANE CLIENT

In this appendix, we further illustrate the public membrane by describing a simple client program that uses it. First, we define the following auxiliary function, getint. This function presupposes that we have some invariant on a private integer location, \(r\), and that we want to periodically update \(r\) by “merging” its contents with those of its shadow location, using some merge function \(f\) that preserves the invariant.

\[
\text{getint} \triangleq \lambda m f r. \text{let} n_1 = !r \text{ in let} x_2 = \text{shadowread} m r \text{ in} \\
\quad \text{if isint} x_2 \text{ then} \\
\quad \quad \text{let} n_3 = f n_1 x_2 \text{ in} \\
\quad \quad \text{let} _ = \text{if} n_1 \neq n_3 \text{ then} r \leftarrow n_3 \text{ else} () \text{ in} \\
\quad \quad \text{let} _ = \text{if} x_2 \neq n_3 \text{ then} \text{shadowwrite} m r n_3 \text{ else} () \text{ in} \\
\quad n_3 \\
\quad \text{else} (\text{shadowwrite} m r n_1; n_1)
\]

If the private location \(r\) and its shadow contain integers \(n_1\) and \(n_2\), the code applies \(f n_1 n_2\) to obtain the “merged” integer \(n_3\), updates \(r\) and its shadow, and returns \(n_3\); otherwise, the code overwrites the shadow’s contents with (and returns) \(n_1\).

Our illustrative client program is a private up/down counter with increment and decrement functions, and private lower and upper limits. The invariant we have in mind is that the counter always remains within these limits. However, the counter also exposes shadow locations for the lower and upper limits to its untrusted context. The context can freely update these shadow lower and upper limits. However, these changes are propagated to the private locations (using getint).

\(^{16}\)We discuss the straightforward functions makelock’, release, acquire, and syncwith in Appendix F.
only if the changes maintain the intended invariant.

\[
\text{client} \triangleq \text{let } m = \text{makepub()} \text{ in } \\
\text{let } lo = \text{pubref } m \text{ in let } c = \text{ref } 0 \text{ in let } hi = \text{pubref } m \text{ in } \\
\text{let sync} = \text{makesync()} \text{ in } \\
\text{let use} = \lambda _. \text{sync(\_.) assert } !lo \leq !c; \text{assert } !c \leq !hi \\
\text{let getlimits} = \lambda _. \text{let } n = !c \text{ in } \\
\text{let } a = \text{getint}(\lambda n_1 n_2. \text{if } n_2 \leq n \text{ then } n_2 \text{ else } n_1) \text{ lo in } \\
\text{let } b = \text{getint}(\lambda n_1 n_2. \text{if } n \leq n_2 \text{ then } n_2 \text{ else } n_1) \text{ hi in } \\
(a, b) \\
\text{let } \text{decr} = \lambda _. \text{sync(\_.) let } n = (!c) - 1 \text{ in } \\
\text{let } b = \text{fst(getlimits())} \leq n \text{ in } \\
\text{if } b \text{ then } c \leftarrow n \text{ else } (); b \\
\text{let } \text{incr} = \lambda _. \text{sync(\_.) let } n = (!c) + 1 \text{ in } \\
\text{let } b = n \leq \text{snd(getlimits())} \text{ in } \\
\text{if } b \text{ then } c \leftarrow n \text{ else } (); b \\
\text{let } \text{lo}' = \text{pubwrap } m \text{ lo in let } \text{hi}' = \text{pubwrap } m \text{ hi in } \\
(\text{use, lo}, \text{hi}, \text{incr}, \text{decr})
\]

The client’s private state comprises a public membrane \(m\), a private counter \(c\) and private limits \(lo\) and \(hi\), subject to the representation invariant \(!lo \leq !c \leq !hi\) (i.e., the counter respects its limits). The client returns (i) a simple function \(use\) that, when applied, simply asserts the representation invariant (and that could be extended to offer some service to the counter’s clients), (ii) the shadow locations \(lo'\) and \(hi'\) that substitute \(lo\) and \(hi\) for the surrounding untrusted code (the untrusted code may freely change these at any time) and (iii) increment and decrement functions, \(\text{incr}\) and \(\text{decr}\), that either change the counter \(c\) and return \(\text{true}\) or (when changing the counter would violate its current limits) return \(\text{false}\). Prior to changing the counter, \(\text{incr}\) and \(\text{decr}\) copy the shadow locations \(lo'\) and \(hi'\) into \(lo\) and \(hi\) using the function \(\text{getint}\), if doing so will not break the representation invariant.

Using the rules of OCPL, we prove that client is robustly safe, meaning that no context can exploit the returned values \(use\), \(\text{incr}\), \(\text{decr}\), \(lo'\) and \(hi'\) in a way that violates the assertions about the representation invariant in \(use\). To do this, we first prove that client satisfies the following triple: \(\top\) \{client \(\{x. \text{lowval }x\}\)\}. One key step in this proof is showing that \(lo'\) and \(hi'\) are both low. This follows because \(lo'\) and \(hi'\) are obtained from \(\text{pubwrap}\) and the postcondition of \(\text{pubwrap}\) (rule \text{PubWrapSpec}, Fig. 11) says that the value it returns is low. Robust safety follows immediately from Theorem RobustSafety applied to this triple.

\section{Library Functions}

In this appendix, we specify the implementations of locks, finite maps, and finite bijections used in our examples. As nothing here is original or terribly interesting, we relegate many implementation and verification details to our Coq development.

\textit{Locks and synchronization.} We specify locks and synchronization in Fig. 13. Rules \texttt{LockedExclusive-AcquireSpec} constitute a CAP-style lock specification (Dinsdale-Young et al. 2010) satisfied by, for
Initially locked locks.

LockedExclusive
locked γ * locked γ ⊩ ⊥

MakeLockSpec
{⊤} makelock' () {lk γ, ret lk. islock lk R * locked γ}

ReleaseSpec
{islock γ lk R * locked γ * R} release lk {ret (). ⊤}

AcquireSpec
{islock γ lk R} acquire lk {ret (). locked γ * R}

Initially unlocked locks.

MakeLockSpec
{R} makelock () {lk γ, ret lk. islock lk R}

Synchronization.

IsSync

issync sync R ≜ □ ∀Q. {R} e {v. R * Q} ⊮

SyncWithSpec

{⊤} sync (λ_· e) {v. Q}

{islock γ lk R} syncwith lk {v. issync v R}

MakeSyncSpec

{R} makesync () {v. issync v R}

Fig. 13. Lock and synchronization interface

example, the evident implementation of spin locks using CAS. The assertion islock γ lk R represents knowledge that the value lk is a lock with abstract name γ protecting resource R. The assertion locked γ represents ownership of the lock named γ. Thus, for any resource R,

- makelock' () returns a new, locked lock lk protecting R (MakeLock' Spec);
- release lk gives up lock lk and resource R, returning unit (ReleaseSpec); and
- acquire lk grabs lock lk and the resource R it protects, returning unit (AcquireSpec).

This specification is atypical in that makelock’ constructs a locked lock, a feature needed only to argue that our caretaker specification is just a specialized lock spec (see the implementation of “blocking caretakers” in Appendix D). MakeLockSpec embodies a typical CAP-style lock specification: it specifies the evident wrapper around makelock’ and release that consumes resource R in order to construct a new, unlocked lock lk protecting resource R.

Rules IsSync—MakeSyncSpec specify the following simple library, readily verified against any lock implementation.

syncwith ≜ λlk f. acquire lk; let r = f () in release lk; r
makesync ≜ λ_. let lk = makelock () in syncwith lk

Given a lock lk and a thunk f, the function syncwith applies f while holding the lock, returning whatever f() returns. The function makesync returns syncwith lk for a fresh lock lk. In specifying makesync, we follow Turon et al. (2013). Definition IsSync is the key. It says that to prove sync (λ_· e) returns a value satisfying Q, it suffices to show that e, when given access to the resource R protected by sync’s lock, (i) returns a value satisfying Q and (ii) preserves R.
Finite maps. We specify functional finite maps in Fig. 14. The spec is parameterized by an injection \( \iota : X \rightarrow Val \) from some type \( X \) (our examples use locations) to values. The assertion
\[
\text{ismap } \text{map} \ f
\]
denotes knowledge that value \( \text{map} \) represents the finite partial function \( f : X \xrightarrow{\text{fin}} Val \). \text{MapEmptySpec} says that value \( \text{mapempty} \) represents the empty map. \text{MapInsertNewSpec} specifies a special case of insertion. \text{MapLookupPartialSpec} specifies a partial lookup operation. The evident implementation of finite maps using association lists satisfies this spec when \( X \) enjoys decidable equality.

Finite bijections. We specify functional finite partial bijections in Fig. 15. As with finite maps, the spec is parameterized by an injection \( \iota : X \rightarrow Val \) from some type \( X \) to values. The assertion
\[
\text{isbij } \text{bij} \ f \ f' \ 
\]
denotes knowledge that value \( \text{bij} \) represents the finite partial bijection \( f, f' : X \xrightarrow{\text{fin}} Val \) in the sense that for every \( x \in X \) and value \( v' \), if \( f \ x = v' \), then there exists \( x' \in X \) such that \( v' = \iota \ x' \) and \( f' \ x' = \iota \ x \). Representing a bijection as a pair of finite maps (as specified in Fig. 14) leads to a natural implementation of this specification.