

Verifiable C

*Applying the Verified Software Toolchain
to C programs*

*Version 2.0
January 15, 2018*

Andrew W. Appel

with Lennart Beringer, Qinxiang Cao, Josiah Dodds

Contents

Verifiable C	i
Contents	ii
1 Overview	5
2 Installation	7
3 Workflow, loadpaths	8
4 Verifiable C and clightgen	9
5 ASTs: abstract syntax trees	10
6 Use the IDE	12
7 Functional model, API spec	13
8 Proof of the sumarray program	17
9 start_function	18
10 forward	21
11 If, While, For	24
12 While loops	25
13 PROP() LOCAL() SEP()	27
14 Entailments	28
15 Array subscripts	30
16 At the end of the loop body	32
17 Returning from a function	34
18 Global variables and main()	35
19 Function calls	37
20 Tying all the functions together	38
21 Separation logic: EX, *, emp, !!	39
22 EX, Intros, Exists	40
23 Integers: nat, Z, int	42
24 Values: Vint, Vptr	44
25 C types	45
26 CompSpecs	46
27 retype	47
28 Uninitialized data, default_val	48
29 data_at	49

30	field_at	50
31	reptype', repinj	51
32	LOCAL defs: temp, lvar, gvar	53
33	go_lower	54
34	saturate_local	55
35	field_compatible, field_address	56
36	value_fits	58
37	cancel	60
38	entailer!	61
39	normalize	62
40	assert_PROP	66
41	Welltypedness of variables	67
42	Shares	68
43	Pointer comparisons	70
44	Proof of the reverse program	71
45	Alternate proof of reverse	75
46	Global variables	76
47	For loops (special case)	77
48	For loops (general case)	78
49	Manipulating preconditions	79
50	The Frame rule	81
51	32-bit Integers	82
52	CompCert C abstract syntax	85
53	C light semantics	87
54	Splitting arrays	89
55	sublist	90
56	Later	92
57	Mapsto and func_ptr	93
58	with_library: Library functions	94
59	malloc/free	95
60	exit	96
61	Function pointers	97

62	Axioms of separation logic	99
63	Obscure higher-order axioms	100
64	Proving larg(ish) programs	101
65	Separate compilation, <code>semax_ext</code>	103
66	Catalog of tactics/lemmas	104

1 Overview

Verifiable C is a language and program logic for reasoning about the functional correctness of C programs. The *language* is a subset of CompCert C light; it is a dialect of C in which side-effects and loads have been factored out of expressions. The *program logic* is a higher-order separation logic, a kind of Hoare logic with better support for reasoning about pointer data structures, function pointers, and data abstraction.

Verifiable C is *foundationally sound*. That is, it is proved (with a machine-checked proof in the Coq proof assistant) that,

Whatever observable property about a C program you prove using the Verifiable C program logic, that property will actually hold on the assembly-language program that comes out of the C compiler.

This soundness proof comes in two parts: The program logic is proved sound with respect to the semantics of CompCert C, by a team of researchers primarily at Princeton University; and the C compiler is proved correct with respect to those same semantics, by a team of researchers primarily at INRIA. This chain of proofs from top to bottom, connected in Coq at specification interfaces, is part of the *Verified Software Toolchain*.



To use Verifiable C, one must have had some experience using Coq, and some familiarity with the basic principles of Hoare logic. These can be obtained by studying Pierce’s *Software Foundations* interactive textbook, and doing the exercises all the way to chapter “Hoare2.”

It is also useful to read the brief introductions to Hoare Logic and Separation Logic, covered in Appel’s *Program Logics for Certified Compilers*, [Chapters 2 and 3](#) (those chapters available free, follow the link).

PROGRAM LOGICS FOR CERTIFIED COMPILERS (Cambridge University Press, 2014) describes *Verifiable C* version 1.1. If you are interested in the semantic model, soundness proof, or memory model of VST, the book is well worth reading. But it is not a reference manual.

More recent VST versions differ in several ways from what the PLCC book describes.

- In the LOCAL component of an assertion, one writes `temp i v` instead of ``(eq v) (eval_id i)`.
- In the SEP component of an assertion, backticks are not used (predicates are not lifted).
- In general, the backtick notation is rarely needed.
- The type-checker now has a more refined view of char and short types.
- `field_mapsto` is now called `field_at`, and it is dependently typed.
- `typed_mapsto` is renamed `data_at`, and last two arguments are swapped.
- `umapsto` (“untyped mapsto”) no longer exists.
- `mapsto sh t v w` now permits either ($w = \text{Vundef}$) or the value w belongs to type t . This permits describing uninitialized locations, i.e., `mapsto_sh t v = mapsto_sh t v Vundef`. For function calls, one uses `forward.call` instead of `forward`.
- C functions may fall through the end of the function body, and this is (per the C semantics) equivalent to a return statement.

2 Installation

The Verified Software Toolchain runs on Linux, Mac, or Windows. You will need to install:

Coq 8.7.1, from coq.inria.fr. Follow the standard installation instructions.

CompCert 3.2, from <http://compcert.inria.fr/download.html>. Build the *clightgen* tool, using these commands: `./configure -clightgen x86_32-linux; make`. You might replace `x86_32-linux` with `x86_32-macosx` or `x86_32-cygwin`. Verifiable C should work on other 32-bit architectures as well, but has not been extensively tested. Verifiable C has not been tested on 64-bit architectures, but a near-future release will support these.

VST 2.0, from vst.cs.princeton.edu, or else an appropriate version from <https://github.com/PrincetonUniversity/VST>. After unpacking, read the `BUILD_ORGANIZATION` file (or simply `make -j`).

Note on the Windows (cygwin) installation of CompCert: To build CompCert you'll need an up to date version of the [menhir parser generator](#). To work around a cygwin incompatibility in the menhir build, touch `src/.versioncheck` before doing `make`.

3 Workflow, loadpaths

Within VST, the `progs` directory contains some sample C programs with their verifications. The workflow is:

- Write a C program $F.c$.
- Run `clightgen -normalize $F.c$` to translate it into a Coq file $F.v$.
- Write a verification of $F.v$ in a file such as `verif_` $F.v$. That latter file must import both $F.v$ and the VST *Floyd*¹ program verification system, `VST.floyd.proofauto`.

LOAD PATHS. Interactive development environments (CoqIDE or Proof General) will need their load paths properly initialized through command-line arguments. Running `make` in `vst` creates a file `.loadpath` with the right arguments *for proof development of the VST itself or of its progs/examples*. For example,

```
coqide `cat .loadpath` progs/verif_reverse.v &
```

IN NORMAL USE (if you are not simply browsing the `progs` examples) your own files ($F.c$, $F.v$, `verif_` $F.v$) will not be inside the VST directory. You will need to run `coqc` or `coqide` (or Proof General) with “coq flags” to access the VST components. For this, use the file `.loadpath-export` or `._CoqProject-export`, created by `make` in VST.

Example:

```
cd my-own-directory
coqide `cat my/path/to/VST/.loadpath-export` myfile.v
```

FOR MORE INFORMATION, See the heading **USING PROOF GENERAL AND COQIDE** in the file `BUILD_ORGANIZATION`.

¹Named after Robert W. Floyd (1936–2001), a pioneer in program verification.

4 Verifiable C and *clightgen*

Verifiable C is a *program logic* (higher-order impredicative concurrent separation logic) for C programs with these restrictions:

- No casting between integers and pointers.
- No goto statements.
- Only structured switch statements (no Duff’s device).

CompCert’s *clightgen* tool translates C into abstract syntax trees (ASTs) of CompCert’s *Clight* intermediate language. You find *clightgen* in the root directory of your CompCert installation, after doing `make clightgen`.

Suppose you have a C source program broken into three files `x.c` `y.c` `z.c`.

```
clightgen -normalize x.c y.c z.c
```

This produces the files `x.v` `y.v` `z.v` containing Coq representations of ASTs.

Clightgen invokes the standard macro-preprocessor (to handle `define` and `include`), parses, type-checks, and produces ASTs. We translate all three files in one call to *Clightgen*, so that the global names in the C program (“extern” identifiers) will have consistent symbol-table indexes (`ident` values) across all three files.

Although your C programs may have side effects inside subexpressions, and memory dereferences inside subexpressions or if-tests, the *program logic* does not permit this. Therefore, *clightgen* transforms your programs before you apply the program logic:

- Factors out function calls and assignments from inside subexpressions (by moving them into their own assignment statements).
- Factors `&&` and `||` operators into `if` statements (to capture short circuiting behavior).
- When the `-normalize` flag is used, factors each memory dereference into a top level expression, i.e. `x=a[b[i]]`; becomes `t=b[i]; x=a[t];`.

5 *ASTs: abstract syntax trees*

We will introduce Verifiable C by explaining the proof of a simple C program: adding up the elements of an array.

```

unsigned sumarray(unsigned a[], int n) {
  int i; unsigned s;
  i=0;
  s=0;
  while (i<n) {
    s+=a[i];
    i++;
  }
  return s;
}

unsigned four[4] = {1,2,3,4};

int main(void) {
  unsigned s;
  s = sumarray(four,4);
  return (int)s;
}

```

You can examine this program in `VST/progs/sumarray.c`. Then look at `progs/sumarray.v` to find the output of CompCert's *clightgen* utility: it is the abstract syntax tree (AST) of the C program, expressed in Coq. In `sumarray.v` there are definitions such as,

Definition `_main` : ident := 54%positive.

Definition `_s` : ident := 50%positive.

...

Definition `f_sumarray` := { |

fn_return := tint; ...

fn_params := ((_a, (tptr tint)) :: (_n, tint) :: nil);

fn_temps := ((_i, tint) :: (_s, tint) :: (_x, tint) :: nil);

fn_body :=

(Ssequence

(Sset _i (Econst_int (Int.repr 0) tint))

(Ssequence (Sset _s (Econst_int (Int.repr 0) tint)) (Ssequence ...))) |}.

Definition `prog` : Clight.program := { | ... f_sumarray ... |}.

In general it's never necessary to read the AST file such as `sumarray.v`. But it's useful to know what kind of thing is in there. C-language identifiers such as `main` and `s` are represented in ASTs as positive numbers (for efficiency); the definitions `_main` and `_s` are abbreviations for these. The AST for `sumarray` is in the function-definition `f_sumarray`.

In the source program `sumarray.c`, the function `sumarray`'s return type is `int`. In the abstract syntax (`sumarray.v`), the `fn.return` component of the function definition is `tint`, or equivalently (by **Definition**) `Tint l32 Signed noattr`. The `Tint` constructor is part of the abstract syntax of C type-expressions, defined by CompCert as,

Inductive `type : Type :=`

- | `Tvoid`: `type`
- | `Tint`: `intsize` \rightarrow `signedness` \rightarrow `attr` \rightarrow `type`
- | `Tpointer`: `type` \rightarrow `attr` \rightarrow `type`
- | `Tstruct`: `ident` \rightarrow `attr` \rightarrow `type`
- | `...` .

See also [Chapter 25](#) (C types).

6 *Use the IDE*

[Chapter 7](#) through [Chapter 20](#) are meant to be read while you have the file `progs/verif_sumarray.v` open in a window of your interactive development environment for Coq. You can use Proof General, CoqIDE, or any other IDE that supports Coq.

Reading these chapters will be much less informative if you cannot see the proof state as each chapter discusses it.

Before starting the IDE, review [Chapter 3](#) (Workflow) to see how to set up load paths.

7 *Functional model, API spec*

A program without a specification cannot be incorrect, it can only be surprising.
(Paraphrase of J. J. Horning, 1982)

The file `progs/verif_sumarray.v` contains the specification of `sumarray.c` and the proof of correctness of the C program with respect to that specification. For larger programs, one would typically break this down into three or more files:

1. Functional model (often in the form of a Coq function)
2. API specification
3. Function-body correctness proofs, one per file.

We start `verif_sumarray.v` with some standard boilerplate:

```
Require Import VST.floyd.proofauto.
Require Import VST.progs.sumarray.
Instance CompSpecs : compspecs. make_compspecs prog. Defined.
Definition Vprog : varspecs. mk_varspecs prog. Defined.
```

The first line imports Verifiable C and its *Floyd* proof-automation library. The second line imports the AST of the program to be proved. Lines 3 and 4 are identical in any verification: see [Chapter 26](#) and [Chapter 46](#).

To prove correctness of `sumarray.c`, we start by writing a *functional spec* of adding-up-a-sequence, then an *API spec* of adding-up-an-array-in-C.

FUNCTIONAL MODEL. A *mathematical model* of this program is the sum of a sequence of integers: $\sum_{i=0}^{n-1} x_i$. It's conventional in Coq to use list to represent a sequence; we can represent the sum with a list-fold:

Definition `sum_Z` : list Z → Z := fold_right Z.add 0.

A functional model contains not only definitions; it's also useful to include theorems about this mathematical domain:

Lemma `sum_Z_app`: $\forall a\ b, \text{sum_Z } (a++b) = \text{sum_Z } a + \text{sum_Z } b.$

Proof. `intros. induction a; simpl; omega. Qed.`

The data types used in a functional model can be any kind of mathematics at all, as long as we have a way to relate them to the integers, tuples, and sequences used in a C program. But the mathematical integers `Z` and the 32-bit modular integers `Int.int` are often relevant. Notice that this functional spec does not depend on `sumarray.v` or even on anything in the Verifiable C libraries. This is typical, and desirable: the functional model is about mathematics, not about C programming.

THE APPLICATION PROGRAMMER INTERFACE (API) of a C program is expressed in its header file: function prototypes and data-structure definitions that explain how to call upon the modules' functionality. In *Verifiable C*, an *API specification* is written as a series of *function specifications* (funspecs) corresponding to the function prototypes.

Definition `sumarray_spec` : `ident * funspec` :=
`DECLARE _sumarray`
`WITH a: val, sh : share, contents : list Z, size: Z`
`PRE [_a OF (tptr tuint), _n OF tint]`
`PROP(readable_share sh;`
`$0 \leq \text{size} \leq \text{Int.max_signed};$`
`Forall (fun x $\Rightarrow 0 \leq x \leq \text{Int.max_unsigned}$) contents)`
`LOCAL(temp _a a; temp _n (Vint (Int.repr size)))`
`SEP(data_at sh (tarray tuint size) (map Vint (map Int.repr contents)) a)`
`POST [tuint]`
`PROP()`
`LOCAL(temp ret_temp (Vint (Int.repr (sum_Z contents))))`
`SEP(data_at sh (tarray tuint size) (map Vint (map Int.repr contents)) a).`

The funspec begins, **Definition** `f_spec` := `DECLARE _f ...` where `f` is the name of the C function, and `_f` : `ident` is Coq's name for the identifier that denotes `f` in the AST of the C program (see [page 10](#)).

A function is specified by its *precondition* and its *postcondition*. The WITH clause quantifies over Coq values that may appear in both the precondition and the postcondition. The precondition is parameterized by the C-language function parameters, and the postcondition is parameterized by a identifier `ret_temp`, which is short for, “the temporary variable holding the return value.”

Function preconditions, postconditions, and loop invariants are *assertions* about the state of variables and memory at a particular program point. In an assertion $\text{PROP}(\vec{P}) \text{ LOCAL}(\vec{Q}) \text{ SEP}(\vec{R})$, the propositions in the sequence \vec{P} are all of Coq type `Prop`. They describe things that are true independent of program state. In the function precondition above, the statement $0 \leq \text{size} \leq \text{Int.max_signed}$ is true *just within the scope of the quantification of the variable size*; it is bound by WITH, and spans the PRE and POST assertions.

The LOCAL propositions \vec{Q} are *variable bindings* of type `localdef`. Here, the function-parameters a and n are treated as nonaddressable local variables, or “temp” variables. The `localdef (temp _a a)` says that (in this program state) the contents of C local variable `_a` is the Coq value a . In general, a C scalar variable holds something of type `val`; this type is defined by CompCert as,

Inductive `val`: `Type` := `Vundef`: `val` | `Vint`: `int` → `val` | `Vlong`: `int64` → `val`
 | `Vfloat`: `float` → `val` | `Vsingle`: `float32` → `val` | `Vptr`: `block` → `ptrofs` → `val`.

The SEP conjuncts \vec{R} are *spatial assertions* in separation logic. In this case, there’s just one, a `data_at` assertion saying that at address a in memory, there is a data structure of type *array[size] of unsigned integers*, with access-permission `sh`, and the contents of that array is the sequence `map Vint (map Int.repr contents)`.

THE POSTCONDITION is introduced by `POST [tuint]`, indicating that this function returns a value of type `unsigned int`. There are no PROP statements in this postcondition—no forever-true facts hold now, that

weren't already true on entry to the function. The LOCAL *must not mention* the function parameters or local variables, because they are destroyed on function exit (and because your local variable names will not be meaningful to the function's caller); it will only mention the return-temporary `ret_temp`. The SEP clause mentions all the spatial resources from the precondition, minus ones that have been freed (deallocated), plus ones that have been malloc'd (allocated).

So, overall, the specification for `sumarray` is this: “At any call to `sumarray`, there exist values $a, sh, contents, size$ such that sh gives at least read-permission; $size$ is representable as a nonnegative 32-bit signed integer; function-parameter `_a` contains value a and `_n` contains the 32-bit representation of $size$; and there's an array in memory at address a with permission sh containing $contents$. The function returns a value equal to `sum_int(contents)`, and leaves the array unaltered.”

INTEGER OVERFLOW. In Verifiable C's signed integer arithmetic, you must prove (if the system cannot prove automatically) that no overflow occurs. In unsigned integers, arithmetic is treated as modulo- 2^n (where n is typically 32 or 64), and overflow is not an issue. See [Chapter 23](#). The function `Int.repr`: $\mathbb{Z} \rightarrow \text{int}$ truncates mathematical integers into 32-bit integers by taking the (sign-extended) low-order 32 bits. `Int.signed`: $\text{int} \rightarrow \mathbb{Z}$ injects back into the signed integers.

This program uses unsigned arithmetic for the s and the array contents, and uses signed arithmetic for i .

The postcondition guarantees that the value returned is `Int.repr (sum_Z contents)`. But what if $\sum s \geq 2^{32}$, so the sum doesn't fit in a 32-bit signed integer? Then `Int.unsigned(Int.repr (sum_Z contents)) \neq (sum_Z contents)`. In general, for a claim about `Int.repr(x)` to be *useful*, one also needs a claim that $0 \leq x \leq \text{Int.max_unsigned}$ or $\text{Int.min_signed} \leq x \leq \text{Int.max_signed}$. The caller of this function will probably need to prove $0 \leq \text{sum_Z contents} \leq \text{Int.max_unsigned}$ in order to make much use of the postcondition.

8 *Proof of the sumarray program*

To prove correctness of a whole program,

1. Collect the function-API specs together into Gprog: list funspec.
2. Prove that each function satisfies its own API spec (with a `semax_body` proof).
3. Tie everything together with a `semax_func` proof.

In `progs/verif_sumarray.v`, the first step is easy:

Definition `Gprog := ltac:(with_library prog [sumarray_spec; main_spec])`.

The function specs, built using `DECLARE`, are listed in the argument to `with_library`. [Chapter 58](#) describes `with_library`.

In addition to `Gprog`, the API spec contains `Vprog`, the list of global-variable type-specs. This is computed automatically by the `mk_varspecs` tactic, as shown at the beginning of `verif_sumarray.v`.

Each C function can call any of the other C functions in the API, so each `semax_body` proof is a client of the entire API spec, that is, `Vprog` and `Gprog`. You can see that in the statement of the `semax_body` lemma for the `_sumarray` function:

Lemma `body_sumarray: semax_body Vprog Gprog f_sumarray sumarray_spec`.

Here, `f_sumarray` is the actual function body (AST of the C code) as parsed by `clightgen`; you can read it in `sumarray.v`. You can read `body_sumarray` as saying, *In the context of `Vprog` and `Gprog`, the function body `f_sumarray` satisfies its specification `sumarray_spec`*. We need the context in case the `sumarray` function refers to a global variable (`Vprog` provides the variable's type) or calls a global function (`Gprog` provides the function's API spec).

9 start_function

The predicate `semax_body` states the Hoare triple of the function body, $\Delta \vdash \{Pre\}c\{Post\}$. *Pre* and *Post* are taken from the funspec for *f*, *c* is the body of *F*, and the type-context Δ is calculated from the global type-context overlaid with the parameter- and local-types of the function.

To prove this, we begin with the tactic `start_function`, which takes care of some simple bookkeeping and expresses the Hoare triple to be proved.

Lemma `body_sumarray`: `semax_body Vprog Gprog f_sumarray sumarray_spec`.

Proof.

`start_function`.

The proof goal now looks like this:

```

Espec : OracleKind
a : val
sh : share
contents : list Z
size : Z
Delta_specs := abbreviate : PTree.t funspec
Delta := abbreviate : tycontext
SH : readable_share sh
H : 0 ≤ size ≤ Int.max_signed
H0 : Forall (fun x : Z ⇒ 0 ≤ x ≤ Int.max_unsigned) contents
POSTCONDITION := abbreviate : ret_assert
MORE_COMMANDS := abbreviate : statement
----- (1/1)
semax Delta
  (PROP ()
    LOCAL(temp _a a; temp _n (Vint (Int.repr size)))
    SEP(data_at sh (tarray tuint size) (map Vint (map Int.repr contents)) a))
  (Ssequence (Sset _i (Econst.int (Int.repr 0) tint)) MORE_COMMANDS)
  POSTCONDITION

```

First we have *Espec*, which you can ignore for now (it characterizes the outside world, but `sumarray.c` does not do any I/O). Then `a,sh,contents,size` are exactly the variables of the `WITH` clause of `sumarray_spec`.

The two abbreviations `Delta_spec`, `Delta` are the type-context in which Floyd's proof tactics will look up information about the types of the program's variables and functions. The hypotheses `SH,H,H0` are exactly the `PROP` clause of `sumarray_spec`'s precondition. The `POSTCONDITION` is exactly the `POST` part of `sumarray_spec`.

To see the contents of an abbreviation, either (1) set your IDE to show implicit arguments, or (2) unfold abbreviate in `POSTCONDITION`.

Below the line we have one proof goal: the Hoare triple of the function body. In general, any C statement c might satisfy a Hoare-logic judgment $\Delta \vdash \{P\} c \{R\}$ when, in global context Δ , started in a state satisfying precondition P , statement c is sure not to crash and, if it terminates, the final state will satisfy R . We write the Hoare judgement in Coq as `semax (Δ : tycontext) (P : environ \rightarrow mpred) (c : statement) (R : ret_assert).`

Δ is a *type context*, giving types of function parameters, local variables, and global variables; and *specifications* (`funspec`) of global functions.

P is the precondition;

c is a command in the C language; and

R is the postcondition. Because a c statement can exit in different ways (fall-through, continue, break, return), a `ret_assert` has predicates for all of these cases.

Right after `start_function`, the command c is the entire function body.

Because we do *forward* Hoare-logic proof, we won't care about the postcondition until we get to the end of c , so here we hide it away in an abbreviation. Here, the command c is a long sequence starting with `i=0;...more`, and we hide the *more* in an abbreviation `MORE_COMMANDS`.

The precondition of this semax has LOCAL and SEP parts taken directly from the funspec (the PROP clauses have been moved above the line). The statement (Sset _i (Econst.int (Int.repr 0) tint)) is the AST generated by clightgen from the C statement `i=0;`.

10 forward

We do Hoare logic proof by forward symbolic execution. On [page 18](#) we show the proof goal at the beginning of the `sumarray` function body. In a forward Hoare logic proof of $\{P\} i = 0; \text{more} \{R\}$ we might first apply the sequence rule,

$$\frac{\{P\} i = 0; \{Q\} \quad \{Q\} \text{more} \{R\}}{\{P\} i = 0; \text{more} \{R\}}$$

assuming we could derive some appropriate assertion Q . For many kinds of statements (assignments, return, break, continue) this is done automatically by the forward tactic, which applies a strongest-postcondition style of proof rule to derive Q . When we execute forward here, the resulting proof goal is,

Espec, a, sh, contents, size, Delta_spec, SH, H, H0 *as before*

Delta := abbreviate : tycontext

POSTCONDITION := abbreviate : ret_assert

MORE_COMMANDS := abbreviate : statement

----- (1/1)

semax Delta

(PROP ())

LOCAL(temp _i (Vint (Int.repr 0)); temp _a a;

temp _n (Vint (Int.repr size)))

SEP(data_at sh (tarray tuint size) (map Vint (map Int.repr contents)) a))

(Ssequence (Sset _s (Econst_int (Int.repr 0) tuint)) MORE_COMMANDS)

POSTCONDITION

Notice that the precondition of this `semax` is really the *postcondition* Q of the `i=0;` statement; it is the precondition of the *next* statement, `s=0;.` It's much like the precondition of `i=0;` what has changed?

- The `LOCAL` part contains `temp _i (Vint (Int.repr 0))` in addition to what it had before; this says that the local variable i contains integer value zero.

- the command is now $s=0;more$, where `MORE_COMMANDS` no longer contains $s=0$;
- Delta has changed; it now records the information that i is initialized.

Applying the forward again will go through $s=0$; to yield a proof goal with a `LOCAL` binding for the `_s` variable.

FORWARD WORKS ON SEVERAL KINDS OF C COMMANDS. In each of the following cases, x must be a nonaddressable local variable, a temp.

$c_1; c_2$ Sequence of commands. The forward tactic will work on c_1 first.

$(c_1; c_2); c_3$ In this case, forward will re-associate the commands using the `seq.assoc` axiom, and work on $c_1; (c_2; c_3)$.

$x=E$; Assignment statement. Expression E must not contain memory dereferences (loads or stores using `*prefix`, `suffix[]`, or `->` operators). No restrictions on the form of the precondition (except that it must be in canonical form, `PROP/LOCAL/SEP`). The expression `&p→next` is permitted, since it does not actually load or store (it just computes an address).

$x= *E$; Memory load.

$x= a[E]$; Array load.

$x= E \rightarrow fld$; Field load.

$x= E \rightarrow f_1.f_2$; Nested field load; see [Chapter 30](#).

$x= E \rightarrow f_1[i].f_2$; Fields and subscripts; see [Chapter 30](#).

$E_1 = E_2$; Memory store. Expression E_2 must not dereference memory. Expression E_1 must be equivalent to a single memory store via some access *path* (see [Chapter 30](#)), and the precondition must contain an appropriate storable `data_at` or `field_at`.

if (E) C_1 else C_2 For an if-statement, use `forward_if` and (perhaps) provide a postcondition.

while (E) C For a while-loop, use the `forward_while` tactic ([page 25](#)) and provide a loop invariant.

break; The forward tactic works.

continue; The forward tactic works.

return E ; Expression E must not dereference memory, and the presence/absence of E must match the nonvoid/void return type of the function. The proof goal left by forward is to show that the precondition (with appropriate substitution for the abstract variable `ret_var`) entails the function's postcondition.

$x = f(a_1, \dots, a_n)$; For a function call, use `forward_call` (see [Chapter 19](#)).

11 If, While, For

To do forward proof through if-statements, while-loops, and for-loops, you need to provide additional information: join-postconditions, loop invariants, etc. The tactics are `forward_if`, `forward_while`, `forward_for`, `forward_for_simple_bound`.

If you're not sure which tactic to use, and with how many arguments, just use `forward`, and the error message will make a suggestion.

- **if** e **then** s_1 **else** s_2 ; $s_3 \dots$
Use `forward_if` Q , where Q is the *join postcondition*, the precondition of statement s_3 .
- **if** e **then** s_1 **else** s_2 ; $\}$ \dots When the if-statement appears at the end of a block, so the postcondition is already known, you can do `forward_if`. That is, you don't need to supply a join postcondition if POSTCONDITION is fully instantiated, without any unification variables. You can unfold abbreviate `in` POSTCONDITION to see.
- **while** (e) s ; \dots (no break statements in s)
You write `forward_while` Q , where Q is a loop invariant. See [Chapter 12](#).
- **while** (e) s ; \dots (with break statements in s)
You must treat this as if it were **for** $(;e;)$ s , and use the `forward_for` tactic; see below.
- **for** $(e_1; i < e_2; i++)$ s (no break statements in s)
If e_2 is loop-invariant, you may be able to use `forward_for_simple_bound`; see [Chapter 47](#).
- **for** $(e_1; e_2; e_3)$ s (no break statements in s)
Use `forward_for` Q Q' , where Q is the loop invariant and Q' is the assertion that goes right before the *increment* command, e_3 . See [Chapter 48](#).
- **for** $(e_1; e_2; e_3)$ s ; \dots
If s contains break statements *and* there are more commands after the loop, you will need to write `forward_for` Q Q' R , where Q and Q' are as above, and R is the join-postcondition. See [Chapter 48](#).

12 While loops

To prove a *while* loop by forward symbolic execution, you use the tactic `forward_while`, and you must supply a loop invariant. Take the example of the `forward_while` in `progs/verif_sumarray.v`. The proof goal is,

```

Espec, Delta_specs, Delta
a : val, sh : share, contents : list Z, size : Z
SH : readable_share sh
H : 0 ≤ size ≤ Int.max_signed
H0 : Forall (fun x : Z ⇒ 0 ≤ x ≤ Int.max_unsigned) contents
POSTCONDITION := abbreviate : ret_assert
MORE_COMMANDS, LOOP_BODY := abbreviate : statement
-----(1/1)

```

```

semax Delta
(PROP ()
  LOCAL(temp _s (Vint (Int.repr 0)); temp _i (Vint (Int.repr 0));
    temp _a a; temp _n (Vint (Int.repr size)))
  SEP(data_at sh (tarray tuint size) (map Vint (map Int.repr contents)) a))
(Ssequence
  (Swhile (Ebinop Olt (Etempvar _i tint) (Etempvar _n tint) tint)
    LOOP_BODY)
  MORE_COMMANDS)
POSTCONDITION

```

A loop invariant is an assertion, almost always in the form of an existential `EX...PROP(...)``LOCAL(...)``SEP(...)`. Each iteration of the loop has a state characterized by a different value of some iteration variable(s), the `EX` binds that value. The invariant for the `sumarray` loop is,

```

EX i: Z,
  PROP(0 ≤ i ≤ size)
  LOCAL(temp _a a; temp _i (Vint (Int.repr i)); temp _n (Vint (Int.repr size));
    temp _s (Vint (Int.repr (sum_Z (sublist 0 i contents)))))
  SEP(data_at sh (tarray tuint size) (map Vint (map Int.repr contents)) a).

```

The existential binds i , the iteration-dependent value of the local variable named $_i$. In general, there may be any number of EX quantifiers.

The `forward_while` tactic will generate four subgoals to be proved:

1. the precondition (of the whole loop) implies the loop invariant;
2. the loop-condition expression type-checks (i.e., guarantees to evaluate successfully);
3. the postcondition of the loop body implies the loop invariant;
4. the loop invariant (and negation of the loop condition) is a strong enough precondition to prove the `MORE_COMMANDS` after the loop.

Let's take a look at that first subgoal:

(above-the-line hypotheses elided) —1/4

```

ENTAIL Delta,
  PROP()
  LOCAL(temp _s (Vint (Int.repr 0)); temp _i (Vint (Int.repr 0));
        temp _a a; temp _n (Vint (Int.repr size)))
  SEP(data_at sh (tarray tuint size) (map Vint (map Int.repr contents)) a)
 $\vdash$  EX  $i : \mathbb{Z}$ ,
  PROP( $0 \leq i \leq \text{size}$ )
  LOCAL(temp _a a; temp _i (Vint (Int.repr i));
        temp _n (Vint (Int.repr size));
        temp _s (Vint (Int.repr (sum_Z (sublist 0 i contents)))))
  SEP(data_at sh (tarray tuint size) (map Vint (map Int.repr contents)) a)
  
```

This is an *entailment* goal; [Chapter 14](#) shows how to prove such goals.

13 PROP() LOCAL() SEP()

Each element of a SEP clause is a *spatial predicate*, that is, a predicate on some part of the memory. The Coq type for a spatial predicate is `mpred`; it can be thought of as `mem → Prop` (but is not quite the same, for quite technical semantic reasons).

The SEP represents the *separating conjunction* of its spatial predicates. When we write spatial predicates outside of a PROP/LOCAL/SEP, we use `*` instead of semicolon to indicate separating conjunction.

The LOCAL part of an assertion is a *local* predicate, that is, a predicate about the values of local variables. It's an *ordinary conjunction* (not separating) of its individual temp and lvar clauses. You can think of a local predicate as Coq type `environ → Prop`, where `environ` is the type of run-time local-variable frames.

A program assertion (precondition, postcondition, loop invariant, etc.) is a predicate *both* on its local-var *environ* and its memory. Its Coq type is `environ → mpred`. If you do the Coq command, `Check (PROP())LOCAL()SEP())` then Coq replies, `environ → mpred`. We call assertions of this type *lifted predicates*.

The *canonical form* of a lifted assertion is `PROP(\vec{P})LOCAL(\vec{Q})SEP(\vec{R})`, where \vec{P} is a list of propositions (`Prop`), where \vec{Q} is a list of local-variable definitions (`localdef`), and \vec{R} is a list of base-level assertions (`mpred`). Each list is semicolon-separated.

The existential quantifier `EX` can also be used on canonical forms, e.g., `EX x:T, PROP(\vec{P})LOCAL(\vec{Q})SEP(\vec{R})`.

14 Entailments

An *entailment* in separation logic, $P \vdash Q$, says that any state satisfying P must also satisfy Q . In Verifiable C, if P and Q are mpreds, then any mem satisfying P must also satisfy Q . If P and Q are *lifted predicates*, then any $\text{environ} \times \text{mem}$ satisfying P must also satisfy Q .

Usually we write lifted entailments as $\text{ENTAIL } \Delta, P \vdash Q$ in which Δ is the global type context, providing additional constraints on the state.

Verifiable C's *rule of consequence* is,

$$\frac{\text{ENTAIL } \Delta, P \vdash P' \quad \text{semax } \Delta P' \text{ c } Q' \quad \text{ENTAIL } \Delta, Q' \vdash Q}{\text{semax } \Delta P \text{ c } Q}$$

Using this axiom (called `semax_pre_post`) on a proof goal $\text{semax } \Delta P \text{ c } Q$ yields three subgoals: another `semax` and two (lifted) entailments, $\text{ENTAIL } \Delta, P \vdash P'$ and $\text{ENTAIL } \Delta, Q \vdash Q'$. P and Q are typically in the form $\text{PROP}(\vec{P})\text{LOCAL}(\vec{Q})\text{SEP}(\vec{R})$, perhaps with some EX quantifiers in the front. The turnstile \vdash is written in Coq as `|--`.

Let's consider the entailment arising from `forward_while` in the `progs/verif_sumarray.v` example:

$H : 0 \leq \text{size} \leq \text{Int.max_signed}$
(other above-the-line hypotheses elided) 1/4

ENTAIL Delta,
 PROP()
 LOCAL(temp _s (Vint (Int.repr 0)); temp _i (Vint (Int.repr 0));
 temp _a a; temp _n (Vint (Int.repr size)))
 SEP(data_at sh (tarray tuint size) (map Vint (map Int.repr contents)) a)
 $\vdash \text{EX } i : \mathbb{Z},$
 PROP($0 \leq i \leq \text{size}$)
 LOCAL(temp _a a; temp _i (Vint (Int.repr i));
 temp _n (Vint (Int.repr size));
 temp _s (Vint (Int.repr (sum_Z (sublist 0 i contents)))))
 SEP(data_at sh (tarray tuint size) (map Vint (map Int.repr contents)) a)

We instantiate the existential with the only value that works here, zero: **Exists 0**. [Chapter 22](#) explains how to handle existentials with **Intros** and **Exists**.

Now we use the `entailer!` tactic to solve as much of this goal as possible (see [Chapter 38](#)). In this case, the goal solves entirely automatically. In particular, $0 \leq i \leq \text{size}$ solves by `omega`; `sublist 0 0 contents` rewrites to `nil`; and `sum.Z nil` simplifies to `0`.

THE SECOND SUBGOAL of `forward_while` in `progs/verif_sumarray.v` is a *type-checking entailment*, of the form `ENTAIL Δ, PQR ⊢ tc_expr Δ e` where e is (the abstract syntax of) a C expression; in the particular case of a *while* loop, e is the negation of the loop-test expression. The assertion `tc_expr Δ e` says that executing e won't crash: all the variables it references exist and are initialized; and it doesn't divide by zero, et cetera.

In this case, the entailment concerns the expression $\neg(i < n)$,

`ENTAIL Delta, PROP(...) LOCAL(...) SEP(...)`

`⊢ tc_expr Delta`

`(Eunop Onotbool (Ebinop Olt (Etempvar _i tint) (Etempvar _n tint) tint)`
`tint)`

This solves completely via the `entailer!` tactic. To see why that is, instead of doing `entailer!`, do `unfold tc_expr; simpl`. You'll see that the right-hand side of the entailment simplifies down to `!!True`, (equivalent to `TT`, the “true” `mpred`). That's because the typechecker is *calculational*, as [Chapter 25 of Program Logics for Certified Compilers](#) explains.

15 Array subscripts

THE THIRD SUBGOAL of `forward_while` in `progs/verif_sumarray.v` is the *body* of the while loop: $\{x=a[i]; s+=x; i++;\}$.

This can be handled by three forward commands, but the first one needs a bit of extra help. To see why, try doing forward just *before* the `assert_PROP` instead of after. You'll see an error message saying that it can't prove $0 \leq i < \text{Zlength contents}$. Indeed, the command `x=a[i];` is safe only if *i* is in-bounds of the array *a*.

Let's examine the proof goal:

SH : readable_share *sh*

H : $0 \leq \text{size} \leq \text{Int.max_signed}$

H0 : Forall (fun x : Z $\Rightarrow 0 \leq x \leq \text{Int.max_unsigned}$) contents

i : Z

HRE : *i* < size

H1 : $0 \leq i \leq \text{size}$

----- (1/1)

semex Delta

(PROP ())

 LOCAL(temp _a *a*; temp _i (Vint (Int.repr *i*));

 temp _n (Vint (Int.repr size));

 temp _s (Vint (Int.repr (sum_Z (sublist 0 *i* contents)))))

 SEP(data_at *sh* (tarray tuint size) (map Vint (map Int.repr contents)) *a*))

(Ssequence

 (Sset _x

 (Ederef

 (Ebinop Oadd (Etempvar _a (tptr tuint)) (Etempvar _i tint)

 (tptr tuint)) tuint)) MORE_COMMANDS) POSTCONDITION

The Coq variable *i* was introduced automatically by `forward_while` from the existential variable, the EX *i*:Z of the loop invariant.

Going forward through $x=a[i]$; will be enabled by the `data_at` in the precondition, as long as the subscript value is less than the length of contents. One important property of `data_at` π (tarray τ n) σ p is that $n = \text{Zlength}(\sigma)$. If we had that fact above the line, then (using assumptions HRE and H) it would be easy to prove $0 \leq i < \text{Zlength contents}$.

Therefore, we write,

```
assert_PROP (Zlength contents = size). {
  entailer!. do 2 rewrite Zlength_map. reflexivity.
}
```

[Chapter 40](#) describes `assert_PROP`, which (like Coq's standard `assert`) will put `Zlength contents=size` above the line. The *first* subgoal of `assert_PROP` requires us to prove the proposition, *using facts from the current Hoare precondition* (which would not be accessible to Coq's standard `assert`). The reason this one is so easily provable is that `entailer!` extracts the $n = \text{Zlength}(\sigma)$ fact from `data_at` and puts it above the line.

The *second* subgoal is just like the subgoal we had before doing `assert_PROP`, but with the new proposition above the line. Now that `H_2: Zlength contents = size` is above the line, forward succeeds on the array subscript.

Two more forward commands take us to the end of the loop body.

16 At the end of the loop body

In `progs/verif_sumarray.v`, at the comment “Now we have reached the end of the loop body,” it is time to prove that the *current* precondition (which is the postcondition of the loop body) entails the loop invariant. This is the proof goal:

$H : 0 \leq \text{size} \leq \text{Int.max_signed}$

$H0 : \text{Forall } (\text{fun } x : Z \Rightarrow 0 \leq x \leq \text{Int.max_unsigned}) \text{ contents}$

$HRE : i < \text{size}$

$H1 : 0 \leq i \leq \text{size}$

(other above-the-line hypotheses elided)

ENTAIL Delta,

PROP()

LOCAL(temp _i (Vint (Int.add (Int.repr i) (Int.repr 1))));

temp _s

(force_val

(sem_add_default tint tint

(Vint (Int.repr (sum_Z (sublist 0 i contents)))))

(Znth i (map Vint (map Int.repr contents)) Vundef));

temp _x (Znth i (map Vint (map Int.repr contents)) Vundef);

temp _a a; temp _n (Vint (Int.repr size)))

SEP(data_at sh (tarray tuint size) (map Vint (map Int.repr contents)) a)

⊢ EX $a_0 : Z$,

PROP($0 \leq a_0 \leq \text{size}$)

LOCAL(temp _a a; temp _i (Vint (Int.repr a_0)));

temp _n (Vint (Int.repr size));

temp _s (Vint (Int.repr (sum_Z (sublist 0 a_0 contents)))))

SEP(data_at sh (tarray tuint size) (map Vint (map Int.repr contents)) a)

The right-hand side of this entailment is just the loop invariant. As usual at the end of a loop body, there is an existentially quantified variable that must be instantiated with an iteration-dependent value. In this case it's obvious: the quantified variable represents the contents of C local variable `_i`, so we do, **Exists** (`i+1`).

The resulting entailment has many trivial parts and a nontrivial residue. The usual way to get to the hard part is to run `entailer!`, which we do now. After clearing away the irrelevant hypotheses, we have:

$H : 0 \leq \text{Zlength contents} \leq \text{Int.max_signed}$

$\text{HRE} : i < \text{Zlength contents}$

$\text{H1} : 0 \leq i \leq \text{Zlength contents}$

----- $(1/1)$

$\text{Vint (Int.repr (sum_Z (sublist 0 (i + 1) contents)))} =$

$\text{Vint (Int.repr (sum_Z (sublist 0 i contents) + \text{Znth } i \text{ contents } 0))}$

Applying `f.equal` twice, leaves the goal,

$\text{sum_Z (sublist 0 (i + 1) contents)} =$

$\text{sum_Z (sublist 0 i contents) + Znth } i \text{ contents } 0$

Now the lemma `sublist.split` is helpful here:

`sublist.split`: $\forall l m h a, \quad 0 \leq l \leq m \leq h \leq |a| \rightarrow$

$\text{sublist } l \ h \ a = \text{sublist } l \ m \ a ++ \text{sublist } m \ h \ a$

So we do, rewrite `(sublist.split 0 i (i+1))` by `omega`. A bit more rewriting with the theory of `sum_Z` and `sublist` finishes the proof.

See also: [Chapter 55](#) (`sublist`).

17 Returning from a function

In `progs/verif_sumarray.v`, at the comment “After the loop,” we have reached the return statement. The forward tactic works here, leaving a proof goal that the precondition of the return entails the postcondition of the function-spec. (Sometimes the entailment solves automatically, leaving no proof goal at all.) The goal is a *lowered* entailment (on `mpred` assertions).

```
H4 : Forall (value_fits tuint) (map Vint (map Int.repr contents))
H2 : field_compatible (Tarray tuint (Zlength ...) noattr) [] a
  (other above-the-line hypotheses elided)
-----
data_at sh (tarray tuint (Zlength ...)) (map Vint (map Int.repr contents)) a
⊢ !!(Vint (Int.repr (sum_Z contents)) =
    Vint (Int.repr (sum_Z (sublist 0 i contents))))
```

The left-hand side of this entailment is a spatial predicate (`data_at`). Purely nonspatial facts (H4 and H2) derivable from it have already been inferred and moved above the line by `saturate_local` (see [Chapter 34](#)).

In general the right-hand side of a lowered entailment is $!!P \ \&\& \ R$, where P is a conjunction of propositions (`Prop`) and R is a separating conjunction of spatial predicates. The `!!` operator converts a `Prop` into an `mpred`.

This entailment’s right-hand side has no spatial predicates. That’s because, in the `sumarray` function, the SEP clause of the `funspec`’s postcondition had exactly the same `data_at` clause as we see here in the entailment precondition, and the entailment-solver called by `forward` has already cleared it away.

We can proceed by using `entailer`! The remaining subgoal solves easily in the theory of `sublists`. The proof of the function `sumarray` is now complete.

18 *Global variables and* `main()`

C programs may have “extern” global variables, either with explicit initializers or initialized by default. Any function that accesses a global variable must have the appropriate spatial assertions in its funspec’s precondition (and postcondition). But the main function is special: it has spatial assertions for *all* the global variables. Then it may pass these on, piecemeal, to the functions it calls on an as-needed basis.

The function-spec for the sumarray program’s main is,

Definition `main_spec` :=

```

DECLARE _main WITH u : unit
  PRE [ ] main_pre prog u
  POST [ tint ]
    (* application-specific postcondition *)
  PROP()
  LOCAL(temp ret_temp (Vint (Int.repr (1+2+3+4))))
  SEP(TT).
```

The first four lines are always the same for any program. `main_pre` calculates the precondition automatically from the list of extern global variables and initializers of the program.

Now, when we prove that main satisfies its funspec,

Lemma `body_main`: `semax_body Vprog Gprog f_main main_spec`.

Proof.

`start_function`.

the `start_function` tactic “unpacks” `main_pre` into an assertion:

```

v_four : val
----- (1/1)
semax Delta
  (PROP () LOCAL(gvar _four v_four)
    SEP(data_at Ews (tarray tuint 4)
      (map Vint [Int.repr 1; Int.repr 2; Int.repr 3; Int.repr 4]) v_four))
  (... function body ...)
POSTCONDITION

```

The `LOCAL` clause means that the C global variable `_four` is at memory address `v_four`. See [Chapter 32](#).

The `SEP` clause means that there's data of type “array of 4 integers” at address `v_four`, with access permission `Ews` and contents `[1;2;3;4]`. `Ews` stands for “external write share,” the standard access permission of extern global writable variables. See [Chapter 42](#).

The `sumarray` program's `main_spec` *postcondition* is specific to this program: we say that `main` returns the value `1 + 2 + 3 + 4`.

The `postcondition`'s `SEP` clause says `TT`; we cannot say simply `SEP()` because that is equivalent to `emp` in separation logic, enforcing the empty resource. But memory is not empty: it still contains all the initialized extern global variable `four`. So we give a looser spatial postcondition, `TT` (equivalent to `True` in separation logic).

19 Function calls

Continuing our example, the **Lemma** `body_main` in `verif_sumarray.v`:

Now it's time to prove the function-call statement, `s = sumarray(four,4)`. When proving a function call, one must supply a *witness* for the `WITH` clause of the function-spec. The `_sumarray` function's `WITH` clause ([page 14](#)) starts,

```
Definition sumarray_spec :=
DECLARE _sumarray
  WITH a: val, sh : share, contents : list Z, size: Z
```

so the type of the witness will be `(val*(share*(list Z * Z)))`. To choose the witness, examine your actual parameter values (along with the precondition of the funspec) to see what witness would be consistent; here, we use `(v_four,Ews,four_contents,4)` as follows:

```
forward_call (v_four,Ews,four_contents,4).
```

The `forward_call` tactic (usually) leaves subgoals: you must prove that your current precondition implies the funspec's precondition. Here, these solve easily, as shown in the proof script.

Finally, we are at the return statement. See [Chapter 17](#). In this case, the `forward` tactic is able to prove (using a form of the `entailer` tactic) that the current assertion implies the postcondition of `_main`.

20 Tying all the functions together

We build a whole-program proof by composing together the proofs of all the function bodies. Consider `Gprog`, the list of all the function-specifications:

Definition `Gprog : funspecs := sumarray_spec :: main_spec :: nil.`

Each `semax_body` proof says, assuming that *all the functions I might call* behave as specified, then *my own function-body* indeed behaves as specified:

Lemma `body_sumarray: semax_body Vprog Gprog f_sumarray sumarray_spec.`

Note that *all the functions I might call* might even include “myself,” in the case of a recursive or mutually recursive function.

This might seem like circular reasoning, but (for partial correctness) it is actually sound—by the miracle of step-indexed semantic models, as explained in Chapters 18 and 39 of *Program Logics for Certified Compilers*.

The rule for tying the functions together is called `semax_func`, and its use is illustrated in this theorem, the main proof-of-correctness theorem for the program `sumarray.c`:

Lemma `prog_correct: semax_prog prog Vprog Gprog.`

Proof.

`prove_semax_prog.`

`semax_func_cons body_sumarray.`

`semax_func_cons body_main.`

Qed.

The calls to `semax_func_cons` must appear in the same order as the functions appear in `prog.(prog_defs)`.

21 Separation logic: EX, *, emp, !!

These are the operators and primitives of *spatial predicates*, that is, the kind that can appear as conjuncts of a SEP.

$R ::=$	emp	empty
	TT	True
	FF	False
	$R_1 * R_2$	separating conjunction
	$R_1 \&\& R_2$	ordinary conjunction
	field_at $\pi \tau \vec{fld} \ v \ p$	“field maps-to”
	data_at $\pi \tau \ v \ p$	“maps-to”
	array_at $\tau \ \pi \ v \ lo \ hi$	array slice
	!! P	pure proposition
	EX $x : T, R$	existential quantification
	ALL $x : T, R$	universal quantification
	$R_1 \parallel R_2$	disjunction
	wand $R \ R'$	magic wand $R \multimap R'$
	...	other operators, including user definitions

22 EX, Intros, Exists

In a canonical-form lifted assertion, existentials can occur at the outside, or in one of the base-level conjuncts within the SEP clause. The left-hand side of this assertion has both:

```
ENTAIL  $\Delta$ ,      (* this example in progs/tutorial1.v *)
  EX  $x:Z$ ,
    PROP( $0 \leq x$ ) LOCAL(temp _i (Vint (Int.repr x)))
    SEP(EX  $y:Z$ ,  $!!(x < y) \ \&\& \text{data\_at } \pi \text{ tint (Vint (Int.repr } y)) \ p$ )
 $\vdash$  EX  $u: Z$ ,
  PROP( $0 < u$ ) LOCAL()
  SEP(data_at  $\pi$  tint (Vint (Int.repr  $u$ ))  $p$ )
```

To prove this entailment, one can first move x and y “above the line” by the tactic **Intros** a b:

```
 $a: Z$ 
 $b: Z$ 
 $H: 0 \leq a$ 
 $H0: a < b$ 
```

```
ENTAIL  $\Delta$ ,
  PROP() LOCAL(temp _i (Vint (Int.repr  $a$ )))
  SEP(data_at  $\pi$  tint (Vint (Int.repr  $b$ ))  $p$ )
 $\vdash$  EX  $u: Z$ ,
  PROP( $0 < u$ ) LOCAL()
  SEP(data_at  $\pi$  tint (Vint (Int.repr  $u$ ))  $p$ )
```

One might just as well say **Intros** $x \ y$ to use those names instead of $a \ b$. Note that the propositions (previously hidden inside existential quantifiers) have been moved above the line by **Intros**. Also, if there had been any separating-conjunction operators $*$ within the SEP clause, those will be “flattened” into semicolon-separated conjuncts within SEP.

Sometimes, even when there are no existentials to introduce, one wants

to move PROP propositions above the line and flatten the $*$ operators into semicolons. One can just say **Intros** with no arguments to do that.

If you want to Intro an existential *without* PROP-introduction and $*$ -flattening, you can just use **Intro** a , instead of **Intros** a .

Then, instantiate u by **Exists** b .

$a: Z$

$b: Z$

$H: 0 \leq a$

$H0: a < b$

ENTAIL Δ ,
 PROP() LOCAL(temp _i (Vint (Int.repr a)))
 SEP(data.at π tint (Vint (Int.repr b)) p)
 \vdash PROP($0 < b$) LOCAL()
 SEP(data.at π tint (Vint (Int.repr b)) p)

This entailment proves straightforwardly by entailer!.

23 Integers: nat, Z, int

Coq's standard library has the natural numbers `nat` and the integers `Z`.

C-language integer values are represented by the type `Int.int` (or just `int` for short), which are 32-bit two's complement signed or unsigned integers with mod-2³² arithmetic. [Chapter 51](#) describes the operations on the `int` type.

For most purposes, specifications and proofs of C programs should use `Z` instead of `int` or `nat`. Subtraction doesn't work well on naturals, and that screws up many other kinds of arithmetic reasoning. *Only when you are doing direct natural-number induction* is it natural to use `nat`, and so you might then convert using `Z.to_nat` to do that induction.

Conversions between `Z` and `int` are done as follows:

`Int.repr: Z → int.`
`Int.unsigned: int → Z.`
`Int.signed: int → Z.`

with the following lemmas:

$$\begin{array}{l}
 \text{Int.repr_unsigned} \frac{}{\text{Int.repr(Int.unsigned } z) = z} \\
 \text{Int.unsigned_repr} \frac{0 \leq z \leq \text{Int.max_unsigned}}{\text{Int.unsigned(Int.repr } z) = z} \\
 \text{Int.repr_signed} \frac{}{\text{Int.repr(Int.signed } z) = z} \\
 \text{Int.signed_repr} \frac{\text{Int.min_signed} \leq z \leq \text{Int.max_signed}}{\text{Int.signed(Int.repr } z) = z}
 \end{array}$$

`Int.repr` truncates to a 32-bit twos-complement representation (losing information if the input is out of range). `Int.signed` and `Int.unsigned` are different injections back to `Z` that never lose information.

When doing proofs about signed integers, you must prove that your integers never overflow; when doing proofs about unsigned integers, it's still a good idea to prove that you avoid overflow. That is, if the C variable `_x` contains the value `Vint (Int.repr x)`, then make sure `x` is in the appropriate range. Let's assume that `_x` is a signed integer, i.e. declared in C as `int x`; then the hypothesis is,

`H: Int.min_signed ≤ x ≤ Int.max_signed (* this example in progs/tutorial1.v *)`

If you maintain this hypothesis “above the line”, then Floyd's tactical proof automation can solve goals such as `Int.signed (Int.repr x) = x`. Also, to solve goals such as,

```
...
H2 : 0 ≤ n ≤ Int.max_signed (* this example in progs/tutorial1.v *)
```

```
...
```

```
-----
```

```
Int.min_signed ≤ 0 ≤ n
```

you can use the `reple_signed` tactic, which is basically just `omega` with knowledge of the values of `Int.min_signed`, `Int.max_signed`, and `Int.max_unsigned`.

To take advantage of this, put conjuncts into the `PROP` part of your function precondition such as `0 ≤ i < n; n ≤ Int.max_signed`. Then the `start_function` tactic will move them above the line, and the other tactics mentioned above will make use of them.

To see an example in action, look at `progs/verif_sumarray.v`. The `funspec`'s precondition contains,

```
PROP(... 0 ≤ size ≤ Int.max_signed;
        Forall (fun x ⇒ 0 ≤ x ≤ Int.max_unsigned) contents)
```

to ensure that `size` is representable as a nonnegative signed integer, and each element of `contents` is representable as an unsigned.

24 *Int, Int8, Int16, Int64, Ptrofs*

C programs use signed and unsigned integers of various sizes: 8-bit (signed char, unsigned char), 16-bit (signed short, unsigned short), 32-bit (int, unsigned int), 64-bit (long, unsigned long).

A C compiler may be “32-bit” in which case `sizeof(void*)=4` or “64-bit” in which case `sizeof(void*)=8`. The macro `size_t` is defined in the C standard library as a typedef for the appropriate signed integer, typically unsigned int on a 32-bit system and unsigned long on a 64-bit system.

To talk about integer *values* in all of these sizes, which have n -bit modular arithmetic (if unsigned) or n -bit twos-complement arithmetic (if signed), CompCert has several instantiations of the Integers module:

Int8 for char (signed or unsigned)
Int16 for short (signed or unsigned)
Int for int (signed or unsigned)
Int64 for long (signed or unsigned)
Ptrofs for `size_t`

where **Ptrofs** is isomorphic to the **Int** module (in 32-bit systems) and to the **Int64** module (in 64-bit systems). You pronounce “Ptrofs” as “pointer offset” because it is frequently used to indicate the distance between two pointers into the same object.

The following definitions are used for shorthand:

Definition `int = Int.int`.

Definition `int64 = Int64.int`.

Definition `ptrofs = Ptrofs.int`.

25 Values: Vint, Vptr

Definition $\text{block} : \text{Type} := \text{positive}$.

Inductive $\text{val} : \text{Type} :=$

```

| Vundef: val
| Vint: int → val
| Vlong: int64 → val
| Vfloat: float → val
| Vsingle: float32 → val
| Vptr: block → ptrofs → val.

```

Vundef is the *undefined* value—found, for example, in an uninitialized local variable.

Vint(i) is an integer value, where i is a CompCert 32-bit integer. These 32-bit integers can also represent short (16-bit) and char (8-bit) values.

Vfloat(f) is a 64-bit floating-point value.

Vsingle(f) is a 32-bit floating-point value.

Vptr $b\ z$ is a pointer value, where b is an abstract block number and z is an offset within that block. Different *malloc* operations, or different extern global variables, or stack-memory-resident local variables, will have different abstract block numbers. Pointer arithmetic must be done within the same abstract block, with $(\text{Vptr } b\ z) + (\text{Vint } i) = \text{Vptr } b\ (z + i)$. Of course, the C-language $+$ operator first multiplies i by the size of the array-element that $\text{Vptr } b\ z$ points to.

Vundef is not always treated as distinct from a defined value. For example, $p \mapsto \text{Vint } 5 \vdash p \mapsto \text{Vundef}$, where \mapsto is the `data_at` operator ([Chapter 29](#)). That is, $p \mapsto \text{Vundef}$ really means $\exists v, p \mapsto v$. Vundef could mean “truly uninitialized” or it could mean “initialized but arbitrary.”

26 *C types*

CompCert C describes C's type system with inductive data types.

Inductive signedness := Signed | Unsigned.

Inductive intsize := l8 | l16 | l32 | lBool.

Inductive floatsize := F32 | F64.

Record attr : Type := mk_attr {
 attr_volatile: bool; attr_alignas: option N
 }.

Definition noattr := { | attr_volatile := false; attr_alignas := None | }.

Inductive type : Type :=

| Tvoid: type
 | Tint: intsize → signedness → attr → type
 | Tlong: signedness → attr → type
 | Tfloat: floatsize → attr → type
 | Tpointer: type → attr → type
 | Tarray: type → Z → attr → type
 | Tfunction: typelist → type → calling_convention → type
 | Tstruct: ident → attr → type
 | Tunion: ident → attr → type

with typelist : Type :=

| Tnil: typelist
 | Tcons: type → typelist → typelist.

We have abbreviations for commonly used types:

Definition tint = Tint l32 Signed noattr.

Definition tuint = Tint l32 Unsigned noattr.

Definition tschar = Tint l8 Signed noattr.

Definition tuchar = Tint l8 Unsigned noattr.

Definition tarray (t: type) (n: Z) = Tarray t n noattr.

Definition tptr (t: type) := Tpointer t noattr.

27 CompSpecs

The C language has a namespace for struct- and union-identifiers, that is, *composite types*. In this example, struct foo {int value; struct foo *tail} a,b; the “global variables” namespace contains a,b, and the “struct and union” namespace contains foo.

When you use CompCert clightgen to parse myprogram.c into myprogram.v, the main definition it produces is prog, the AST of the entire C program:

Definition prog : Clight.program := { | prog_types := composites; ... | }.

To interpret the meaning of a type expression, we need to look up the names of its struct identifiers in a *composite* environment. This environment, along with various well-formedness theorems about it, is built from prog as follows:

Require Import VST.floyd.proofauto. (** Import Verifiable C library **)

Require Import myprogram. (** AST of my program **)

Instance CompSpecs : compspecs. **Proof.** make_compspecs prog. **Defined.**

The make_compspecs tactic automatically constructs the *composite specifications* from the program. As a typeclass Instance, CompSpecs is supplied automatically as an implicit argument to the functions and predicates that interpret the meaning of types:

Definition sizeof {env: composite_env} (t: type) : Z := ...

Definition data_at {cs: compspecs} (sh: share) (t: type) (v: val) := ...

@sizeof (@cenv.cs CompSpecs) (Tint I32 Signed noattr) = 4.

sizeof (Tint I32 Signed noattr) = 4.

sizeof (Tstruct _foo noattr) = 8.

@data_at CompSpecs sh t v ⊢ data_at sh t v

When you have two separately compiled .c files, each will have its own prog and its own compspecs. See [Chapter 65](#).

28 retype

For each C-language data type, we define a *representation type*, the Type of Coq values that represent the contents of a C variable of that type.

Definition `retype {cs: compspecs} (t: type) : Type := ...`

Lemma `retype_ind: $\forall (t: \text{type}),$`

`retype t =`

`match t with`

`| Tvoid \Rightarrow unit`

`| Tint _ _ \Rightarrow val`

`| Tlong _ _ \Rightarrow val`

`| Tfloat _ _ \Rightarrow val`

`| Tpointer _ _ \Rightarrow val`

`| Tarray t0 _ _ \Rightarrow list (retype t0)`

`| Tfunction _ _ _ \Rightarrow unit`

`| Tstruct id _ \Rightarrow retype_structlist (co_members (get_co id))`

`| Tunion id _ \Rightarrow retype_unionlist (co_members (get_co id))`

`end`

`retype_structlist` is the right-associative cartesian product of all the (retypes of) the fields of the struct. For example,

`struct list {int hd; struct list *tl};`

`struct one {struct list *p};`

`struct three {int a; struct list *p; double x};`

`retype (Tstruct _list noattr) = (val*val)`

`retype (Tstruct _one noattr) = val`

`retype (Tstruct _three noattr) = (val*(val*val))`

We use `val` instead of `int` for the retype of an integer variable, because the variable might be uninitialized, in which case its value will be `Vundef`.

29 *Uninitialized data*, default_val

CompCert represents uninitialized atomic (integer, pointer, float) values as `Vundef` : `val`.

The dependently typed function `default_val` calculates the undefined value for any C type:

`default_val`: $\forall \{cs: \text{compspecs}\} (t: \text{type}), \text{reptype } t.$

For any C type t , the default value for variables of type t will have Coq type `(reptype t)`.

For example:

```
struct list {int hd; struct list *tl;};
```

```
default_val tint = Vundef
```

```
default_val (tptr tint) = Vundef
```

```
default_val (tarray tint 4) = [Vundef; Vundef; Vundef; Vundef]
```

```
default_val (tarray  $t$   $n$ ) = list_repeat (Z.to_nat  $n$ ) (default_val  $t$ )
```

```
default_val (Tstruct _list noattr) = (Vundef, Vundef)
```

30 data_at

Consider a C program with these declarations:

```
struct list {int hd; struct list *tl;} L;
int f(struct list a[5], struct list *p) { ... }
```

Assume these definitions in Coq:

Definition t_list := Tstruct _list noattr.

Definition t_arr := Tarray t_list 5 noattr.

Somewhere inside `f`, we might have the assertion,

```
PROP() LOCAL(temp _a  $\alpha$ , temp _p  $p$ , gvar _L  $L$ )
SEP(data_at Ews t_list (Vint (Int.repr 0), nullval)  $L$ ;
    data_at  $\pi$  t_arr (list_repeat (Z.to_nat 5) (Vint (Int.repr 1),  $p$ ))  $\alpha$ ;
    data_at  $\pi$  t_list (default_val t_list)  $p$ )
```

This assertion says, “Local variable `_a` contains address α , `_p` contains address p , global variable `_L` is at address L . There is a struct list at L with permission-share `Ews` (“extern writable share”), whose `hd` field contains 0 and whose `tl` contains a null pointer. At address α there is an array of 5 list structs, each with `hd`=1 and `tl`= p , with permission π ; and at address p there is a single list cell that is uninitialized¹, with permission π .”

In pencil-and-paper separation logic, we write $q \mapsto i$ to mean `data_at Tsh tint (Vint (Int.repr i)) q` . We write $L \mapsto (0, \text{NULL})$ to mean `data_at Tsh t_list (Vint (Int.repr 0), nullval) L` . We write $p \mapsto (_, _)$ to mean `data_at π t_list (default_val t_list) p` .

In fact, the definition `data_at_` is useful for the situation $p \mapsto _$:

Definition data_at_ {cs: compspecs} sh t p := data_at sh t (default_val t) p.

¹Uninitialized, or initialized but we don’t know or don’t care what its value is

31 field_at

Consider the example in `progs/nest2.c`

```
struct a {double x1; int x2;};
struct b {int y1; struct a y2;};
struct b p;
```

The command `i = p.y2.x2;` does a nested field load. We call `y2.x2` the *field path*. The precondition for this command might include the assertion,

```
LOCAL(gvar _pb pb)  SEP(data_at sh t_struct_b (u,(v,w)) pb)
```

The postcondition (after the load) would include the new LOCAL fact, `temp _i w`.

The tactic `(unfold_data_at 1%nat)` changes the SEP part of the assertion as follows:

```
SEP(field_at Ews t_struct_b (DOT _y1) (Vint u) pb;
    field_at Ews t_struct_b (DOT _y2) (Vfloat v, Vint w) pb)
```

and then doing `(unfold_field_at 2%nat)` unfolds the second `field_at`,

```
SEP(field_at Ews t_struct_b (DOT _y1) (Vint u) pb;
    field_at Ews t_struct_b (DOT _y2 DOT _x1) (Vfloat v) pb;
    field_at Ews t_struct_b (DOT _y2 DOT _x2) (Vint w) pb)
```

The third argument of `field_at` represents the *path* of structure-fields that leads to a given substructure. The empty path (`nil`) works too; it “leads” to the entire structure. In fact, `data_at π τ v p` is just short for `field_at π τ nil v p` .

Arrays and structs may be nested together, in which case the field path may also contain array subscripts at the appropriate places, using the notation `SUB i` along with `DOT field`.

32 retype', repinj

This chapter is advanced material, describing a feature that is sometimes convenient but never necessary. You can skip this chapter.

```
struct a {double x1; int x2;};
struct b {int y1; struct a y2;} p;
repinj:  $\forall t: \text{type}, \text{retype}' t \rightarrow \text{retype } t$ 
retype t_struct_b = (val*(val*val))
retype' t_struct_b = (int*(float*int))
repinj t_struct_b (i,(x,j)) = (Vint i, (Vfloat x, Vint j))
```

The retype function maps C types to the the corresponding Coq types of (possibly uninitialized) values. When we know a variable is definitely initialized, it may be more natural to use int instead of val for integer variables, and float instead of val for double variables. The retype' function maps C types to the Coq types of (definitely initialized) values.

Definition $\text{retype}' \{cs: \text{compspecs}\} (t: \text{type}) : \text{Type} := \dots$

Lemma $\text{retype}'_ind: \forall (t: \text{type}),$
 $\text{retype } t =$

```
match t with
| Tvoid  $\Rightarrow$  unit
| Tint _ _  $\Rightarrow$  int
| Tlong _ _  $\Rightarrow$  Int64.int
| Tfloat _ _  $\Rightarrow$  float
| Tpointer _ _  $\Rightarrow$  pointer_val
| Tarray t0 _ _  $\Rightarrow$  list (retype' t0)
| Tfunction _ _ _  $\Rightarrow$  unit
| Tstruct id _  $\Rightarrow$  retype'_structlist (co_members (get_co id))
| Tunion id _  $\Rightarrow$  retype'_unionlist (co_members (get_co id))
end
```

The function repinj maps an initialized value to the type of possibly uninitialized values:

Definition $\text{repinj} \{cs: \text{compspecs}\} (t: \text{type}) : \text{retype}' t \rightarrow \text{retype } t := \dots$

The program `progs/nest2.c` (verified in `progs/verif_nest2.v`) illustrates the use of `retype'` and `repinj`.

```
struct a {double x1; int x2;};
struct b {int y1; struct a y2;} p;

int get(void) { int i; i = p.y2.x2; return i; }
void set(int i) { p.y2.x2 = i; }
```

Our API spec for `get` reads as,

Definition `get_spec` :=

```
DECLARE _get
  WITH v : retype' t_struct_b, p : val
  PRE []
    PROP() LOCAL(gvar _p p)
    SEP(data_at Ews t_struct_b (repinj _ v) p)
  POST [ tint ]
    PROP() LOCAL(temp ret_temp (Vint (snd (snd v))))
    SEP(data_at Ews t_struct_b (repinj _ v) p).
```

In this program, `retype' t_struct_b = (int*(float*int))`, and `repinj t_struct_b (i, (x, j)) = (Vint i, (Vfloat x, Vint j))`.

One could also have specified `get` without `retype'` at all:

Definition `get_spec` :=

```
DECLARE _get
  WITH i: Z, x: float, j: int, p : val
  PRE []
    PROP() LOCAL(gvar _p p)
    SEP(data_at Ews t_struct_b (Vint (Int.repr i), (Vfloat x, Vint j)) p)
  POST [ tint ]
    PROP() LOCAL(temp ret_temp (Vint j))
    SEP(data_at Ews t_struct_b (Vint (Int.repr i), (Vfloat x, Vint j)) p).
```

33 LOCAL *defs*: temp, lvar, gvar

The LOCAL part of a PROP()LOCAL()SEP() assertion is a list of localdefs that bind variables to their values or addresses.

Inductive localdef : Type :=
 | temp: ident → val → localdef
 | lvar: ident → type → val → localdef
 | gvar: ident → val → localdef
 | sgvar: ident → val → localdef
 | localprop: Prop → localdef.

temp *i v* binds a nonaddressable local variable *i* to its value *v*.

lvar *i t v* binds an *addressable* local variable *i* (of type *t*) to its *address v*.

gvar *i v* binds a *visible global* variable *i* to its *address v*.

sgvar *i v* binds a *possibly shadowed global* variable *i* to its *address v*.

The *contents* of an addressable (local or global) variable is on the heap, and can be described in the SEP clause.

```
int g=2;
int f(void) { int g; int *p = &g; g=6; return g; }
```

In this program, the global variable *g* is shadowed by the local variable *g*. In an assertion inside the function body, one could write

```
PROP() LOCAL(temp _p q; lvar _g tint q; sgvar _g p}
SEP(data_at Ews tint (Vint (Int.repr 2)) p;
    data_at Tsh tint (Vint (Int.repr 6)) q)
```

to describe a shadowed global variable *_g* that is still there in memory but (temporarily) cannot be referred to by its name in the C program.

34 go_lower

Normally one does not use this tactic directly, it is invoked as the first step of entailer or entailer!

Given a lifted entailment $\text{ENTAIL } \Delta, \text{PROP}(\vec{P}) \text{ LOCAL}(\vec{Q}) \text{ SEP}(\vec{R}) \vdash S$, one often wants to prove it at the base level: that is, with all of \vec{P} moved above the line, with all of \vec{Q} out of the way, just considering the base-level separation-logic conjuncts \vec{R} .

When $\Delta, \vec{P}, \vec{Q}, \vec{R}$ are *concrete*, the `go_lower` tactic does this. Concrete means that the \vec{P}, \vec{Q} are nil-terminated lists (not Coq variables) that every element of \vec{Q} is manifestly a `localdef` (not hidden in Coq abstractions), the identifiers in \vec{Q} are (computable to) ground terms, and the analogous (tree) property for Δ . It is not necessary that $\Delta, \vec{P}, \vec{Q}, \vec{R}$ be fully *ground terms*: Coq variables (and other Coq abstractions) can appear anywhere in \vec{P} and \vec{R} and in the *value* parts of Δ and \vec{Q} . When the entailment is not fully concrete, or when there existential quantifiers outside `PROP`, the tactic `old_go_lower` can still be useful.

`go_lower` moves the propositions \vec{P} above the line; when a proposition is an equality on a Coq variable, it substitutes the variable.

For each `localdef` in \vec{Q} (such as `temp i v`), `go_lower` looks up i in Δ to derive a type-checking fact (such as `tc_val t v`), then introduces it above the line and simplifies it. For example, if t is `tptr tint`, then the typechecking fact simplifies to `is_pointer_or_null v`.

Then it proves the `localdefs` in S , if possible. If there are still some local-environment dependencies remaining in S , it introduces a variable `rho` to stand for the run-time environment.

The remaining goal will be of the form $\vec{R} \vdash S'$, with the semicolons in \vec{R} replaced by the separating conjunction `*`. S' is the residue of S after lowering to the base separation logic and deleting its (provable) `localdefs`.

35 *saturate_local*

Normally one does not use this tactic directly, it is invoked by *entailer* or *entailer!*

To prove an entailment $R_1 * R_2 * \dots * R_n \vdash!! (P'_1 \wedge \dots \wedge P'_n) \&\& R'_1 * \dots * R'_m$, first extract all the *local (nonspatial)* facts from $R_1 * R_2 * \dots * R_n$, use them (along with other propositions above the line) to prove $P'_1 \wedge \dots \wedge P'_n$, and then work on the separation-logic (spatial) conjuncts $R_1 * \dots * R_n \vdash R'_1 * \dots * R'_m$.

An example local fact: $\text{data_at } \text{Ews } (\text{tarray tint } n) \ v \ p \vdash!! (\text{Zlength } v = n)$. That is, the value v in an array “fits” the length of the array.

The Hint database *saturate_local* contains all the local facts that can be extracted from *individual* spatial conjuncts:

field_at_local_facts:

$$\begin{aligned} \text{field_at } \pi \ t \ \text{path} \ v \ p \vdash!! & (\text{field_compatible } t \ \text{path} \ p \\ & \wedge \text{value_fits } (\text{nested_field_type } t \ \text{path}) \ v) \\ \text{data_at } \pi \ t \ v \ p \vdash!! & (\text{field_compatible } t \ \text{nil} \ p \wedge \text{value_fits } t \ v) \end{aligned}$$

memory_block_local_facts:

$$\text{memory_block } \pi \ n \ p \vdash!! \text{isptr } p$$

The assertion $(\text{Zlength } v = n)$ is actually a consequence of *value_fits* when t is an array type. See [Chapter 36](#).

If you create user-defined spatial terms (perhaps using EX, *data_at*, etc.), you can add hints to the *saturate_local* database as well.

The tactic *saturate_local* takes a proof goal of the form $R_1 * R_2 * \dots * R_n \vdash S$ and adds *saturate-local* facts for *each* of the R_i , though it avoids adding duplicate hypotheses above the line.

36 *field_compatible, field_address*

CompCert C light comes with an “address calculus.” Consider this example:

```
struct a {double x1; int x2;};
struct b {int y1; struct a y2;};
struct a *pa; int *q = &(pa→y2.x2);
```

Suppose the value of `_pa` is p . Then the value of `_q` is $p + \delta$; how can we reason about δ ?

Given type t such as `Tstruct _b noattr`, and $path$ such as `(DOT _y2 DOT _x2)`, then `(nested_field_type t path)` is the type of the field accessed by that path, in this case `tint`; `(nested_field_offset t path)` is the distance (in bytes) from the base of t to the address of the field, in this case (on a 32-bit machine) 12 or 16, depending on the field-alignment conventions of the target machine (and the compiler).

On the Intel x86 architecture, where doubles need not be 8-byte-aligned, we have,

$$\text{data_at } \pi \text{ t_struct_b } (i, (f, j)) \ p \vdash \\ \text{data_at } \pi \text{ tint } i \ p * \text{data_at } \pi \text{ t_struct_a } (f, j) \ (\text{offset_val } p \ 12)$$

but the converse is not valid:

$$\text{data_at } \pi \text{ tint } i \ p * \text{data_at } \pi \text{ t_struct_a } (f, j) \ (\text{offset_val } p \ 12) \\ \not\vdash \text{data_at } \pi \text{ t_struct_b } (i, (f, j)) \ p$$

The reasons: we don’t know that $p + 12$ satisfies the alignment requirements for struct b; we don’t know whether $p + 12$ crosses the end-of-memory boundary. That entailment *would* be valid in the presence of this hypothesis: `field_compatible t_struct_b nil p : Prop`.

which says that an entire struct b value *can* fit at address p . Note that

this is a *nonspatial* assertion about *addresses*, independent of the *contents* of memory.

In order to assist with reasoning about reassembly of data structures, `saturate_local` (and therefore `entailer`) puts `field_compatible` assertions above the line; see [Chapter 34](#).

Sometimes one needs to name the address of an internal field—for example, to pass just that field to a function. In that case, one *could* use `field_offset`, but it is better to use `field_address`:

```
Definition field_address (t: type) (path: list gfield) (p: val) : val :=
  if field_compatible_dec t path p
  then offset_val (Int.repr (nested_field_offset t path)) p
  else Vundef
```

That is, `field_address` has “baked in” the fact that the offset is “compatible” with the base address (is aligned, has not crossed the end-of-memory boundary). Therefore we get a valid converse for the example above:

```
data_at  $\pi$  tint i p
  * data_at  $\pi$  t_struct_a (f,j) (field_address t_struct_b (DOT _y2 DOT _x2) p)
 $\vdash$  data_at  $\pi$  t_struct_b (i,(f,j)) p
```

FIELD_ADDRESS VS FIELD_ADDRESS0. You use `field_address t path p` to indicate that *p* points to **at least one** thing of the appropriate field type for *t.path*, that is, the type `nested_field_type t path`.

Sometimes when dealing with arrays, you want a pointer that might possibly point just one past the end of the array; that is, points to **at least zero** things. In this case, use `field_address0 t path p`, which is built from `field_compatible0`. It has slightly looser requirements for how close *p* can be to the end of memory.

37 *value_fits*

The spatial maps-to assertion, $\text{data_at } \pi \ t \ v \ p$, says that there's a value v in memory at address p , filling the data structure whose C type is t (with permission π). A corollary is $\text{value_fits } t \ v$: v is a value that actually *can* reside in such a C data structure.

Value_fits is a recursive, dependently typed relation that is easier described by its induction relation; here, we present a simplified version that assumes that all types t are not volatile:

```

value_fits t v = tc_val' t v    (when t is an integer, float, or pointer type)
value_fits (tarray t' n) v = (Zlength v = Z.max 0 n)  $\wedge$  Forall (value_fits t') v
value_fits (Tstruct i noattr) (v1, (v2, (... , vn))) =
    value_fits (field_type f1 v1)  $\wedge$  ...  $\wedge$  value_fits (field_type fn vn)
    (when the fields of struct i are f1, ..., fn)

```

The predicate tc_val' says,

Definition $\text{tc_val' } (t: \text{type}) (v: \text{val}) := v \neq \text{Vundef} \rightarrow \text{tc_val } t \ v$.

Definition $\text{tc_val } (t: \text{type}) : \text{val} \rightarrow \text{Prop} :=$

```

match t with
| Tvoid  $\Rightarrow$  False
| Tint sz sg _  $\Rightarrow$  is_int sz sg
| Tlong _ _  $\Rightarrow$  is_long
| Tfloat F32 _  $\Rightarrow$  is_single
| Tfloat F64 _  $\Rightarrow$  is_float
| Tpointer _ _ | Tarray _ _ _ | Tfunction _ _ _  $\Rightarrow$  is_pointer_or_null
| Tstruct _ _ | Tunion _ _  $\Rightarrow$  isptr
end

```

So, an atomic value (int, float, pointer) fits *either* when it is Vundef or when it type-checks. We permit Vundef to “fit,” in order to accommodate partially initialized data structures in C.

Since τ is usually concrete, `tc_val τ v` immediately unfolds to something like,

```
TC0: is_int l32 Signed (Vint i)
TC1: is_int l8 Unsigned (Vint c)
TC2: is_int l8 Signed (Vint d)
TC3: is_pointer_or_null p
TC4: is_ptr q
```

TC0 says that i is a 32-bit signed integer; this is a tautology, so it will be automatically deleted by `go_lower`.

TC1 says that c is a 32-bit signed integer whose value is in the range of unsigned 8-bit integers (unsigned char). TC2 says that d is a 32-bit signed integer whose value is in the range of signed 8-bit integers (signed char). These hypotheses simplify to,

```
TC1: 0 ≤ Int.unsigned c ≤ Byte.max_unsigned
TC2: Byte.min_signed ≤ Int.signed c ≤ Byte.max_signed
```

38 *cancel*

The *cancel* tactic proves associative-commutative rearrangement goals such as $(A_1 * A_2) * ((A_3 * A_4) * A_5) \vdash A_4 * (A_5 * A_1) * (A_3 * A_2)$.

If the goal has the form $(A_1 * A_2) * ((A_3 * A_4) * A_5) \vdash (A_4 * B_1 * A_1) * B_2$ where there is only a partial match, then *cancel* will remove the matching conjuncts and leave a subgoal such as $A_2 * A_3 * A_5 \vdash B_1 * B_2$.

cancel solves $(A_1 * A_2) * ((A_3 * A_4) * A_5) \vdash A_4 * \text{TT} * A_1$ by absorbing $A_2 * A_3 * A_5$ into TT . If the goal has the form

$$\frac{F := ?224 : \text{list}(\text{environ} \rightarrow \text{mpred})}{(A_1 * A_2) * ((A_3 * A_4) * A_5) \vdash A_4 * (\text{fold_right sepcon emp } F) * A_1}$$

where F is a *frame* that is an abbreviation for an uninstantiated logical variable of type $\text{list}(\text{environ} \rightarrow \text{mpred})$, then the *cancel* tactic will perform *frame inference*: it will unfold the definition of F , instantiate the variable (in this case, to $A_2 :: A_3 :: A_5 :: \text{nil}$), and solve the goal. The frame may have been created by $\text{evar}(F : \text{list}(\text{environ} \rightarrow \text{mpred}))$, as part of forward symbolic execution through a function call.

WARNING: *cancel* can turn a provable entailment into an unprovable entailment. Consider this:

$$\frac{A * C \vdash B * C}{A * D * C \vdash C * B * D}$$

This goal is provable by first rearranging to $(A * C) * D \vdash (B * C) * D$. But *cancel* may aggressively cancel C and D , leaving $A \vdash B$, which is not provable. You might wonder, what kind of crazy hypothesis is $A * C \vdash B * C$; but indeed such “context-dependent” cancellations do occur in the theory of linked lists; see PLCC Chapter 19.

CANCEL DOES *not* USE $\beta\eta$ equality, as that could be slow in some cases. That means sometimes *cancel* leaves a residual subgoal $A \vdash A'$ where $A =_\beta A'$; sometimes the only differences are in (invisible) implicit arguments. You can apply *derives_refl* to solve such residual goals.

39 *entailer!*

The *entailer* and *entailer!* tactics simplify (or solve entirely) entailments in either the lifted or base-level separation logic. The *entailer* never turns a provable entailment into an unprovable one; *entailer!* is more aggressive and more efficient, but sometimes (rarely) turns a provable entailment into an unprovable one. We recommend trying *entailer!* first.

When *go_lower* is applicable, the *entailers* start by applying it (see [Chapter 33](#)).

Then: *saturate_local* (see [Chapter 34](#)).

NEXT: on each side of the entailment, gather the propositions to the left: $R_1 * (!!P_1 \&\& (!!P_2 \&\& R_2))$ becomes $!!(P_1 \wedge P_2) \&\& (R_1 * R_2)$.

Move all left-hand-side propositions above the line; substitute variables. Autorewrite with *entailer_rewrite*, a *modest* hint database. If the r.h.s. or its first conjunct is a “*valid_pointer*” goal (or one of its variants), try to solve it.

At this point, *entailer* tries *normalize* and (if progress) back to NEXT; *entailer!* applies *cancel* to the spatial terms and *prove_it_now* to each propositional conjunct.

The result is that either the goal is entirely solved, or a residual entailment or proposition is left for the user to prove.

40 *normalize*

The `normalize` tactic performs autorewrite **with** norm and several other transformations. **Normalize can be slow: use Intros and entailer if they can do the job.**

The norm rewrite-hint database uses several sets of rules.

Generic separation-logic simplifications.

$$\begin{array}{llll}
 P * \text{emp} = P & \text{emp} * P = P & P \&\& \text{TT} = P & \text{TT} \&\& P = P \\
 P \&\& \text{FF} = \text{FF} & \text{FF} \&\& P = \text{FF} & P * \text{FF} = \text{FF} & \text{FF} * P = \text{FF} \\
 P \&\& P = P & (\text{EX } _ : A, P) = P & \text{local 'True} = \text{TT}
 \end{array}$$

Pull EX and !! out of *-conjunctions.

$$\begin{array}{ll}
 (\text{EX } x : A, P) * Q = \text{EX } x : A, P * Q & (\text{EX } x : A, P) \&\& Q = \text{EX } x : A, P \&\& Q \\
 P * (\text{EX } x : A, Q) = \text{EX } x : A, P * Q & P \&\& (\text{EX } x : A, Q) = \text{EX } x : A, P \&\& Q \\
 P * (!!Q \&\& R) = !!Q \&\& (P * R) & (!!Q \&\& P) * R = !!Q \&\& (P * R)
 \end{array}$$

Delete auto-provable propositions.

$$P \rightarrow (!!P \&\& Q = Q) \qquad P \rightarrow (!!P = \text{TT})$$

Integer arithmetic.

$$\begin{array}{llllll}
 n + 0 = n & 0 + n = n & n * 1 = n & 1 * n = n & \text{sizeof tuchar} = 1 \\
 \text{align } n \ 1 = n & (z > 0) \rightarrow (\text{align } 0 \ z = 0) & (z \geq 0) \rightarrow (\text{Z.max } 0 \ z = z)
 \end{array}$$

Int32 arithmetic.

$$\text{Int.sub } x \ x = \text{Int.zero}$$

$$\text{Int.sub } x \ \text{Int.zero} = x$$

$$\text{Int.add } x \ (\text{Int.neg } x) = \text{Int.zero}$$

$$\text{Int.add } x \ \text{Int.zero} = x$$

$$\text{Int.add } \text{Int.zero } x = x$$

$$x \neq y \rightarrow \text{offset_val}(\text{offset_val } v \ i) \ j = \text{offset_val } v \ (\text{Int.add } i \ j)$$

$$\text{Int.add}(\text{Int.repr } i)(\text{Int.repr } j) = \text{Int.repr}(i + j)$$

$$\text{Int.add}(\text{Int.add } z \ (\text{Int.repr } i)) \ (\text{Int.repr } j) = \text{Int.add } z \ (\text{Int.repr}(i + j))$$

$$z > 0 \rightarrow (\text{align } 0 \ z = 0)$$

$$\text{force_int}(\text{Vint } i) = i$$

$$(\text{min_signed} \leq z \leq \text{max_signed}) \rightarrow \text{Int.signed}(\text{Int.repr } z) = z$$

$$(0 \leq z \leq \text{max_unsigned}) \rightarrow \text{Int.unsigned}(\text{Int.repr } z) = z$$

$$(\text{Int.unsigned } i < 2^n) \rightarrow \text{Int.zero_ext } n \ i = i$$

$$(-2^{n-1} \leq \text{Int.signed } i < 2^{n-1}) \rightarrow \text{Int.sign_ext } n \ i = i$$

map, fst, snd, ...

$$\text{map } f \ (x :: y) = f \ x :: \text{map } f \ y$$

$$\text{map } \text{nil} = \text{nil}$$

$$\text{fst}(x, y) = x$$

$$\text{snd}(x, y) = y$$

$$(\text{isptr } v) \rightarrow \text{force_ptr } v = v$$

$$\text{isptr } (\text{force_ptr } v) = \text{isptr } v$$

$$(\text{is_pointer_or_null } v) \rightarrow \text{ptr_eq } v \ v = \text{True}$$

Unlifting.

$$'f \ \rho = f \ \text{[when } f \text{ has arity 0]}$$

$$'f \ a_1 \ \rho = f \ (a_1 \ \rho) \ \text{[when } f \text{ has arity 1]}$$

$$'f \ a_1 \ a_2 \ \rho = f \ (a_1 \ \rho) \ (a_2 \ \rho) \ \text{[when } f \text{ has arity 2, etc.]}$$

$$(P * Q)\rho = P\rho * Q\rho$$

$$(P \ \&\& \ Q)\rho = P\rho \ \&\& \ Q\rho$$

$$(!P)\rho = !!P$$

$$!!(P \wedge Q) = !!P \ \&\& \ !!Q$$

$$(\text{EX } x : A, P x)\rho = \text{EX } x : A, P x \rho$$

$$(\text{EX } x : B, P x) = \text{EX } x : B, '(P x)$$

$$'(P * Q) = 'P * 'Q$$

$$'(P \ \&\& \ Q) = 'P \ \&\& \ 'Q$$

Type checking and miscellaneous.

$$\text{tc_andp tc_TT } e = e \qquad \text{tc_andp } e \text{ tc_TT} = e$$

$$\text{eval_id } x \text{ (env_set } \rho \ x \ v) = v$$

$$x \neq y \rightarrow (\text{eval_id } x \text{ (env_set } \rho \ y \ v) = \text{eval_id } x \ v)$$

$$\text{isptr } v \rightarrow (\text{eval_cast_neutral } v = v)$$

$$(\exists t. \text{tc_val } t \ v \wedge \text{is_pointer_type } t) \rightarrow (\text{eval_cast_neutral } v = v)$$

Expression evaluation. (autorewrite with eval, but in fact these are usually handled just by simpl or unfold.)

$$\text{deref_noload}(\text{tarray } t \ n) = (\text{fun } v \Rightarrow v) \qquad \text{eval_expr}(\text{Etempvar } i \ t) = \text{eval_id } i$$

$$\text{eval_expr}(\text{Econst_int } i \ t) = \text{'(Vint } i)$$

$$\text{eval_expr}(\text{Ebinop } op \ a \ b \ t) =$$

$$\text{'(eval_binop } op \ (\text{typeof } a) \ (\text{typeof } b)) \ (\text{eval_expr } a) \ (\text{eval_expr } b)$$

$$\text{eval_expr}(\text{Eunop } op \ a \ t) = \text{'(eval_unop } op \ (\text{typeof } a)) \ (\text{eval_expr } a)$$

$$\text{eval_expr}(\text{Ecast } e \ t) = \text{'(eval_cast}(\text{typeof } e) \ t) \ (\text{eval_expr } e)$$

$$\text{eval_lvalue}(\text{Ederef } e \ t) = \text{'force_ptr } (\text{eval_expr } e)$$

Function return values.

$$\text{get_result}(\text{Some } x) = \text{get_result1}(x) \qquad \text{retval}(\text{get_result1 } i \ \rho) = \text{eval_id } i \ \rho$$

$$\text{retval}(\text{env_set } \rho \ \text{ret_temp } v) = v$$

$$\text{retval}(\text{make_args}(\text{ret_temp} :: \text{nil}) \ (v :: \text{nil}) \ \rho) = v$$

$$\text{ret_type}(\text{initialized } i \ \Delta) = \text{ret_type}(\Delta)$$

Postconditions. (autorewrite with ret_assert.)

$$\text{normal_ret_assert FF ek vl} = \text{FF}$$

$$\text{frame_ret_assert}(\text{normal_ret_assert } P) Q = \text{normal_ret_assert } (P * Q)$$

$$\text{frame_ret_assert } P \text{ emp} = P$$

$$\text{frame_ret_assert } P Q \text{ EK_return vl} = P \text{ EK_return vl} * Q$$

$$\text{frame_ret_assert}(\text{loop1_ret_assert } P Q) R =$$

$$\text{loop1_ret_assert } (P * R)(\text{frame_ret_assert } Q R)$$

$$\text{frame_ret_assert}(\text{loop2_ret_assert } P Q) R =$$

$$\text{loop2_ret_assert } (P * R)(\text{frame_ret_assert } Q R)$$

$$\text{overridePost } P (\text{normal_ret_assert } Q) = \text{normal_ret_assert } P$$

$$\text{normal_ret_assert } P \text{ ek vl} = (!!(\text{ek} = \text{EK_normal}) \&\& (!!(\text{vl} = \text{None}) \&\& P))$$

$$\text{loop1_ret_assert } P Q \text{ EK_normal None} = P$$

$$\text{overridePost } P R \text{ EK_normal None} = P$$

$$\text{overridePost } P R \text{ EK_return} = R \text{ EK_return}$$

IN ADDITION TO REWRITING, normalize applies the following lemmas:

$$P \vdash \text{TT} \quad \text{FF} \vdash P \quad P \vdash P * \text{TT} \quad (\forall x. (P \vdash Q)) \rightarrow (\text{EX } x : A, P \vdash Q)$$

$$(P \rightarrow (\text{TT} \vdash Q)) \rightarrow (!P \vdash Q) \quad (P \rightarrow (Q \vdash R)) \rightarrow (!P \&\& Q \vdash R)$$

and does some rewriting and substitution when P is an equality in the goal, $(P \rightarrow (Q \vdash R))$.

Given the goal $x \rightarrow P$, where x is not a Prop, normalize avoids doing an intro. This allows the user to choose an appropriate name for x .

41 *assert_PROP*

Consider the proof state of `verif_sumarray.v`, just after (`* Prove postcondition of loop body implies loop invariant. *`). We have,

```

H : 0 ≤ i ≤ size
-----
semax Delta
  (PROP () LOCAL(...))
  SEP(data_at sh (tarray tuint size) (map Vint (map Int.repr contents)) a))
  x = x[i]; ...
  POSTCONDITION

```

We desire, above the line, `Zlength contents = size`. This is not provable from anything above the line. But it is provable from the precondition (`PROP/LOCAL/SEP`).

Whenever a pure proposition (`Prop`) is provable from the precondition, you can bring it above the line using `assert_PROP`.

For example, `assert_PROP(Zlength contents = size)` gives you an entailment proof goal:

```

H : 0 ≤ i ≤ size
-----
ENTAIL Delta,
  (PROP () LOCAL(...))
  SEP(data_at sh (tarray tuint size) (map Vint (map Int.repr contents)) a))
  ⊢ !! (Zlength contents = size).

```

Then, typically, you use `entailer` to prove the assertion. For example:

```

assert_PROP (Zlength contents = size). {
  entailer!. do 2 rewrite Zlength_map.reflexivity.
}

```

42 Welltypedness of variables

Verifiable C’s typechecker ensures this about C-program variables: if a variable is initialized, then it contains a value of its declared type.

Function parameters (accessed by Etempvar expressions) are always initialized. Nonaddressable local variables (accessed by Etempvar expressions) and address-taken local variables (accessed by Evar) may be uninitialized or initialized. Global variables (accessed by Evar) are always initialized.

The typechecker keeps track of the initialization status of local nonaddressable variables, *conservatively*: if on all paths from function entry to the current point—assuming that the conditions on if-expressions and while-expressions are uninterpreted/nondeterministic—there is an assignment to variable x , then x is known to be initialized.

Addressable local variables do not have initialization status tracked by the typechecker; instead, this is tracked in the separation logic, by `data.at` assertions such as $v \mapsto _$ (uninitialized) or $v \mapsto i$ (initialized).

Proofs using the forward tactic will typically generate proof obligations (for the user to solve) of the form,

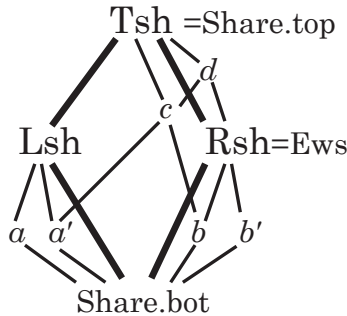
$$\text{ENTAIL } \Delta, \text{PROP}(\vec{P}) \text{ LOCAL}(\vec{Q}) \text{ SEP}(\vec{R}) \vdash \text{PROP}(\vec{P}') \text{ LOCAL}(\vec{Q}') \text{ SEP}(\vec{R}')$$

Δ keeps track of which nonaddressable local variables are initialized; says that all references to local variables contain values of the right type; and says that all addressable locals and globals point to an appropriate block of memory.

Using `go_lower` or `entailer` on an `ENTAIL` goal causes a `tc_val` assertion to be placed above the line for each initialized tempvar. As explained at [page 58](#), this `tc_val` may be simplified into an `is_int` hypothesis, or even removed if vacuous.

43 Shares

Operators such as `data_at` take a *permission share*, expressing whether the assertion grants read permission, write permission, or some other fractional permission.



The *top* share, written `Tsh` or `Share.top`, gives total permission: to deallocate any cells within the footprint of this mapsto, to read, to write.

<code>Share.split Tsh = (Lsh, Rsh)</code>	
<code>Share.split Lsh = (a, a')</code>	<code>Share.split Rsh = (b, b')</code>
$a' \oplus b = c$	$\text{lub}(c, \text{Rsh}) = a' \oplus \text{Rsh} = d$
$\forall sh. \text{writable_share } sh \rightarrow \text{readable_share } sh$	
<code>writable_share Ews</code>	<code>readable_share b</code>
<code>writable_share d</code>	<code>readable_share c</code>
<code>writable_share Tsh</code>	$\neg \text{readable_share Lsh}$

Any share may be split into a *left half* and a *right half*. The left and right of the top share are given distinguished names `Lsh`, `Rsh`.

The right-half share of the top share (or any share containing it such as `d`) is sufficient to grant *write permission* to the data: “the right share is the write share.” A thread of execution holding only `Lsh`—or subshares of it such as `a, a'`—can neither read or write the object, but such shares are not completely useless: holding any nonempty share prevents other threads from deallocating the object.

Any subshare of `Rsh`, in fact any share that overlaps `Rsh`, grants *read*

permission to the object. Overlap can be tested using the `glb` (greatest lower bound) operator.

Whenever $(\text{data_at } sh \ t \ w \ v)$ holds, then the share sh must include at least a read share, thus this gives permission to load memory at address v to get a value w of type t .

To make sure sh has enough permission to write (i.e., $Rsh \subset sh$, we can say `writable_share sh : Prop`.

To test whether a share sh is empty or nonempty, use `sepalg.identity sh` or `sepalg.nonidentity sh`.

Memory obtained from `malloc` comes with the top share Tsh . Writable extern global variables and stack-allocated addressable locals (which of course must not be deallocated) come with the “extern writable share” Ews which is equal to Rsh . Read-only globals come with a half-share of Rsh .

Sequential programs usually have little need of any shares except the Tsh and Ews . However, many function specifications can be parameterized over any share (example: `sumarray_spec` on [page 14](#)); that kind of generalized specification makes the functions usable in more contexts.

In C it is undefined to test deallocated pointers for equality or inequalities, so the Hoare-logic rule for pointer comparison also requires some permission-share; see [page 70](#).

44 *Pointer comparisons*

In C, if p and q are expressions of type pointer-to-something, testing $p=q$ or $p!=q$ is defined only if: p is NULL, or points within a currently allocated object, or points at the end of a currently allocated object; and similarly for q . Testing $p<q$ (etc.) has even stricter requirements: p and q must be pointers into the *same* allocated object.

Verifiable C enforces this by creating “type-checking” conditions for the evaluation of such pointer-comparison expressions. Before reasoning about the result of evaluating expression $p==q$, you must first prove $\text{tc_expr} \Delta (\text{Ebinop Oeq} (\text{Etempvar } _p (\text{tptr tint})) (\text{Etempvar } _q (\text{tptr tint})))$, where tc_expr is the type-checking condition for that expression. This simplifies into an entailment with the current precondition on the left, and $\text{denote_tc_comparable } p \ q$ on the right.

The entailer(!) has a solver for such proof goals. It uses the hint database `valid_pointer`. It relies on spatial terms on the l.h.s. of the entailment, such as `data.at $\pi \ t \ v \ p$` which guarantees that p points to something.

The file `progs/verif_ptr_compare.v` illustrates pointer comparisons.

45 *Proof of the reverse program*

Program Logics for Certified Compilers, Chapter 3 shows a program that reverses a linked list (destructively, in place), along with a proof of correctness. (Chapters 2 and 3 available free [here](#).)

That proof is based on a general notion of *list segments*. Here we show a simpler proof that does not use segments, but see [Chapter 45](#) for proof that corresponds to Chapters 3 and 27 of PLCC.

The C program is in `progs/reverse.c`:

```
struct list {unsigned head; struct list *tail;};

struct list *reverse (struct list *p) {
  struct list *w, *t, *v;
  w = NULL;
  v = p;
  while (v) { t = v->tail;  v->tail = w;  w = v;  v = t; }
  return w;
}
```

Please open your CoqIDE or Proof General to `progs/verif_reverse2.v`. As usual, in `progs/verif_reverse2.v` we import the clightgen-produced file `reverse.v` and then build `CompSpecs` and `Vprog` (see [page 13](#), [Chapter 26](#), [Chapter 46](#)).

For the struct list used in this program, we can define the notion of *linked list* $x \overset{\sigma}{\rightsquigarrow} \text{nil}$ with a recursive definition:

```
Fixpoint listrep (sigma: list val) (x: val) : mpred :=
match sigma with
| h::hs => EX y:val, data_at Tsh t.struct_list (h,y) x * listrep hs y
| nil => !! (x = nullval) && emp
end.
```


That is, $\text{listrep } \sigma \ x$ describes a null-terminated linked list starting at pointer p , with permission-share Tsh , representing the sequence σ .

The API spec (see also [Chapter 7](#)) for reverse is,

Definition $\text{reverse_spec} :=$

DECLARE $_reverse$

WITH σ : list val, p : val

PRE [$_p$ OF (tptr t_struct_list)]

PROP() LOCAL(temp $_p \ p$) SEP (listrep $\sigma \ p$)

POST [(tptr t_struct_list)]

EX q :val, PROP() LOCAL(temp $_p \ q$) SEP (listrep (rev σ) q).

The precondition says (for p the function parameter) $p \xrightarrow{\sigma} \text{nil}$, and the postcondition says that (for q the return value) $q \xrightarrow{\text{rev } \sigma} \text{nil}$.

In your IDE, enter the Lemma body $_reverse$ and move after the start_function tactic. As expected, the precondition for the function-body is

PROP() LOCAL(temp $_p \ p$) SEP(listrep $\sigma \ p$).

After forward through two assignment statements ($w=\text{NULL}; v=p;$) the LOCAL part also contains $\text{temp } _v \ p; \text{ temp } _w \ (\text{Vint } (\text{Int.repr } 0))$.

The loop invariant for the while loop is quite similar to the one given in PLCC Chapter 3 page 20:

$$\exists \sigma_1, \sigma_2. \ \sigma = \text{rev}(\sigma_1) \cdot \sigma_2 \ \wedge \ v \xrightarrow{\sigma_2} 0 * w \xrightarrow{\sigma_1} 0$$

It's quite typical for loop invariants to existentially quantify over the values that are different iteration-to-iteration. We represent this in PROP/LOCAL/SEP notation as,

EX σ_1 : list val, EX σ_2 : list val, EX w : val, EX v : val,

PROP($\sigma = \text{rev } \sigma_1 ++ \sigma_2$)

LOCAL(temp $_w \ w; \text{ temp } _v \ v$)

SEP(listrep $\sigma_1 \ w; \text{ listrep } \sigma_2 \ v$).

We apply `forward_while` with this invariant, and (as usual) we have four subgoals: (1) precondition implies loop invariant, (2) loop invariant implies typechecking of loop-termination test, (3) loop body preserves invariant, and (4) after the loop.

(1) To prove the precondition implies the loop invariant, we instantiate σ_1 with `nil` and σ_2 with σ ; we instantiate w with `NULL` and v with p . But this leaves the goal,

```
ENTAIL  $\Delta$ , PROP() LOCAL(temp _v p; temp _w nullval; temp _p p)
  SEP(listrep  $\sigma$  p)
 $\vdash$  PROP( $\sigma = \text{rev } [] ++ \sigma$ ) LOCAL(temp _w nullval; temp _v p)
  SEP(listrep [] nullval; listrep  $\sigma$  p)
```

The PROP and LOCAL parts are trivially solvable by the entailer. We can remove the SEP conjunct `(listrep [] nullval)` by unfolding that occurrence of `listrep`, leaving `!!(nullval=nullval)&&emp`.

(2) The type-checking condition is not trivial, as it is a pointer comparison (see [Chapter 43](#)), but the entailer! solves it anyway.

(3) The loop body starts by assuming the *loop invariant* and the truth of the *loop test*. Their propositional parts have already been moved above the line at the comment `(* loop body preserves invariant *)`. That is, HRE: `isptr v` says that the loop test is true, and H: $\sigma = \text{rev } \sigma_1 ++ \sigma_2$ is from the invariant.

The first statement in the loop body, `t=v→tail`; loads from the list cell at v . But our SEP assertion for v is, `listrep σ_2 v`. The assertion `listrep σ_2 v` is not a `data_at` that we can load from. So we can unfold this occurrence of `listrep`, but *still* there is no `data_at` unless we know that σ_2 is $h :: r$ for some h, r .

We destruct σ_2 leaving two cases: $\sigma_2 = \text{nil}$ and $\sigma_2 = h :: r$. The first case is a contradiction—by the definition of `listrep`, we must have $v == \text{nullptr}$, but that's incompatible with `isptr(v)` above the line.

In the second case, we have (below the line) $\exists y, \dots$ that binds the value of the tail-pointer of the first cons cell. We move that above the line by **Intros** y .

NOW THAT THE FIRST LIST-CELL IS UNFOLDED, it's easy to go forward through the four commands of the loop body. Now we are (* at end of loop body, re-establish invariant *).

We choose values to instantiate the existentials: **Exists** $(h :: \sigma_1, r, v, y)$. (Note that `forward_while` has uncurried the four separate EX quantifiers into a single 4-tuple EX.) Then `entailer!` leaves two subgoals:

-----(1/2)	
rev σ_1 ++ $h :: r = (\text{rev } \sigma_1 \text{ ++ } [h]) \text{ ++ } r$	
----- (2/2)	
listrep σ_1 $w * \text{field_at Tsh t_struct_list [] } (h, w) v * \text{listrep } r y$	
⊢ listrep $(h :: \sigma_1) v * \text{listrep } r y$	

Indeed, `entailer!` always leaves at most two subgoals: at most one propositional goal, and at most one cancellation (spatial) goal. Here, the propositional goal is easily dispatched in the theory of (Coq) lists.

The second subgoal requires unrolling the r.h.s. list segment, by unfolding the appropriate instance of `listrep`. Then we appropriately instantiate some existentials, call on the *entailer!* again, and the goal is solved.

(4) After the loop, we must prove that the loop invariant *and the negation of the loop-test condition* is a sufficient precondition for the next state-ment(s). In this case, the next statement is a return; one can *always* go forward through a return, but now we have to prove that our current assertion implies the function postcondition. This is fairly straightforward.

46 *Alternate proof of reverse*

Chapter 27 of PLCC describes a proof of the same list-reverse program, based on a general theory of list segments. That proof is shown in `progs/verif_reverse.v`.

The general theory is in `progs/list_dt.v`. It accommodates list segments over any C struct type, no matter how many fields. Here, we import the `LsegSpecial` module of that theory, covering the “ordinary” case appropriate for the `reverse.c` program.

Require Import VST.progs.list_dt. **Import** LsegSpecial.

Then we *instantiate* that theory for our particular struct list by providing the `listspec` operator with the *names* of the struct (`_list`) and the link field (`_tail`).

Instance LS: listspec _list _tail.

Proof. `eapply mk_listspec; reflexivity.` **Defined.**

All other fields (in this case, just `_head`) are treated as “data” fields.

Now, `lseg LS π σ p q` is a list segment starting at pointer p , ending at q , with permission-share π and contents σ .

In general, with multiple data fields, the type of σ is constructed via `reptype` (see [Chapter 27](#)). In this example, with one data field, the type of σ computes to `list val`.

47 *Global variables*

In the C language, “extern” global variables live in the same namespace as local variables, but they are shadowed by any same-name local definition. In the C light operational semantics, global variables live in the same namespace as *addressable* local variables (both referenced by the expression-abstract-syntax constructor *Evar*), but in a different namespace from *nonaddressable* locals (expression-abstract-syntax constructor *Etempvar*).¹

In the program-AST produced by *clightgen*, globals (and their initializers) are listed as *Gvars* in the *prog.defs*. These are accessed (automatically) in two ways by the Verifiable C program logic. First, their names and types are gathered into *Vprog* as shown on [page 13](#) (try the Coq command `Print Vprog` to see this list). Second, their initializers are translated into *data_at* conjuncts of separation logic as part of the *main_pre* definition (see [page 35](#)).

When proving *semax_body* for the main function, the *start_function* tactic takes these definitions from *main_pre* and puts them in the precondition of the function body. In some cases this is done using the more-primitive *mapsto* operator², in other cases it uses the higher-level (and more standard) *data_at*.³

¹This difference in namespace treatment cannot matter in a program translated by *CompCert clightgen* from C, because no as-translated expression will exercise the difference.

²For example, examine the proof state in `progs/verif_reverse.v` immediately after *start_function* in Lemma `body_main`; and see the conversion to *data_at* done by the `setup_globals` lemma in that file.

³For example, examine the proof state in `progs/verif_sumarray.v` immediately after *start_function* in Lemma `body_main`.

48 For loops (special case)

MANY FOR-LOOPS HAVE THE FORM, `for (init; i < hi; i++) body` such that the expression `hi` will evaluate to the same value every time around the loop. This upper-bound expression need not be a literal constant, it just needs to be invariant.

For these loops you can use the tactic,

```
forward_for_simple_bound n (EX i:Z, PROP( $\vec{P}$ ) LOCAL( $\vec{Q}$ ) SEP( $\vec{R}$ )).
forward_for_simple_bound n (EX i:Z, EX x:A, PROP( $\vec{P}$ ) LOCAL( $\vec{Q}$ ) SEP( $\vec{R}$ )).
```

where n is the upper bound: a Coq value of type Z such that `hi` will evaluate to n . This tactic generates simpler subgoals than the general `forward_for` tactic.

The loop invariant is $(\text{EX } i:Z, \text{PROP}(\vec{P}) \text{ LOCAL}(\vec{Q}) \text{ SEP}(\vec{R}))$, where i is the value (in each iteration) of the loop iteration variable `i`. You *must* have an existential quantifier for the *value* of the loop-iteration variable. You *may* have a second \exists for a value of your choice that depends on i .

You must omit from Q any mention of the loop iteration variable `i`. The tactic will insert the binding `temp i i`. You need not write `i ≤ hi` in P , the tactic will insert it.

AN EXAMPLE of a for-loop proof is in `progs/verif_sumarray2.v`. This is an alternate implementation of `progs/sumarray.c` (see [Chapter 12](#)) that uses a for loop instead of a while loop:

```
unsigned sumarray(unsigned a[], int n) { /* sumarray2.c */
    int i; unsigned s=0;
    for (i=0; i<n; i++) { s += a[i]; }
    return s;
}
```

Also see `progs/verif_min.v` for several approaches to the specification/verification of another for-loop.

49 For loops (general case)

The C-language for loop has the general form, *for* (*init*; *test*; *incr*) *body*. Let *Inv* be the loop invariant, established by the initializer and preserved by the body-plus-increment. Let *PreInc* be the assertion just before the increment. *Post* is the join-postcondition of the loop; you don't need to provide it if *either* (1) there are no break statements in the loop, or (2) the postcondition is already provided in your proof context (typically because a close-brace follows the entire loop). Depending on whether you need *Post*, verify the loop with,

forward_for Inv PreInc. or *forward_for Inv PreInc Post.*

This is demonstrated in the lemma *body_sumarray_alt* in the file *progs/verif_sumarray2.v*.

```
unsigned sumarray(unsigned a[], int n) {
  int i; unsigned s;
  s=0;
  for (i=0;
      /* Inv : loop invariant */
        i<n; i++) {
    s += a[i];
    /* PreInc : pre-increment invariant */
  }
  /* Post : loop postcondition */
  return s;
}
```

The *Inv* and *PreInc* should have type $A \rightarrow \text{environ} \rightarrow \text{mpred}$, where A is the type of some iteration-dependent quantity (in this example, \mathbb{Z} , to hold the value of loop iteration variable i), and $\text{environ} \rightarrow \text{mpred}$ is the usual type of assertions.

50 Manipulating preconditions

In some cases you cannot go forward until the precondition has a certain form. For example, to go forward through $t=v \rightarrow \text{tail}$; there must be a `data.at` or `field.at` in the SEP clause of the precondition that gives a value for `_tail` field of `t`. As [page 74](#) describes, a listrep can be unfolded to expose such a SEP conjunct.

Faced with the proof goal, $\text{semax } \Delta \text{ (PROP}(\vec{P})\text{LOCAL}(\vec{Q})\text{SEP}(\vec{R})) \text{ } c \text{ } \textit{Post}$ where $\text{PROP}(\vec{P})\text{LOCAL}(\vec{Q})\text{SEP}(\vec{R})$ does not match the requirements for forward symbolic execution, you have several choices:

- Use the rule of consequence explicitly:
apply `semax_pre` **with** $\text{PROP}(\vec{P}')\text{LOCAL}(\vec{Q}')\text{SEP}(\vec{R}')$,
then prove $\text{ENTAIL } \Delta, \vec{P}; \vec{Q}; \vec{R} \vdash \vec{P}'; \vec{Q}'; \vec{R}'$.
- Use the rule of consequence implicitly, by using tactics ([page 80](#)) that modify the precondition.
- Do rewriting in the precondition, either directly by the standard `rewrite` and `change` tactics, or by `normalize` ([page 62](#)).
- Extract propositions and existentials from the precondition, by using `Intros` ([page 40](#)) or `normalize`.
- Flatten stars into semicolons, in the SEP clause, by `Intros`.
- Use the freezer ([page 101](#)) to temporarily “frame away” spatial conjuncts.

TACTICS FOR MANIPULATING PRECONDITIONS. In many of these tactics we select specific conjuncts from the SEP items, that is, the semicolon-separated list of separating conjuncts. These tactic refer to the list by zero-based position number, 0,1,2,...

For example, suppose the goal is a semax or entailment containing $\text{PROP}(\vec{P})\text{LOCAL}(\vec{Q})\text{SEP}(a;b;c;d;e;f;g;h;i;j)$. Then:

focus_SEP $i\ j\ k$. Bring items $\#i,j,k$ to the front of the SEP list.

focus_SEP 5. *results in* $\text{PROP}(\vec{P})\text{LOCAL}(\vec{Q})\text{SEP}(f;a;b;c;d;e;g;h;i;j)$.

focus_SEP 0. *results in* $\text{PROP}(\vec{P})\text{LOCAL}(\vec{Q})\text{SEP}(a;b;c;d;e;f;g;h;i;j)$.

focus_SEP 1 3. *results in* $\text{PROP}(\vec{P})\text{LOCAL}(\vec{Q})\text{SEP}(b;d;a;c;e;f;g;h;i;j)$

focus_SEP 3 1. *results in* $\text{PROP}(\vec{P})\text{LOCAL}(\vec{Q})\text{SEP}(d;b;a;c;e;f;g;h;i;j)$

gather_SEP $i\ j\ k$. Bring items $\#i,j,k$ to the front of the SEP list and conjoin them into a single element.

gather_SEP 5. *results in* $\text{PROP}(\vec{P})\text{LOCAL}(\vec{Q})\text{SEP}(f;a;b;c;d;e;g;h;i;j)$.

gather_SEP 1 3. *results in* $\text{PROP}(\vec{P})\text{LOCAL}(\vec{Q})\text{SEP}(b*d;a;c;e;f;g;h;i;j)$

gather_SEP 3 1. *results in* $\text{PROP}(\vec{P})\text{LOCAL}(\vec{Q})\text{SEP}(d*b;a;c;e;f;g;h;i;j)$

replace_SEP $i\ R$. Replace the i th element the SEP list with the assertion R , and leave a subgoal to prove.

replace_SEP 3 R . *results in* $\text{PROP}(\vec{P})\text{LOCAL}(\vec{Q})\text{SEP}(a;b;c;R;e;f;g;h;i;j)$.

with subgoal $\text{PROP}(\vec{P})\text{LOCAL}(\vec{Q})\text{SEP}(d) \vdash R$.

replace_in_pre $S\ S'$. Replace S with S' anywhere it occurs in the precondition then leave $(\vec{P};\vec{Q};\vec{R}) \vdash (\vec{P};\vec{Q};\vec{R})[S'/S]$ as a subgoal.

frame_SEP $i\ j\ k$. Apply the frame rule, keeping only elements i,j,k of the SEP list. See [Chapter 50](#).

51 The Frame rule

Separation Logic supports the Frame rule,

$$\text{Frame} \frac{\{P\} c \{Q\}}{\{P * F\} c \{Q * F\}}$$

To use this in a forward proof, suppose you have the proof goal,

$\text{semax} \Delta \text{PROP}(\vec{P})\text{LOCAL}(\vec{Q})\text{SEP}(R_0;R_1;R_2) \ c_1;c_2;c_3 \ \text{Post}$

and suppose you want to “frame out” R_2 for the duration of $c_1;c_2$, and have it back again for c_3 . First you rewrite by `seq_assoc` to yield the goal

$\text{semax} \Delta \text{PROP}(\vec{P})\text{LOCAL}(\vec{Q})\text{SEP}(R_0;R_1;R_2) \ (c_1;c_2);c_3 \ \text{Post}$

Then eapply `semax_seq'` to peel off the first command $(c_1;c_2)$ in the new sequence:

$\text{semax} \Delta \text{PROP}(\vec{P})\text{LOCAL}(\vec{Q})\text{SEP}(R_0;R_1;R_2) \ c_1;c_2 \ ?88$

$\text{semax} \Delta' \ ?88 \ c_3 \ \text{Post}$

Then `frame_SEP 0 2` to retain only $R_0;R_2$.

$\text{semax} \Delta \text{PROP}(\vec{P})\text{LOCAL}(\vec{Q})\text{SEP}(R_0;R_2) \ c_1;c_2 \ \dots$

Now you’ll see that (in the precondition of the second subgoal) the unification variable `?88` has been instantiated in such a way that R_2 is added back in.

52 32-bit Integers

The VST program logic uses CompCert's 32-bit integer type.

Inductive comparison := Ceq | Cne | Clt | Cle | Cgt | Cge.

Int.wordsize: nat = 32.

Int.modulus : $Z = 2^{32}$.

Int.max_unsigned : $Z = 2^{32} - 1$.

Int.max_signed : $Z = 2^{31} - 1$.

Int.min_signed : $Z = -2^{31}$.

Int.int : Type.

Int.unsigned : int $\rightarrow Z$.

Int.signed : int $\rightarrow Z$.

Int.repr : $Z \rightarrow \text{int}$.

Int.zero := Int.repr 0.

(* Operators of type int \rightarrow int \rightarrow bool *)

Int.eq Int.lt Int.ltu Int.cmp(c:comparison) Int.cmpu(c:comparison)

(* Operators of type int \rightarrow int *)

Int.neg Int.not

(* Operators of type int \rightarrow int \rightarrow int *)

Int.add Int.sub Int.mul Int.divs Int.mods Int.divu Int.modu

Int.and Int.or Int.xor Int.shl Int.shru Int.shr Int.rol Int.ror Int.rolm

Lemma eq-dec: $\forall (x\ y: \text{int}), \{x = y\} + \{x <> y\}$.

Theorem unsigned_range: $\forall i, 0 \leq \text{unsigned } i < \text{modulus}$.

Theorem unsigned_range_2: $\forall i, 0 \leq \text{unsigned } i \leq \text{max_unsigned}$.

Theorem signed_range: $\forall i, \text{min_signed} \leq \text{signed } i \leq \text{max_signed}$.

Theorem repr_unsigned: $\forall i, \text{repr } (\text{unsigned } i) = i$.

Lemma repr_signed: $\forall i, \text{repr } (\text{signed } i) = i$.

Theorem unsigned_repr:

$\forall z, 0 \leq z \leq \text{max_unsigned} \rightarrow \text{unsigned } (\text{repr } z) = z$.

Theorem signed_repr:

$$\forall z, \text{min_signed} \leq z \leq \text{max_signed} \rightarrow \text{signed} (\text{repr } z) = z.$$

Theorem signed_eq_unsigned:

$$\forall x, \text{unsigned } x \leq \text{max_signed} \rightarrow \text{signed } x = \text{unsigned } x.$$

Theorem unsigned_zero: unsigned zero = 0.

Theorem unsigned_one: unsigned one = 1.

Theorem signed_zero: signed zero = 0.

Theorem eq_sym: $\forall x \ y, \text{eq } x \ y = \text{eq } y \ x.$

Theorem eq_spec: $\forall (x \ y : \text{int}), \text{if } \text{eq } x \ y \text{ then } x = y \text{ else } x <> y.$

Theorem eq_true: $\forall x, \text{eq } x \ x = \text{true}.$

Theorem eq_false: $\forall x \ y, x <> y \rightarrow \text{eq } x \ y = \text{false}.$

Theorem add_unsigned: $\forall x \ y, \text{add } x \ y = \text{repr} (\text{unsigned } x + \text{unsigned } y).$

Theorem add_signed: $\forall x \ y, \text{add } x \ y = \text{repr} (\text{signed } x + \text{signed } y).$

Theorem add_commut: $\forall x \ y, \text{add } x \ y = \text{add } y \ x.$

Theorem add_zero: $\forall x, \text{add } x \ \text{zero} = x.$

Theorem add_zero_l: $\forall x, \text{add } \text{zero } x = x.$

Theorem add_assoc: $\forall x \ y \ z, \text{add} (\text{add } x \ y) \ z = \text{add } x (\text{add } y \ z).$

Theorem neg_repr: $\forall z, \text{neg} (\text{repr } z) = \text{repr } (-z).$

Theorem neg_zero: $\text{neg } \text{zero} = \text{zero}.$

Theorem neg_involutive: $\forall x, \text{neg} (\text{neg } x) = x.$

Theorem neg_add_distr: $\forall x \ y, \text{neg} (\text{add } x \ y) = \text{add} (\text{neg } x) (\text{neg } y).$

Theorem sub_zero_l: $\forall x, \text{sub } x \ \text{zero} = x.$

Theorem sub_zero_r: $\forall x, \text{sub } \text{zero } x = \text{neg } x.$

Theorem sub_add_opp: $\forall x \ y, \text{sub } x \ y = \text{add } x (\text{neg } y).$

Theorem sub_idem: $\forall x, \text{sub } x \ x = \text{zero}.$

Theorem sub_add_l: $\forall x \ y \ z, \text{sub} (\text{add } x \ y) \ z = \text{add} (\text{sub } x \ z) \ y.$

Theorem sub_add_r: $\forall x \ y \ z, \text{sub } x (\text{add } y \ z) = \text{add} (\text{sub } x \ z) (\text{neg } y).$

Theorem sub_shifted: $\forall x \ y \ z, \text{sub} (\text{add } x \ z) (\text{add } y \ z) = \text{sub } x \ y.$

Theorem sub_signed: $\forall x \ y, \text{sub } x \ y = \text{repr} (\text{signed } x - \text{signed } y).$

Theorem `mul_commut`: $\forall x\ y, \text{mul } x\ y = \text{mul } y\ x$.

Theorem `mul_zero`: $\forall x, \text{mul } x\ \text{zero} = \text{zero}$.

Theorem `mul_one`: $\forall x, \text{mul } x\ \text{one} = x$.

Theorem `mul_assoc`: $\forall x\ y\ z, \text{mul } (\text{mul } x\ y)\ z = \text{mul } x\ (\text{mul } y\ z)$.

Theorem `mul_add_distr_l`: $\forall x\ y\ z, \text{mul } (\text{add } x\ y)\ z = \text{add } (\text{mul } x\ z)\ (\text{mul } y\ z)$.

Theorem `mul_signed`: $\forall x\ y, \text{mul } x\ y = \text{repr } (\text{signed } x * \text{signed } y)$.

and many more axioms for the bitwise operators, shift operators, signed/unsigned division and mod operators.

53 CompCert C abstract syntax

The CompCert verified C compiler translates standard C source programs into an abstract syntax for *CompCert C*, and then translates that into abstract syntax for *C light*. Then VST Separation Logic is applied to the C light abstract syntax. C light programs proved correct using the VST separation logic can then be compiled (by CompCert) to assembly language.

C light syntax is defined by these Coq files from CompCert:

Integers. 32-bit (and 8-bit, 16-bit, 64-bit) signed/unsigned integers.

Floats. IEEE floating point numbers.

Values. The val type: integer + float + pointer + undefined.

AST. Generic support for abstract syntax.

Ctypes. C-language types and structure-field-offset computations.

Clight. C-light expressions, statements, and functions.

You will see C light abstract syntax constructors in the Hoare triples (semax) that you are verifying. We summarize the constructors here.

Inductive `expr : Type :=`

<code>(* 1 *)</code>	<code>Econst_int: int → type → expr</code>
<code>(* 1.0 *)</code>	<code>Econst_float: float → type → expr (* double precision *)</code>
<code>(* 1.0f0 *)</code>	<code>Econst_single: float → type → expr (* single precision *)</code>
<code>(* 1L *)</code>	<code>Econst_long: int64 → type → expr</code>
<code>(* x *)</code>	<code>Evar: ident → type → expr</code>
<code>(* x *)</code>	<code>Etempvar: ident → type → expr</code>
<code>(* *e *)</code>	<code>Ederef: expr → type → expr</code>
<code>(* &e *)</code>	<code>Eaddrrof: expr → type → expr</code>
<code>(* ~e *)</code>	<code>Eunop: unary_operation → expr → type → expr</code>
<code>(* e + e *)</code>	<code>Ebinop: binary_operation → expr → expr → type → expr</code>
<code>(* (int)e *)</code>	<code>Ecast: expr → type → expr</code>
<code>(* e.f *)</code>	<code>Efield: expr → ident → type → expr.</code>

Inductive unary_operation := Onotbool | Onotint | Oneg | Oabsfloat.

Inductive binary_operation := Oadd | Osub | Omul | Odiv | Omod
| Oand | Oor | Oxor | Oshl | Oeq | One | Olt | Ogt | Ole | Oge.

Inductive statement : Type :=

$(* \text{ /**/}; *)$	Sskip : statement
$(* E_1 = E_2; *)$	Sassign : $\text{expr} \rightarrow \text{expr} \rightarrow \text{statement}$ (<i>* memory store *</i>)
$(* x = E; *)$	Sset : $\text{ident} \rightarrow \text{expr} \rightarrow \text{statement}$ (<i>* tempvar assign *</i>)
$(* x = f(\dots); *)$	Scall: $\text{option ident} \rightarrow \text{expr} \rightarrow \text{list expr} \rightarrow \text{statement}$
$(* x = b(\dots); *)$	Sbuiltin: $\text{option ident} \rightarrow \text{external_function} \rightarrow \text{typelist} \rightarrow$ $\text{list expr} \rightarrow \text{statement}$
$(* s_1; s_2 *)$	Ssequence : $\text{statement} \rightarrow \text{statement} \rightarrow \text{statement}$
$(* \text{ if() else } \{ \} *)$	Sifthenelse : $\text{expr} \rightarrow \text{statement} \rightarrow \text{statement} \rightarrow \text{statement}$
$(* \text{ for } (;;s_2) s_1 *)$	Sloop: $\text{statement} \rightarrow \text{statement} \rightarrow \text{statement}$
$(* \text{ break}; *)$	Sbreak : statement
$(* \text{ continue}; *)$	Scontinue : statement
$(* \text{ return } E; *)$	Sreturn : $\text{option expr} \rightarrow \text{statement}$
	Sswitch : $\text{expr} \rightarrow \text{labeled_statements} \rightarrow \text{statement}$
	Slabel : $\text{label} \rightarrow \text{statement} \rightarrow \text{statement}$
	Sgoto : $\text{label} \rightarrow \text{statement}$.

54 *C light semantics*

The operational semantics of C light statements and expressions is given in `compcert/cfrontend/Clight.v`. We do not expose these semantics *directly* to the user of Verifiable C. Instead, the *statement* semantics is reformulated as `semax`, an axiomatic (Hoare-logic style) semantics. The *expression* semantics is reformulated in `veric/expr.v` and `veric/Cop2.v` as a *computational*¹ *big-step evaluation semantics*. In each case, a soundness proof relates the Verifiable C semantics to the CompCert Clight semantics.

Rules for `semax` are given in `veric/SeparationLogic.v`—but you rarely use these rules directly. Instead, derived lemmas regarding `semax` are proved in `floyd/*.v` and Floyd’s forward tactic applies them (semi)automatically.

The following functions (from `veric/expr.v`) define expression evaluation:

```
eval_id {CS: compspecs} (id: ident) : environ → val.
    (* evaluate a tempvar *)
eval_var {CS: compspecs} (id: ident) (ty: type) : environ → val.
    (* evaluate an lvar or gvar, addressable local or global variable *)
eval_cast (t t': type) (v: val) : val.
    (* cast value v from type t to type t', but beware! There are
       three types involved, including native type of v. *)
eval_unop (op: unary_operation) (t1 : type) (v1 : val) : val.
eval_binop {CS: compspecs} (op: binary_operation) (t1 t2: type) (v1 v2: val): val.
eval_lvalue {CS: compspecs} (e: expr) : environ → val.
    (* evaluate an l-expression, one that denotes a loadable/storable place *)
eval_expr {CS: compspecs} (e: expr) : environ → val.
    (* evaluate an r-expression, one that is not storable *)
```

The *environ* argument is for looking up the values of local and global variables. However, in most cases where Verifiable C users see `eval_lvalue` or `eval_expr`—in subgoals generated by the forward tactic—all the variables

¹that is, defined by Fixpoint instead of by Inductive.

have already been substituted by values. Thus the environment is not needed.

The expression-evaluation functions call upon several helper functions from `veric/Cop2.v`:

```

sem_cast: type → type → val → option val.
sem_cast_* (* several helper functions for sem_cast *)
bool_val: type → val → option bool.
bool_val_*: (* helper functions *)
sem_notbool: type → val → option val.
sem_neg: type → val → option val.
sem_sub {CS: compspecs}: type → type → val → val → option val.
sem_sub_*: (* helper functions *)
sem_add {CS: compspecs}: type → type → val → val → option val.
sem_add_*: (* helper functions *)
sem_mul: type → type → val → val → option val.
sem_div: type → type → val → val → option val.
sem_mod: type → type → val → val → option val.
sem_and: type → type → val → val → option val.
sem_or: type → type → val → val → option val.
sem_xor: type → type → val → val → option val.
sem_shl: type → type → val → val → option val.
sem_shr: type → type → val → val → option val.
sem_cmp: comparison → type → type → (...) → val → val → option val.
sem_unary_operation: unary_operation → type → val → option val.
sem_binary_operation {CS: compspecs}:
  binary_operation → type → type → mem → val → val → option val.

```

The details are not so important to remember. The main point is that Coq expressions of the form `sem...` *should* simplify away, provided that their arguments are instantiated with concrete operators, concrete constructors `Vint/Vptr/Vfloat`, and concrete C types. The `int` values (etc.) carried inside `Vint/Vptr/Vfloat` *do not* need to be concrete: they can be Coq variables. This is the essence of proof by symbolic execution.

55 Splitting arrays

Consider this example, from the main function of `progs/verif_sumarray2.v`:

`data_at sh (tarray tint k) al p : mpred`

The `data_at` predicate here says that in memory starting at address p there is an array of k slots containing, respectively, the elements of the sequence al .

Suppose we have a function `sumarray(unsigned a[], int n)` that takes an array of length n , and we apply it to a “slice” of p : `sumarray(p+i,k-i)`; where $0 \leq i \leq k$. The precondition of the `sumarray` funspec has `data_at sh (tarray tint n)`. In this case, we would like $a = \&(p[i])$, $n = k - i$, and bl = the sublist of al from i to $k - 1$.

To prove this function-call by `forward_call`, we must split up `(data_at sh (tarray tint k) al p)` into two conjuncts:

`(data_at sh (tarray tint i) (sublist 0 i al) p *
data_at sh (tarray tint (k - i)) (sublist i k al) q),`

where q is the pointer to the array slice beginning at address $p + i$. We write this as, $q = \text{field_address0 (tarray tint k) [ArraySubsc i] p}$. That is, given a pointer p to a data structure described by `(tarray tint k)`, calculate the *address* for subscripting the i th element. (See [Chapter 35](#))

As shown in the `body_main` proof in `progs/verif_sumarray2.v`, the lemma `split_array` proves the equivalence of these two predicates, using the VST-Floyd lemma `split2_data_at_Tarray`. Then the `data_at ... q` predicate can satisfy the precondition of `sumarray`, while the p slice will be part of the “frame” for the function call.

See also: `split3_data_at_Tarray`.

56 *sublist*

Chapter 54 explained that we often need to reason about slices of arrays whose contents are sublists of lists. For that we have a function `sublist i j l` which makes a new list out of the elements $i \dots j - 1$ of list l .

To simplify expressions involving, `sublist`, `++`, `map`, `Zlength`, `Znth`, and `list_repeat`, use **autorewrite with sublist**.

Often, you find equations “above the line” of the form,

H: $n = \text{Zlength } (\text{map } \text{Vint } (\text{map } \text{Int.repr } \text{contents}))$

You may find it useful to do **autorewrite with sublist in * \vdash** to change this to $n = \text{Zlength } \text{contents}$ before proceeding with **(autorewrite with sublist)** below the line.

These rules comprise the *sublist rewrite database*:

`sublist_nil'`: $i = j \rightarrow \text{sublist } i \ j \ l = []$.

`app_nil_l`: $[] ++ l = l$.

`app_nil_r`: $l ++ [] = l$.

`Zlength_rev`: $\text{Zlength } (\text{rev } l) = \text{Zlength } l$.

`sublist_rejoin'`: $0 \leq i \leq j = j' \leq k \leq \text{Zlength } l \rightarrow$

$\text{sublist } i \ j \ l ++ \text{sublist } j' \ k \ l = \text{sublist } i \ k \ l$.

`subsub1`: $a - (a - b) = b$.

`Znth_list_repeat_inrange`: $0 \leq i \leq n \rightarrow \text{Znth } i \ (\text{list_repeat } (\text{Z.to_nat } n) \ a) \ d = a$.

`Zlength_cons`: $\text{Zlength } (a :: l) = \text{Z.succ } (\text{Zlength } l)$.

`Zlength_nil`: $\text{Zlength } [] = 0$.

`Zlength_app`: $\text{Zlength } (l ++ l') = \text{Zlength } l ++ \text{Zlength } l'$.

`Zlength_map`: $\text{Zlength } (\text{map } f \ l) = \text{Zlength } l$.

`list_repeat_0`: $\text{list_repeat } (\text{Z.to_nat } 0) = []$.

`Zlength_list_repeat`: $0 \leq n \rightarrow \text{Zlength } (\text{list_repeat } (\text{Z.to_nat } n)) = n$.

`Zlength_sublist`: $0 \leq i \leq j \leq \text{Zlength } l \rightarrow \text{Zlength } (\text{sublist } i \ j \ l) = j - i$.

`sublist_sublist`: $0 \leq m \rightarrow 0 \leq k \leq i \leq j - m \rightarrow$

$\text{sublist } k \ i \ (\text{sublist } m \ j \ l) = \text{sublist } (k + m) \ (i + m) \ l$.

`sublist_app1`: $0 \leq i \leq j \leq \text{Zlength } l \rightarrow \text{sublist } i \ j \ (l ++ l') = \text{sublist } i \ j \ l$.

sublist_app2: $0 \leq \text{Zlength } l \leq i \rightarrow$

$\text{sublist } i \ j \ (l ++ l') = \text{sublist } (i - \text{Zlength } l) \ (j - \text{Zlength } l) \ l'.$

sublist_list_repeat: $0 \leq i \leq j \leq k \rightarrow$

$\text{sublist } i \ j \ (\text{list_repeat } (\text{Z.to_nat } k) \ v) = \text{list_repeat } (\text{Z.to_nat } (j - i)) \ v.$

sublist_same: $i = 0 \rightarrow j = \text{Zlength } l \rightarrow \text{sublist } i \ j \ l = l.$

app_Znth1: $i < \text{Zlength } l \rightarrow \text{Znth } i \ (l ++ l') \ d = \text{Znth } i \ l \ d.$

app_Znth2: $i \geq \text{Zlength } l \rightarrow \text{Znth } i \ (l ++ l') \ d = \text{Znth } i - \text{Zlength } l \ l' \ d.$

Znth_sublist: $0 \leq i \rightarrow 0 \leq j < k - i \rightarrow \text{Znth } j \ (\text{sublist } i \ k \ l) \ d = \text{Znth } (j + i) \ l \ d.$

along with miscellaneous \mathbb{Z} arithmetic:

$$n - 0 = n \quad 0 + n = n \quad n + 0 = n \quad n \leq m \rightarrow \max(n, m) = m$$

$$n + m - n = m \quad n + m - m = n \quad m - n + n = m \quad n - n = 0$$

$$n + m - (n + p) = m - p \quad \text{etcetera.}$$

57 Later

Many of the Hoare rules (e.g., see PLCC, page 161) have the operator \triangleright (pronounced “later”) in their precondition:

$$\text{semax_set_forward} \frac{}{\Delta \vdash \{\triangleright P\} \ x := e \ \{\exists v. x = (e[v/x]) \wedge P[v/x]\}}$$

The modal assertion $\triangleright P$ is a slightly weaker version of the assertion P . It is used for reasoning by induction over how many steps left we intend to run the program. The most important thing to know about \triangleright later is that P is stronger than $\triangleright P$, that is, $P \vdash \triangleright P$; and that operators such as $*$, $\&\&$, ALL (and so on) commute with later: $\triangleright(P * Q) = (\triangleright P) * (\triangleright Q)$.

This means that if we are trying to apply a rule such as `semax_set_forward`; and if we have a precondition such as

`local (tc_expr Δ e) $\&\&$ \triangleright local (tc_temp_id id t Δ e) $\&\&$ ($P_1 * \triangleright P_2$)`

then we can use the rule of consequence to *weaken* this precondition to

`\triangleright (local (tc_expr Δ e) $\&\&$ local (tc_temp_id id t Δ e) $\&\&$ ($P_1 * P_2$))`

and then apply `semax_set_forward`. We do the same for many other kinds of command rules.

This weakening of the precondition is done automatically by the forward tactic, as long as there is only one \triangleright later in a row at any point among the various conjuncts of the precondition.

A more sophisticated understanding of \triangleright is needed to build proof rules for recursive data types and for some kinds of object-oriented programming; see PLCC Chapter 19.

58 Mapsto and func_ptr

Aside from the standard operators and axioms of separation logic, the core separation logic has just two primitive spatial predicates:

Parameter `address_mapsto`:

`memory_chunk` \rightarrow `val` \rightarrow `share` \rightarrow `share` \rightarrow `address` \rightarrow `mpred`.

Parameter `func_ptr` : `funspec` \rightarrow `val` \rightarrow `mpred`.

`func_ptr` ϕ `v` means that value v is a pointer to a function with specification ϕ ; see [Chapter 61](#).

`address_mapsto` expresses what is typically written $x \mapsto y$ in separation logic, that is, a singleton heap containing just value y at address x .

From this, we construct two low-level derived forms:

`mapsto (sh:share) (t:type) (v w: val) : mpred` describes a singleton heap with just one value w of (C-language) type t at address v , with permission-share sh .

`mapsto_ (sh:share) (t:type) (v:val) : mpred` describes an *uninitialized* singleton heap with space to hold a value of type t at address v , with permission-share sh .

From these primitives, `field_at` and `data_at` are constructed.

59 *with_library*: Library functions

A CompCert C program is implicitly linked with dozens of “built-in” and library functions. In the .v file produced by clightgen, the prog-defs component of your prog lists these as External definitions, along with the Internal definitions of your own functions. *Every one of these needs exactly one funspec*, of the form DECLARE...WITH..., and this funspec must be *proved* with a semax_ext proof.

Fortunately, if your program does not use a given library function f , then the funspec DECLARE _f WITH...PRE[...] False POST... with a **False** precondition is easy to prove! The tactic with_library prog [$s_1; s_2; \dots; s_n$] augments your explicit funspec-list [$s_1; s_2; \dots; s_n$] with such trivial funspecs for the other functions in the program prog.

Definition Gprog := ltac:(with_library prog [sumarray_spec; main_spec]).

YOU MAY WISH to use standard library functions such as malloc, free, exit. These are axiomatized (with external funspecs) in floyd.library. To use them, **Require Import** VST.floyd.library *after* you import floyd.proofauto. This imports a (floyd.library.)with_library tactic hiding the standard (floyd.forward.)with_library tactic; the new one includes *axiomatized* specifications for malloc, free, exit, etc. We haven’t proved the implementations against the axioms, so if you don’t trust them, then don’t import floyd.library.

The next chapters explain the specifications of certain standard-library functions.

60 *malloc/free*

The C library's `malloc` and `free` functions have these specifications:

```

DECLARE _malloc
  WITH n:Z
  PRE [ 1%positive OF tuint ] (* parameter 1 has type unsigned int *)
    PROP(0 ≤ n ≤ Int.max_unsigned)
    LOCAL(temp 1%positive (Vint (Int.repr n)))
    SEP()
  POST [ tptr tvoid ] EX p:_,
    PROP()
    LOCAL(temp ret_temp p)
    SEP(if eq.dec p nullval then emp
        else (malloc_token Tsh n p * memory_block Tsh n p)).

DECLARE _free
  WITH p:val, n:Z
  PRE [ 1%positive OF tptr tvoid ]
    PROP()
    LOCAL(temp 1%positive p)
    SEP(malloc_token Tsh n p; memory_block Tsh n p)
  POST [ Tvoid ]
    PROP()
    LOCAL()
    SEP().

```

You must **Import** `VST.floyd.library`. Then these funspecs are made available in your Gprog by the use of the `with_library` tactic ([Chapter 58](#)).

The purpose of the `malloc_token` is to describe the special record-descriptor that tells `free` how big the allocated record was.

See `progs/verif_queue.v` for a demonstration of `malloc/free`.

61 *exit*

Import VST.floyd.library. before you define
Gprog := Itac:(with_library prog [...]).
and you will get:

```
DECLARE _exit  
  WITH u: unit  
  PRE [1%positive OF tint]  
    PROP() LOCAL() SEP()  
  POST [ tvoid ]  
    PROP(False) LOCAL() SEP().
```

62 *Function pointers*

Parameter $\text{func_ptr} : \text{funspec} \rightarrow \text{val} \rightarrow \text{mpred}$.

Definition $\text{func_ptr}' f v := \text{func_ptr } f v \ \&\& \ \text{emp}$.

$\text{func_ptr } \phi v$ means that v is a pointer to a function with funspec ϕ .
 $\text{func_ptr}' \phi v$ is a form more suitable to be a conjunct of a SEP clause.

Verifiable C's program logic is powerful enough to reason expressively about function pointers (see PLCC Chapters 24 and 29). Object-oriented programming with function pointers is illustrated, in two different styles, by the programs `progs/message.c` and `progs/object.c`, and their verifications, `progs/verif_message.c` and `progs/verif_object.c`.

In this chapter, we illustrate using the much simpler program, `progs/funcptr.c`.

```
int myfunc (int i) { return i+1; }
void *a[] = {myfunc};
int main (void) {
  int (*f)(int);
  int j;
  f = &myfunc;
  j = f(3);
  return j;
}
```

The verification, in `progs/verif_funcptr.v`, defines

Definition $\text{myfunc_spec} := \text{DECLARE } _myfunc \ \text{myspec}$.

where myspec is a Definition for a WITH...PRE...POST specification.

Near the beginning of **Lemma** `body_main`, notice that we have `LOCAL(gvar _myfunc p)` in the precondition. That `gvar` is needed by the tactic `make_func_ptr _myfunc`, which adds $\text{func_ptr}' \ \text{myspec } p$ to the

SEP clause. It “knows” to use `myspec` because it looks up `_myfunc` in Delta (which, in turn, is derived from Gprog).

Now, forward through the assignment `f=myfunc` works as you might expect, adding the LOCAL clause `temp _f p`.

To call a function-variable, such as this program’s `j=f(3)`; just use `forward.call` as usual. However, in such a case, `forward.call` will find its `funspec` in a `func_ptr`’ SEP-clause, rather than as a global entry in Delta as for ordinary function calls.

Note: Unfortunately, in order to get the gvar `_myfunc` into the precondition of `main`, there must be some initialized global variable that refers to `myfunc`. That’s the purpose of the (otherwise useless) array `a` in this program. And suppose you wanted to do `make_func_ptr` in some function *other* than `main`. Then you’d need to add this gvar to the LOCAL clause of that function’s precondition, and pass it down from `main`. Both of these infelicities ought to be remedied in a future release.

63 *Axioms of separation logic*

These axioms of separation logic are often useful, although generally it is the automation tactics (entailer, cancel) that apply them.

```

pred_ext: P ⊢ Q → Q ⊢ P → P = Q.
derives_refl: P ⊢ P.
derives_trans: P ⊢ Q → Q ⊢ R → P ⊢ R.
andp_right: X ⊢ P → X ⊢ Q → X ⊢ (P && Q).
andp_left1: P ⊢ R → P && Q ⊢ R.
andp_left2: Q ⊢ R → P && Q ⊢ R.
orp_left: P ⊢ R → Q ⊢ R → P || Q ⊢ R.
orp_right1: P ⊢ Q → P ⊢ Q || R.
orp_right2: P ⊢ R → P ⊢ Q || R.
exp_right: ∀ {B: Type} (x:B) (P: mpred) (Q: B → mpred),
    P ⊢ Q x → P ⊢ EX x:B, Q.
exp_left: ∀ {B: Type} (P:B → mpred) (Q: mpred),
    (∀ x, P x ⊢ Q) → EX x:B, P ⊢ Q.
allp_left: ∀ {B} (P: B → mpred) x Q, P x ⊢ Q → ALL x:B, P ⊢ Q.
allp_right: ∀ {B} (P: mpred) (Q:B → mpred),
    (∀ v, P ⊢ Q v) → P ⊢ ALL x:B, Q.
prop_left: ∀ (P: Prop) Q, (P → (TT ⊢ Q)) → !!P ⊢ Q.
prop_right: ∀ (P: Prop) Q, P → (Q ⊢ !!P).
not_prop_right: ∀ (P: mpred) (Q: Prop), (Q → (P ⊢ FF)) → P ⊢ !(~Q).

sepcon_assoc: (P * Q) * R = P * (Q * R).
sepcon_comm: P Q, P * Q = Q * P.
sepcon_andp_prop: P * (!!Q && R) = !!Q && (P * R).
derives_extract_prop: (P → Q ⊢ R) → !!P && Q ⊢ R.
sepcon_derives: P ⊢ P' → Q ⊢ Q' → P * Q ⊢ P' * Q'.

```

64 *Obscure higher-order axioms*

The wand \multimap operator is “magic wand,” $\text{ewand } \multimap$ is “existential magic wand,” and \triangleright is pronounced “later” and written $|>$ in Coq.

see PLCC, Chapter 19.

$\text{imp_andp_adjoint: } P \&\& Q \vdash R \leftrightarrow P \vdash (Q \longrightarrow R).$
 $\text{wand_sepcon_adjoint: } P * Q \vdash R \leftrightarrow P \vdash Q \multimap R.$
 $\text{ewand_sepcon: } (P * Q) \multimap R = P \multimap (Q \multimap R).$
 $\text{ewand_TT_sepcon: } \forall (P \ Q \ R: A),$
 $\quad (P * Q) \&\& (R \multimap \text{TT}) \vdash (P \&\& (R \multimap \text{TT})) * (Q \&\& (R \multimap \text{TT})).$
 $\text{exclude_elsewhere: } P * Q \vdash (P \&\& (Q \multimap \text{TT})) * Q.$
 $\text{ewand_conflict: } P * Q \vdash \text{FF} \rightarrow P \&\& (Q \multimap R) \vdash \text{FF}$

 $\text{now_later: } P \vdash \triangleright P.$
 $\text{later_K: } \triangleright (P \longrightarrow Q) \vdash (\triangleright P \longrightarrow \triangleright Q).$
 $\text{later_allp: } \forall T \ (F: T \rightarrow \text{mpred}), \triangleright (\text{ALL } x:T, F \ x) = \text{ALL } x:T, \triangleright (F \ x).$
 $\text{later_exp: } \forall T \ (F: T \rightarrow \text{mpred}), \text{EX } x:T, \triangleright (F \ x) \vdash \triangleright (\text{EX } x: F \ x).$
 $\text{later_exp': } \forall T \ (\text{any}:T) \ F, \triangleright (\text{EX } x: F \ x) = \text{EX } x:T, \triangleright (F \ x).$
 $\text{later_imp: } \triangleright (P \longrightarrow Q) = (\triangleright P \longrightarrow \triangleright Q).$
 $\text{loeb: } \triangleright P \vdash P \rightarrow \text{TT} \vdash P.$
 $\text{later_sepcon: } \triangleright (P * Q) = \triangleright P * \triangleright Q.$
 $\text{later_wand: } \triangleright (P \multimap Q) = \triangleright P \multimap \triangleright Q.$
 $\text{later_ewand: } \triangleright (P \multimap Q) = (\triangleright P) \multimap (\triangleright Q).$

65 Proving larg(ish) programs

When your program is not all in one .c file, see also [Chapter 65](#). Whether or not your program is all in one .c file, you can prove the individual function bodies in separate .v files. This uses less memory, and (on a multicore computer with parallel make) saves time. To do this, put your API spec (up to the construction of Gprog in one file; then each `semax_body` proof in a separate file that imports the API spec.

EXTRACTION OF SUBORDINATE SEMAX-GOALS. To ease memory pressure and recompilation time, it is often advisable to partition the proof of a function into several lemmas. Any proof state whose goal is a `semax`-term can be extracted as a stand-alone statement by invoking tactic `semax_subcommand V G F`. The three arguments are as in the statement of surrounding `semax-body` lemma, i.e. are of type *varspecs*, *funspecs*, and *function*.

The subordinate tactic `mkConciseDelta V G F Δ` can also be invoked individually, to concisely display the type context Δ as the application of a sequence of initializations to the host function's `func_tycontext`.

THE FREEZER. A distinguishing feature of separation logic is the frame rule, i.e. the ability to modularly verify a statement w.r.t. its minimal resource footprint. Unfortunately, being phrased in terms of the syntactic program structure, the standard frame rule does not easily interact with forward symbolic execution as implemented by the Floyd tactics (and many other systems), as these continuously rearrange the associativity of statement sequencing to peel off the redex of the next *forward*, and (purposely) hide the program continuation as the abbreviation `MORE_COMMANDS`.

Resolving this conflict, Floyd's *freezer* abstraction provides a means for flexible framing, by implementing a veil that opaquely hides selected items of a SEP clause from non-symbolic treatment by non-freezer tactics.

The freezer abstraction consists of two main tactics, *freeze* $N\ F$ and *thaw* F , where $N : \text{list nat}$ and F is a user-supplied (fresh) Coq name. The result of applying *freeze* $[i_1; \dots; i_n]\ F$ to a semax goal is to remove items i_1, \dots, i_n from the precondition's SEP clause, inserting the item $FRZL\ F$ at the head of the SEP list, and adding a hypothesis $F := \text{abbreviate}$ to Coq's proof context.

The term $FRZL\ F$ participates symbolically in all non-freezer tactics just like any other SEP item, so can in particular be canceled, and included in a function call's frame. Unfolding a freezer is not tied to the associativity structure of program statements but can be achieved by invoking *thaw* F , which simply replaces $FRZL\ F$ by the the list of F 's constituents. As multiple freezers can coexists and freezers can be arbitrarily nested, SEP-clauses R effectively contain forests of freezers, each constituent being thawable independently and freezer-level by freezer-level.

Wrapping single *forward* or *forward_call* commands in a freezer often speeds up the processing time noticably, as invocations of subordinate tactics *entailer*, *cancel*, etc. are supplied with smaller and more symbolic proof goals. In our experience, applying the freezer throughout the proof of an entire function body typically yields a speedup of about 30% on average with improvements of up to 55% in some cases, while also easing the memory pressure and freeing up valuable real estate on the user's screen.

A more invasive implementation of a freezer-like abstraction would refine the $\text{PROP}(P)\ \text{LOCAL}(Q)\ \text{SEP}(R)$ structure to terms of the form $\text{PROP}(P)\ \text{LOCAL}(Q)\ \text{SEP}(R)\ \text{FR}(H)$ where $H : \text{list mpred}$. Again, terms in H would be treated opaquely by all tactics, and freezing/thawing would correspond to transfer rules between R and H . In either case, forward symbolic execution is reconciled with the frame rule, and the use of the mechanism is sound engineering practice as documentation of programmer's insight is combined with performance improvements.

66 *Separate compilation*, `semax_ext`

What to do when your program is spread over multiple `.c` files. See `progs/even.c` and `progs/odd.c` for an example.

CompCert's `clightgen` tool translates your `.c` file into a `.v` file in which each C-language identifier is assigned a positive number in the AST (Abstract Syntax Tree) representation. When you have several `.c` files, you need consistent numbering of the identifiers in the `.v` files. One way to achieve this is to run `clightgen` on all the `.c` files at once:

```
clightgen even.c odd.c
```

This generates `even.v` and `odd.v` with consistent names. (It's not exactly separate compilation, but it will have to suffice for now.)

Now, you can do *modular verification of modular programs*. This is illustrated in,

`progs/verif_evenodd_spec.v` Specifications of the functions.

`progs/verif_even.v` Verification of `even.c`.

`progs/verif_odd.v` Verification of `odd.c`.

Linking of the final proofs is described by Stewart.¹

¹Gordon Stewart, *Verified Separate Compilation for C*, PhD Thesis, Department of Computer Science, Princeton University, April 2015

67 *Catalog of tactics / lemmas*

Below is an alphabetic catalog of the major floyd tactics. In addition to short descriptions, the entries indicate whether a tactic (or tactic notation) is typically user-applied [u], primarily of internal use [i] or is expected to be used at development-time but unlikely to appear in a finished proof script [d]. We also mention major interdependencies between tactics, and their points of definition.

- assert_PROP** P (tactic; [Chapter 40](#)) Put the proposition P above the line, if it is provable from the current precondition.
- cancel** (tactic; [page 60](#)) Deletes identical spatial conjuncts from both sides of a base-level entailment.
- data_at_conflict** p (tactic) equivalent to `field_at_conflict p nil`.
- deadvars!** (tactic) Removes from the LOCAL block of the current precondition, any variables that are irrelevant to the rest of program execution. Fails if there is no such variable.
- derives_refl** (lemma) $A \vdash A$. Useful after `cancel` to handle $\beta\eta$ -equality; see [page 60](#).
- derives_refl'** (lemma) $A = B \rightarrow A \vdash B$.
- drop_LOCAL** n (tactic, where $n : nat$). Removes the n th entry of a the LOCAL block of a `semax` or `ENTAIL` precondition.
- drop_LOCALs** $[_i; _j]$ Removes variables $_i$ and $_j$ from the LOCAL block of a `semax` or `ENTAIL` precondition.
- entailer** (tactic; [page 61](#), [page 28](#)) Proves (lifted or base-level) entailments, possibly leaving a residue for the user to prove.
- entailer!** (tactic; [page 61](#), [page 28](#)) Like `entailer`, but faster and more powerful—however, it sometimes turns a provable goal into an unprovable goal.
- Exists** v (tactic; [Chapter 22](#)) Instantiate an EX existential on the right-hand side of an entailment.
- field_at_conflict** $p fld$ (tactic) Solves an entailment of the form $\dots * \text{field_at } sh \ t \ fld \ v_1 \ p * \dots * \text{field_at } sh \ t \ fld \ v_2 \ p * \dots \vdash _$ based on the contradiction that the same field-assertion cannot `*-separate`. Usually invoked automatically by `entailer` (or `entailer!`)

to prove goals such as $!!(p < q)$. Needs to be able to prove (or compute) the fact that $0 < \text{sizeof}(\text{nested_field_type } t \text{ fld})$;
 for `data_at_conflict` that's equivalent to $0 < \text{sizeof } t$.

forward (tactic; [page 21](#)) Do forward Hoare-logic proof through one C statement (assignment, break, continue, return).

forward_call *ARGS* (tactic; [page 37](#)) Forward Hoare-logic proof through one C function-call, where *ARGS* is a witness for the WITH clause of the funspec.

forward_for (tactic; [page 78](#)) Hoare-logic proof for the for statement, general case.

forward_for_simple_bound *n Inv* (tactic; [page 77](#)) When a for-loop has the form `for (init; i < hi; i++)`, where *n* is the *value* of *hi*, and *Inv* is the loop invariant.

forward_if *Q* (tactic; [page 24](#)) Hoare-logic proof for the if statement, where *Q* may be omitted if at the end of a block, where the postcondition is already given.

forward_while *Inv* (tactic; [Chapter 12](#)) Forward Hoare-logic proof of a while loop, with loop invariant *Inv*.

make_compspecs *prog* (tactic; [page 13](#))

mk_varspecs *prog* (tactic; [page 13](#))

mkConciseDelta *V G F Δ* (tactic; [page 101](#)) Applicable to a proof state with a semax goal. Simplifies the Δ component to the application of a sequence of initializations to the host function's `func_tycontext`. Used to prepare the current proof goal for abstracting/factoring out as a separate lemma.

name *i _i* (tactic) Before `start_function`, suggest the name *i* for the Coq variable associated with the value of C global variable `_i`.

semax_subcommand *V G F* (tactic) Applicable to a proof state with a semax goal. Extracts the current proof state as a stand-alone statement that can be copy-and-pasted to a separate file. The three arguments should be copied from the statement of surrounding semax-body lemma: *V* : varspecs, *G* : funspecs, *F* : function.

start_function (tactic; [Chapter 9](#)) Unpack the funspec's pre- and post-condition into a Hoare triple describing the function body.

sublist_split (lemma; [page 33](#)) Break a sublist into the concatenation of two smaller sublists.

unfold_data_at (tactic; [page 50](#)) When t is a struct (or array) type, break apart `data_at sh t v p` into a separating conjunction of its individual fields (or array elements).

unfold_field_at (tactic; [page 50](#)) Like `unfold_data_at`, but starts with `field_at sh t path v p`.

with_library (tactic; [Chapter 58](#)) Complete the funspecs by inserting stub specifications for all unspecified functions; and (if `Import VST.floyd.library` is done) adding standard specifications for `malloc`, `free`, `exit`.