

Coccinelle

User's manual

release 0.2.3

Julia Lawall and Yoann Padioleau

(with contributions from Rene Rydhof Hansen, Nicolas Palix, Henrik Stuart)

November 14, 2010

Contents

I	User Manual	3
1	Introduction	4
2	Installing Coccinelle	5
2.1	Requirements	5
2.2	Getting Coccinelle	5
2.3	Compiling Coccinelle	5
2.4	Running Coccinelle	5
3	Tutorial	6
4	Examples	7
4.1	Examples	7
4.1.1	Function renaming	7
4.1.2	Removing a function argument	8
4.1.3	Introduction of a macro	9
4.1.4	Look for NULL dereference	11
4.1.5	Reference counter: the of_xxx API	12
4.1.6	Filtering identifiers, declarers or iterators with regular expression	14
4.2	Tips and Tricks	14
4.2.1	How to remove useless parentheses?	14
5	Isomorphisms and <code>standard.iso</code>	16
6	Parsing C, <code>cpp</code>, and <code>standard.h</code>	17
7	Developing a Semantic Patch	18
8	Advanced Features	19
II	Reference Manual	20
9	SmPL grammar	21
9.1	Program	21
9.2	Metavariables for transformations	21
9.3	Metavariables for scripts	24
9.4	Transformation	25
9.5	Types	27
9.6	Function declarations	28
9.7	Declarations	29

9.8	Statements	29
9.9	Expressions	30
9.10	Constant, Identifiers and Types for Transformations	31
10	spatch command line options	32
10.1	Introduction	32
10.2	Selecting and parsing the semantic patch	33
10.2.1	Standalone options	33
10.2.2	The semantic patch	33
10.2.3	Isomorphisms	33
10.2.4	Display options	33
10.3	Selecting and parsing the C files	34
10.3.1	Standalone options	34
10.3.2	Selecting C files	34
10.3.3	Parsing C files	35
10.4	Application of the semantic patch to the C code	37
10.4.1	Feedback at the rule level during the application of the semantic patch	37
10.4.2	Feedback at the CTL level during the application of the semantic patch	37
10.4.3	Actions during the application of the semantic patch	38
10.5	Generation of the result	38
10.6	Other options	39
10.6.1	Version information	39
10.6.2	Help	39
10.6.3	Controlling the execution of Coccinelle	39
10.6.4	Miscellaneous	40
III	Appendix	41

Foreword

This manual documents the release 0.2.3 of Coccinelle. It is organized as follows:

- Part I is an introduction to Coccinelle
- Part II is the reference description of Coccinelle, its language and command line tool.

Conventions

Copyright

Coccinelle copyright is

© 2010, University of Copenhagen DIKU and INRIA.

© 2005-2009, University of Copenhagen DIKU and Ecole des Mines de Nantes.

Coccinelle is open source and can be freely redistributed under the terms of the GNU General Public License version 2. See the file `license.txt` in the distribution for licensing information.

Copyright © 2010, Nicolas Palix, Julia Lawall, and Gilles Muller

Copyright © 2008, 2009, Yoann Padioleau, Nicolas Palix, Julia Lawall, and Gilles Muller

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

Availability

Coccinelle can be freely downloaded from <http://coccinelle.lip6.fr>.

This website contains also additional information and a wiki website.

Part I

User Manual

Chapter 1

Introduction

Coccinelle is a tool to help automate repetitive source-to-source style-preserving program transformations on C source code, like for instance to perform some refactorings. Coccinelle is presented as a command line tool called `spatch` that takes as input the name of a file containing the specification of a program transformation, called a *semantic patch*, and a set of C files, and then performs the transformation on all those C files.

To make it easy to express those transformations, Coccinelle proposes a WYSISWYG approach where the C programmer can leverage the things he already knows: the C syntax and the patch syntax. Indeed, with Coccinelle transformations are written in specific language called SmPL, for Semantic Patch Language, which as the name suggests is very close to the syntax of a patch, but which does not work at a line level, than traditional patches do. but a more high level, or semantic level.

Here is an example of a simple program transformation. To replace every calls to `foo` of any expression x to a call to `bar`, create a semantic patch file `ex1.cocci` (semantic patches usually ends with the `.cocci` filename extension) containing:

```
@@ expression x; @@  
  
- foo(x)  
+ bar(x)
```

Then to “apply” the specified program transformation to a set of C files, simply do:

```
$ spatch -sp_file ex1.cocci *.c
```

Coccinelle primarily targets ANSI C, and supports some GCC extensions. It has only partial support for K&R C. K&R function declarations are only recognized if the parameter declarations are indented. Furthermore, the parameter names are subsequently considered to be type names, due to confusion with function prototypes, in which a name by itself is indeed the name of a type.

Chapter 2

Installing Coccinelle

2.1 Requirements

2.2 Getting Coccinelle

2.3 Compiling Coccinelle

2.4 Running Coccinelle

Chapter 3

Tutorial

Chapter 4

Examples

4.1 Examples

This section presents a range of examples. Each example is presented along with some C code to which it is applied. The description explains the rules and the matching process.

4.1.1 Function renaming

One of the primary goals of Coccinelle is to perform software evolution. For instance, Coccinelle could be used to perform function renaming. In the following example, every occurrence of a call to the function `foo` is replaced by a call to the function `bar`.

Before	Semantic patch	After
<pre>1 #DEFINE TEST "foo" 2 3 printf("foo"); 4 5 int main(int i) { 6 //Test 7 int k = foo(); 8 9 if(1) { 10 foo(); 11 } else { 12 foo(); 13 } 14 15 foo(); 16 }</pre>	<pre>1 @@ 2 3 @@ 4 5 6 - foo() 7 + bar()</pre>	<pre>1 #DEFINE TEST "foo" 2 3 printf("foo"); 4 5 int main(int i) { 6 //Test 7 int k = bar(); 8 9 if(1) { 10 bar(); 11 } else { 12 bar(); 13 } 14 15 bar(); 16 }</pre>

4.1.2 Removing a function argument

Another important kind of evolution is the introduction or deletion of a function argument. In the following example, the rule `rule1` looks for definitions of functions having return type `irqreturn_t` and two parameters. A second *anonymous* rule then looks for calls to the previously matched functions that have three arguments. The third argument is then removed to correspond to the new function prototype.

```
1 @ rule1 @
2 identifier fn;
3 identifier irq, dev_id;
4 typedef irqreturn_t;
5 @@
6
7 static irqreturn_t fn (int irq, void *dev_id)
8 {
9     ...
10 }
11
12 @@
13 identifier rule1.fn;
14 expression E1, E2, E3;
15 @@
16
17 fn(E1, E2
18 - ,E3
19     )
```

drivers/atm/firestream.c at line 1653 before transformation

```
1 static void fs_poll (unsigned long data)
2 {
3     struct fs_dev *dev = (struct fs_dev *) data;
4
5     fs_irq (0, dev, NULL);
6     dev->timer.expires = jiffies + FS_POLL_FREQ;
7     add_timer (&dev->timer);
8 }
```

drivers/atm/firestream.c at line 1653 after transformation

```
1 static void fs_poll (unsigned long data)
2 {
3     struct fs_dev *dev = (struct fs_dev *) data;
4
5     fs_irq (0, dev);
6     dev->timer.expires = jiffies + FS_POLL_FREQ;
7     add_timer (&dev->timer);
8 }
```

4.1.3 Introduction of a macro

To avoid code duplication or error prone code, the kernel provides macros such as `BUG_ON`, `DIV_ROUND_UP` and `FIELD_SIZE`. In these cases, the semantic patches look for the old code pattern and replace it by the new code.

A semantic patch to introduce uses of the `DIV_ROUND_UP` macro looks for the corresponding expression, *i.e.*, $(n + d - 1) / d$. When some code is matched, the metavariables `n` and `d` are bound to their corresponding expressions. Finally, Coccinelle rewrites the code with the `DIV_ROUND_UP` macro using the values bound to `n` and `d`, as illustrated in the patch that follows.

Semantic patch to introduce uses of the `DIV_ROUND_UP` macro

```
1 @ haskernel @
2 @@
3
4 #include <linux/kernel.h>
5
6 @ depends on haskernel @
7 expression n,d;
8 @@
9
10 (
11 - ((n) + (d)) - 1) / (d))
12 + DIV_ROUND_UP(n,d)
13 |
14 - ((n) + ((d) - 1)) / (d))
15 + DIV_ROUND_UP(n,d)
16 )
```

Example of a generated patch hunk

```
1 --- a/drivers/atm/horizon.c
2 +++ b/drivers/atm/horizon.c
3 @@ -698,7 +698,7 @@ got_it:
4         if (bits)
5             *bits = (div<<CLOCK_SELECT_SHIFT) | (pre-1);
6         if (actual) {
7 -             *actual = (br + (pre<<div) - 1) / (pre<<div);
8 +             *actual = DIV_ROUND_UP(br, pre<<div);
9             PRINTD (DBG_QOS, "actual_rate:_%u", *actual);
10        }
11        return 0;
```

The `BUG_ON` macro makes an assertion about the value of an expression. However, because some parts of the kernel define `BUG_ON` to be the empty statement when debugging is not wanted, care must be taken when the asserted expression may have some side-effects, as is the case of a function call. Thus, we create a rule introducing `BUG_ON` only in the case when the asserted expression does not perform a function call.

On a particular piece of code that has the form of a function call is a use of `unlikely`, which informs the compiler that a particular expression is unlikely to be true. In this case, because `unlikely` does not perform any side effects, it is safe to use `BUG_ON`. The second rule takes care of this case. It furthermore disables the isomorphism that allows a call to `unlikely` be replaced with its argument, as then the second rule would be the same as the first one.

```

1 @@
2 expression E, f;
3 @@
4
5 (
6   if (<+... f(...) ...+>) { BUG(); }
7 |
8 - if (E) { BUG(); }
9 + BUG_ON(E);
10 )
11
12 @ disable unlikely @
13 expression E, f;
14 @@
15
16 (
17   if (<+... f(...) ...+>) { BUG(); }
18 |
19 - if (unlikely(E)) { BUG(); }
20 + BUG_ON(E);
21 )

```

For instance, using the semantic patch above, Coccinelle generates patches like the following one.

```

1 --- a/fs/ext3/balloc.c
2 +++ b/fs/ext3/balloc.c
3 @@ -232,8 +232,7 @@ restart:
4         prev = rsv;
5     }
6     printk("Window_map_complete.\n");
7 -     if (bad)
8 -         BUG();
9 +     BUG_ON(bad);
10 }
11 #define rsv_window_dump(root, verbose) \
12     __rsv_window_dump((root), (verbose), __FUNCTION__)

```

4.1.4 Look for NULL dereference

This SmPL match looks for NULL dereferences. Once an expression has been compared to NULL, a dereference to this expression is prohibited unless the pointer variable is reassigned.

Original

```
1 foo = kmalloc(1024);
2 if (!foo) {
3     printk ("Error_%s", foo->here);
4     return;
5 }
6 foo->ok = 1;
7 return;
```

Semantic match

```
1 @@
2 expression E, E1;
3 identifier f;
4 statement S1,S2,S3;
5 @@
6
7 * if (E == NULL)
8 {
9     ... when != if (E == NULL) S1 else S2
10     when != E = E1
11 * E->f
12     ... when any
13     return ...;
14 }
15 else S3
```

Matched lines

```
1 foo = kmalloc(1024);
2 if (!foo) {
3     printk ("Error %s", foo->here);
4     return;
5 }
6 foo->ok = 1;
7 return;
```

4.1.5 Reference counter: the of_xxx API

Coccinelle can embed Python code. Python code is used inside special SmPL rule annotated with `script:python`. Python rules inherit metavariables, such as identifier or token positions, from other SmPL rules. The inherited metavariables can then be manipulated by Python code.

The following semantic match looks for a call to the `of_find_node_by_name` function. This call increments a counter which must be decremented to release the resource. Then, when there is no call to `of_node_put`, no new assignment to the `device_node` variable `n` and a return statement is reached, a bug is detected and the position `p1` and `p2` are initialized. As the Python only depends on the positions `p1` and `p2`, it is evaluated. In the following case, some emacs Org mode data are produced. This example illustrates the various fields that can be accessed in the Python code from a position variable.

```
1 @ r exists @
2 local idexpression struct device_node *n;
3 position p1, p2;
4 statement S1,S2;
5 expression E,E1;
6 @@
7
8 (
9 if (!(n@p1 = of_find_node_by_name(...))) S1
10 |
11 n@p1 = of_find_node_by_name(...)
12 )
13 <... when != of_node_put(n)
14     when != if (...) { <+... of_node_put(n) ...+> }
15     when != true !n || ...
16     when != n = E
17     when != E = n
18 if (!n || ...) S2
19 ...>
20 (
21     return <+...n...+>;
22 |
23 return@p2 ...;
24 |
25 n = E1
26 |
27 E1 = n
28 )
29
30 @ script:python @
31 p1 << r.p1;
32 p2 << r.p2;
33 @@
34
35 print "* TODO [[view:%s::face=ov1-face1::linb=%s::colb=%s::cole=%s][inc.
36     counter:%s::%s]]" % (p1[0].file,p1[0].line,p1[0].column,p1[0].column_end,
37     p1[0].file,p1[0].line)
38 print "[[view:%s::face=ov1-face2::linb=%s::colb=%s::cole=%s][return]]" % (p2
39     [0].file,p2[0].line,p2[0].column,p2[0].column_end)
```

Lines 13 to 17 list a variety of constructs that should not appear between a call to `of_find_node_by_name` and a buggy return site. Examples are a call to `of_node_put` (line 13) and a transition into the then branch of a conditional testing whether `n` is `NULL` (line 15). Any number of conditionals testing whether `n` is `NULL` are allowed as indicated by the use of a nest `<... >` to describe the path between the call to `of_find_node_by_name`, the return and the conditional in the pattern on line 18.

The previously semantic match has been used to generate the following lines. They may be edited using the emacs Org mode to navigate in the code from a site to another.

```
1 * TODO [[view:/linux-next/arch/powerpc/platforms/pseries/setup.c::face=ovl-
   facel::linb=236::colb=18::cole=20][inc. counter:/linux-next/arch/powerpc/
   platforms/pseries/setup.c::236]]
2 [[view:/linux-next/arch/powerpc/platforms/pseries/setup.c::face=ovl-face2::
   linb=250::colb=3::cole=9][return]]
3 * TODO [[view:/linux-next/arch/powerpc/platforms/pseries/setup.c::face=ovl-
   facel::linb=236::colb=18::cole=20][inc. counter:/linux-next/arch/powerpc/
   platforms/pseries/setup.c::236]]
4 [[view:/linux-next/arch/powerpc/platforms/pseries/setup.c::face=ovl-face2::
   linb=245::colb=3::cole=9][return]]
```

Note : Coccinelle provides some predefined Python functions, *i.e.*, `cocci.print_main`, `cocci.print_sec` and `cocci.print_secs`. One could alternatively write the following SmPL rule instead of the previously presented one.

```
1 @ script:python @
2 p1 << r.p1;
3 p2 << r.p2;
4 @@
5
6 cocci.print_main(p1)
7 cocci.print_sec(p2, "return")
```

The function `cocci.print_secs` is used when there is several positions which are matched by a single position variable and that every matched position should be printed.

Any metavariable could be inherited in the Python code. However, accessible fields are not currently equally supported among them.

4.1.6 Filtering identifiers, declarers or iterators with regular expression

If you consider the following SmPL file which uses the regexp functionality to filter the identifiers that contain, begin or end by `foo`,

```
1 @anyid@
2 type t;
3 identifier id;
4 @@
5 t id () {...}
6
7 @script:python@
8 x << anyid.id;
9 @@
10 print "Identifier: %s" % x
11
12 @contains@
13 type t;
14 identifier foo ~= ".*foo";
15 @@
16 t foo () {...}
17
18 @script:python@
19 x << contains.foo;
20 @@
21 print "Contains foo: %s" % x
22
23 @endsby@
24 type t;
25 identifier foo ~= ".*foo$";
26 @@
27
28 t foo () {...}
29
30 @script:python@
31 x << endsby.foo;
32 @@
33 print "Ends by foo: %s" % x
34
35 @beginsby@
36 type t;
37 identifier foo ~= "^foo";
38 @@
39 t foo () {...}
40
41 @script:python@
42 x << beginsby.foo;
43 @@
44 print "Begins by foo: %s" % x
```

and the following C program, on the left, which defines the functions `foo`, `bar`, `foobar`, `barfoobar` and `barfoo`, you will get the result on the right.

```
1 int foo () { return 0; }
2 int bar () { return 0; }
3 int foobar () { return 0; }
4 int barfoobar () { return 0; }
5 int barfoo () { return 0; }
6
7 Identifier: foo
8 Identifier: bar
9 Identifier: foobar
10 Identifier: barfoobar
11 Identifier: barfoo
12 Contains foo: foo
13 Contains foo: foobar
14 Contains foo: barfoobar
15 Contains foo: barfoo
16 Ends by foo: foo
17 Ends by foo: barfoo
18 Begins by foo: foo
19 Begins by foo: foobar
```

4.2 Tips and Tricks

4.2.1 How to remove useless parentheses?

If you want to rewrite any access to a pointer value by a function call, you may use the following semantic patch.

```
1 - a = *b
2 + a = readb(b)
```


However, if for some reason your code looks like `bar = *(foo)`, you will end up with `bar = readb((foo))` as the extra parentheses around `foo` are captured by the metavariable `b`.

In order to generate better output code, you can use the following semantic patch instead.

```
1 - a = *(b)
2 + a = readb(b)
```

And rely on your `standard.iso` isomorphism file which should contain:

```
1 Expression
2 @ paren @
3 expression E;
4 @@
5
6 (E) => E
```

Cocinelle will then consider `bar = *(foo)` as equivalent to `bar = *foo` (but not the other way around) and capture both. Finally, it will generate `bar = readb(foo)` as expected.

Chapter 5

Isomorphisms and `standard.iso`

Chapter 6

Parsing C, cpp, and standard.h

Chapter 7

Developing a Semantic Patch

Chapter 8

Advanced Features

Part II

Reference Manual

Chapter 9

SmPL grammar

This document presents the grammar of the SmPL language used by the Coccinelle tool. For the most part, the grammar is written using standard notation. In some rules, however, the left-hand side is in all uppercase letters. These are macros, which take one or more grammar rule right-hand-sides as arguments. The grammar also uses some unspecified nonterminals, such as `id`, `const`, etc. These refer to the sets suggested by the name, *i.e.*, `id` refers to the set of possible C-language identifiers, while `const` refers to the set of possible C-language constants. A HTML version of this documentation is available online at http://coccinelle.lip6.fr/docs/main_grammar.html.

9.1 Program

```
program ::= include_cocci* changeset+  
include_cocci ::= using string  
| using pathToIsoFile  
| virtual id (, id)*  
changeset ::= metavariables transformation  
| script_metavariables script_code
```

`script_code` is any code in the chosen scripting language. Parsing of the semantic patch does not check the validity of this code; any errors are first detected when the code is executed. Furthermore, `@` should not be use in this code. Spatch scans the script code for the next `@` and considers that to be the beginning of the next rule, even if `@` occurs within e.g., a string or a comment.

`virtual` keyword is used to declare virtual rules. Virtual rules may be subsequently used as a dependency for the rules in the SmPL file. Whether a virtual rule is defined or not is controlled by the `-D` option on the command line.

9.2 Metavariables for transformations

The *rulename* portion of the metavariable declaration can specify properties of a rule such as its name, the names of the rules that it depends on, the isomorphisms to be used in processing the rule, and whether quantification over paths should be universal or existential. The optional annotation *expression* indicates that the pattern is to be considered as matching an expression, and thus can be used to avoid some parsing problems.

The *metadecl* portion of the metavariable declaration defines various types of metavariables that will be used for matching in the transformation section.

```

metavariables ::= @@ metadect* @@
                | @ rulename @ metadect* @@
rulename      ::= id [extends id] [depends on dep] [iso] [disable-iso] [exists] [expression]
dep           ::= pnrule
                | dep && dep
                | dep || dep
pnrule       ::= id
                | !id
                | ever id
                | never id
                | (dep)
iso           ::= using string (, string)*
disable-iso  ::= disable COMMA_LIST(id)
exists       ::= exists
                | forall
COMMA_LIST(elem) ::= elem (, elem)*

```

The keyword `disable` is normally used with the names of isomorphisms defined in `standard.iso` or whatever isomorphism file has been included. There are, however, some other isomorphisms that are built into the implementation of Coccinelle and that can be disabled as well. Their names are given below. In each case, the text describes the standard behavior. Using `disable-iso` with the given name disables this behavior.

- `optional_storage`: A SmPL function definition that does not specify any visibility (i.e., static or extern), or a SmPL variable declaration that does not specify any storage (i.e., auto, static, register, or extern), matches a function declaration or variable declaration with any visibility or storage, respectively.
- `optional_qualifier`: This is similar to `optional_storage`, except that here is it the qualifier (i.e., const or volatile) that does not have to be specified in the SmPL code, but may be present in the C code.
- `value_format`: Integers in various formats, e.g., 1 and 0x1, are considered to be equivalent in the matching process.
- `comm_assoc`: An expression of the form `exp bin_op . . .`, where `bin_op` is commutative and associative, is considered to match any top-level sequence of `bin_op` operators containing `exp` as the top-level argument.

The possible types of metavariable declarations are defined by the grammar rule below. Metavariables should occur at least once in the transformation immediately following their declaration. Fresh metavariables must only be used in + code. These properties are not expressed in the grammar, but are checked by a subsequent analysis. The metavariables are designated according to the kind of terms they can match, such as a statement, an identifier, or an expression. An expression metavariable can be further constrained by its type. A declaration metavariable matches the declaration of one or more variables, all sharing the same type specification (e.g., `int a,b,c=3;`). A field metavariable does the same, but for structure fields.


```

metadecl ::= fresh identifier ids ;
| identifier COMMA_LIST(pmid_with_regexp) ;
| identifier COMMA_LIST(pmid_with_virt_or_not_eq) ;
| parameter [list] ids ;
| parameter list [ id ] ids ;
| parameter list [ const ] ids ;
| type ids ;
| statement [list] ids ;
| declaration [list] ids ;
| field [list] ids ;
| typedef ids ;
| declarer name ids ;
| declarer COMMA_LIST(pmid_with_regexp) ;
| declarer COMMA_LIST(pmid_with_not_eq) ;
| iterator name ids ;
| iterator COMMA_LIST(pmid_with_regexp) ;
| iterator COMMA_LIST(pmid_with_not_eq) ;
| [local] idexpression [ctype] COMMA_LIST(pmid_with_not_eq) ;
| [local] idexpression [{ctypes} *] COMMA_LIST(pmid_with_not_eq) ;
| [local] idexpression *+ COMMA_LIST(pmid_with_not_eq) ;
| expression list ids ;
| expression *+ COMMA_LIST(pmid_with_not_eq) ;
| expression enum * * COMMA_LIST(pmid_with_not_eq) ;
| expression struct * * COMMA_LIST(pmid_with_not_eq) ;
| expression union * * COMMA_LIST(pmid_with_not_eq) ;
| expression COMMA_LIST(pmid_with_not_ceq) ;
| expression list [ id ] ids ;
| expression list [ const ] ids ;
| ctype [ ] COMMA_LIST(pmid_with_not_eq) ;
| ctype COMMA_LIST(pmid_with_not_ceq) ;
| {ctypes} * * COMMA_LIST(pmid_with_not_ceq) ;
| {ctypes} * * [ ] COMMA_LIST(pmid_with_not_eq) ;
| constant [ctype] COMMA_LIST(pmid_with_not_eq) ;
| constant [{ctypes} *] COMMA_LIST(pmid_with_not_eq) ;
| position [any] COMMA_LIST(pmid_with_not_eq_mid) ;

```

It is possible to specify that an expression list or a parameter list metavariable should match a specific number of expressions or parameters.

```

ids          ::= COMMA_LIST(pmid)
pmid         ::= id
              | mid
mid          ::= rulename_id.id
pmid_with_regexp ::= pmid ~ = regexp
pmid_with_not_eq ::= pmid [!= id_or_meta]
                 | pmid [!= { COMMA_LIST(id_or_meta) }]
pmid_with_virt_or_not_eq ::= virtual.id
                          | pmid_with_not_eq
pmid_with_not_ceq ::= pmid [!= id_or_cst]
                   | pmid [!= { COMMA_LIST(id_or_cst) }]
id_or_cst     ::= id
              | integer
id_or_meta    ::= id
              | rulename_id.id
pmid_with_not_eq_mid ::= pmid [!= mid]
                       | pmid [!= { COMMA_LIST(mid) }]

```

Subsequently, we refer to arbitrary metavariables as `metaidty`, where *ty* indicates the *metakind* used in the declaration of the variable. For example, `metaidType` refers to a metavariable that was declared using `type` and stands for any type.

The *c_{type}* and *c_{types}* nonterminals are used by both the grammar of metavariable declarations and the grammar of transformations, and are defined on page 27.

An identifier metavariable with `virtual` as its “rule name” is given a value on the command line. For example, if a semantic patch contains a rule that declares an identifier metavariable with the name `virtual.alloc`, then the command line could contain `-D alloc=kmalloc`. There should not be space around the `=`. An example is in `demos/vm.cocci` and `demos/vm.c`.

9.3 Metavariables for scripts

Metavariables for scripts can only be inherited from transformation rules. In the spirit of scripting languages such as Python that use dynamic typing, metavariables for scripts do not include type declarations.

```

script_metavariables ::= @ script:language [rulename] [depends on dep] @ script_metadecl* @@
                       | @ initialize:language [depends on dep] @
                       | @ finalize:language [depends on dep] @
language             ::= python
                       | ocaml
script_metadecl     ::= id << rulename_id.id ;
                       | id ;

```

Currently, the only scripting languages that are supported are Python and OCaml, indicated using `python` and `ocaml`, respectively. The set of available scripting languages may be extended at some point.

Script rules declared with `initialize` are run before the treatment of any file. Script rules declared with `finalize` are run when the treatment of all of the files has completed. There can be at most one of each per scripting language (thus currently at most one of each). Initialize and finalize script rules do not have access to SmPL metavariables. Nevertheless, a finalize script rule can access any variables initialized by the other script rules, allowing information to be transmitted from the matching process to the finalize rule.

A script metavariable that does not specify an origin, using `<<`, is newly declared by the script. This metavariable should be assigned to a string and can be inherited by subsequent rules as an identifier. In Python, the assignment of such a metavariable *x* should refer to the metavariable as `coccinelle.x`. Examples are in the files `demos/pythontococci.cocci` and `demos/camltococci.cocci`.

In an ocaml script, the following extended form of *script_metadecl* may be used:

```
script_metadecl ::= (id, id) << rulename_id.id ;
                  | id << rulename_id.id ;
                  | id ;
```

In a declaration of the form $(id, id) \ll rulename_id.id ;$, the left component of (id, id) receives a string representation of the value of the inherited metavariable while the right component receives its abstract syntax tree. The file `parsing_c/ast_c.ml` in the Coccinelle implementation gives some information about the structure of the abstract syntax tree. Either the left or right component may be replaced by `_`, indicating that the string representation or abstract syntax trees representation is not wanted, respectively.

9.4 Transformation

The transformation specification essentially has the form of C code, except that lines to remove are annotated with `-` in the first column, and lines to add are annotated with `+`. A transformation specification can also use *dots*, “`...`”, describing an arbitrary sequence of function arguments or instructions within a control-flow path. Dots may be modified with a *when* clause, indicating a pattern that should not occur anywhere within the matched sequence. Finally, a transformation can specify a disjunction of patterns, of the form $(pat_1 \mid \dots \mid pat_n)$ where each $(, \mid$ or $)$ is in column 0 or preceded by `\`.

The grammar that we present for the transformation is not actually the grammar of the SmPL code that can be written by the programmer, but is instead the grammar of the slice of this consisting of the `-` annotated and the unannotated code (the context of the transformed lines), or the `+` annotated code and the unannotated code. For example, for parsing purposes, the following transformation is split into the two variants shown below and each is parsed separately.

```

1  proc_info_func(...) {
2      <...
3  -   hostno
4  +   hostptr->host_no
5      ...>
6  }

1  proc_info_func(...) {
2      <...
3  -   hostno
4      ...>
5  }

1  proc_info_func(...) {
2      <...
3  +   hostptr->host_no
4      ...>
5  }
```

Requiring that both slices parse correctly ensures that the rule matches syntactically valid C code and that it produces syntactically valid C code. The generated parse trees are then merged for use in the subsequent matching and transformation process.

The grammar for the minus or plus slice of a transformation is as follows:

```

transformation ::= include+
                | OPTDOTSEQ(expr, when)
                | OPTDOTSEQ(decl_stmt+, when)
                | OPTDOTSEQ(fundecl, when)
include        ::= #include include_string
when           ::= when != when_code
                | when = rule_elem_stmt
                | when COMMA_LIST(any_strict)
                | when true != expr
                | when false != expr
when_code      ::= OPTDOTSEQ(decl_stmt+, when)
                | OPTDOTSEQ(expr, when)
rule_elem_stmt ::= one_decl
                | expr;
                | return [expr];
                | break;
                | continue;
                | \ (rule_elem_stmt (\ | rule_elem_stmt)+\)
any_strict     ::= any
                | strict
                | forall
                | exists

```

```

OPTDOTSEQ(grammar_ds, when_ds) ::=
    [... [when_ds] grammar_ds (... [when_ds] grammar_ds)* [... [when_ds]]

```

Lines may be annotated with an element of the set $\{-, +, *\}$ or the singleton $?$, or one of each set. $?$ represents at most one match of the given pattern. $*$ is used for semantic match, *i.e.*, a pattern that highlights the fragments annotated with $*$, but does not perform any modification of the matched code. $*$ cannot be mixed with $-$ and $+$. There are some constraints on the use of these annotations:

- Dots, *i.e.* \dots , cannot occur on a line marked $+$.
- Nested dots, *i.e.*, dots enclosed in $<$ and $>$, cannot occur on a line with any marking.

Each element of a disjunction must be a proper term like an expression, a statement, an identifier or a declaration. Thus, the rule on the left below is not a syntactically correct SmPL rule. One may use the rule on the right instead.

<pre> 1 @@ 2 type T; 3 T b; 4 @@ 5 6 (7 writeb(..., 8 9 readb(10) 11 -(T) 12 b) </pre>	<pre> 1 @@ 2 type T; 3 T b; 4 @@ 5 6 (7 read 8 9 write 10) 11 (... , 12 -(T) 13 b) </pre>
--	--

9.5 Types

```

ctypes ::= COMMA_LIST(ctype)
ctype ::= [const_vol] generic_ctype **
          | [const_vol] void *†
          | (ctype (| ctype)*)

const_vol ::= const
            | volatile

generic_ctype ::= ctype_qualif
                | [ctype_qualif] char
                | [ctype_qualif] short
                | [ctype_qualif] int
                | [ctype_qualif] long
                | [ctype_qualif] long long
                | double
                | float
                | size_t
                | ssize_t
                | ptrdiff_t
                | enum id { PARAMSEQ(dot_expr, exp_whencode) [, ] }
                | [struct|union] id [{ struct_decl_list* }]

ctype_qualif ::= unsigned
                | signed

struct_decl_list ::= struct_decl_list_start
struct_decl_list_start ::= struct_decl
                          | struct_decl struct_decl_list_start
                          | ... [when != struct_decl]† [continue_struct_decl_list]

continue_struct_decl_list ::= struct_decl struct_decl_list_start
                             | struct_decl

struct_decl ::= ctype d_ident;
               | fn_ctype (* d_ident) (PARAMSEQ(name_opt_decl, ε));)
               | [const_vol] id d_ident;

d_ident ::= id [[expr]]*
fn_ctype ::= generic_ctype **
             | void **

name_opt_decl ::= decl
                 | ctype
                 | fn_ctype

```

[†] The optional when construct ends at the end of the line.

9.6 Function declarations

```

fundecl ::= [fn_ctype] funinfo* funid ([PARAMSEQ(param,  $\varepsilon$ )] { [stmt_seq] }
funproto ::= [fn_ctype] funinfo* funid ([PARAMSEQ(param,  $\varepsilon$ )]);
funinfo ::= inline
           | storage
storage ::= static
           | auto
           | register
           | extern
funid ::= id
         | metaidld
param ::= type id
         | metaidParam
         | metaidParamList
decl ::= ctype id
         | fn_ctype (* id) ([PARAMSEQ(name_opt_decl,  $\varepsilon$ )]
         | void
         | metaidParam

```

PARAMSEQ(*gram_p*, *when_p*) ::= *COMMA_LIST*(*gram_p* | ... [*when_p*])

To match a function it is not necessary to provide all of the annotations that appear before the function name. For example, the following semantic patch:

```

1 @@
2 @@
3
4 foo() { ... }

```

matches a function declared as follows:

```

1 static int foo() { return 12; }

```

This behavior can be turned off by disabling the `optional_storage` isomorphism. If one adds code before a function declaration, then the effect depends on the kind of code that is added. If the added code is a function definition or CPP code, then the new code is placed before all information associated with the function definition, including any comments preceding the function definition. On the other hand, if the new code is associated with the function, such as the addition of the keyword `static`, the new code is placed exactly where it appears with respect to the rest of the function definition in the semantic patch. For example,

```

1 @@
2 @@
3
4 + static
5 foo() { ... }

```

causes `static` to be placed just before the function name. The following causes it to be placed just before the type

```

1 @@
2 type T;
3 @@
4
5 + static
6 T foo() { ... }

```

It may be necessary to consider several cases to ensure that the added code is placed in the right position. For example, one may need one pattern that considers that the function is declared `inline` and another that considers that it is not.

9.7 Declarations

```
decl_var ::= common_decl
           | [storage] ctype COMMA_LIST(d_ident) ;
           | [storage] [const_vol] id COMMA_LIST(d_ident) ;
           | [storage] fn_ctype ( * d_ident ) ( PARAMSEQ(name_opt_decl,  $\epsilon$ ) ) = initialize ;
           | typedef ctype typedef_ident ;
one_decl ::= common_decl
           | [storage] ctype id;
           | [storage] [const_vol] id d_ident ;
common_decl ::= ctype;
              | funproto
              | [storage] ctype d_ident = initialize ;
              | [storage] [const_vol] id d_ident = initialize ;
              | [storage] fn_ctype ( * d_ident ) ( PARAMSEQ(name_opt_decl,  $\epsilon$ ) ) ;
              | decl_ident ( [COMMA_LIST(expr) ] ) ;
initialize ::= dot_expr
              | metaidInitialiser
              | { [COMMA_LIST(dot_expr) ] }
init_list_elem ::= dot_expr
                  | designator = dot_expr
                  | id : dot_expr
designator ::= . id
              | [ dot_expr ]
              | [ dot_expr ... dot_expr ]
decl_ident ::= DeclarerId
              | metaidDeclarer
```

9.8 Statements

The first rule *statement* describes the various forms of a statement. The remaining rules implement the constraints that are sensitive to the context in which the statement occurs: *single_statement* for a context in which only one statement is allowed, and *decl_statement* for a context in which a declaration, statement, or sequence thereof is allowed.

```

stmt ::= include
      | metaidStmt
      | expr;
      | if (dot_expr) single_stmt [else single_stmt]
      | for ([dot_expr]; [dot_expr]; [dot_expr]) single_stmt
      | while (dot_expr) single_stmt
      | do single_stmt while (dot_expr);
      | iter_ident (dot_expr*) single_stmt
      | switch ([dot_expr]) {case_line* }
      | return [dot_expr];
      | { [stmt_seq] }
      | NEST(decl_stmt+, when)
      | NEST(expr, when)
      | break;
      | continue;
      | id:
      | goto id;
      | {stmt_seq }
single_stmt ::= stmt
              | OR(stmt)
decl_stmt ::= metaidStmtList
              | decl_var
              | stmt
              | OR(stmt_seq)
stmt_seq ::= decl_stmt* [DOTSEQ(decl_stmt+, when) decl_stmt*]
            | decl_stmt* [DOTSEQ(expr, when) decl_stmt*]
case_line ::= default : stmt_seq
              | case dot_expr : stmt_seq
iter_ident ::= iteratorId
               | metaidIterator

```

```

OR(gram_o) ::= ( gram_o (| gram_o)*)
DOTSEQ(gram_d, when_d) ::= ... [when_d] (gram_d ... [when_d])*
NEST(gram_n, when_n) ::= <... [when_n] gram_n (... [when_n] gram_n)* ...>
                       | <+... [when_n] gram_n (... [when_n] gram_n)* ... +>

```

OR is a macro that generates a disjunction of patterns. The three tokens (, |, and) must appear in the leftmost column, to differentiate them from the parentheses and bit-or tokens that can appear within expressions (and cannot appear in the leftmost column). These token may also be preceded by \ when they are used in an other column. These tokens are furthermore different from (, |, and), which are part of the grammar metalanguage.

9.9 Expressions

A nest or a single ellipsis is allowed in some expression contexts, and causes ambiguity in others. For example, in a sequence ...*expr* ..., the nonterminal *expr* must be instantiated as an explicit C-language expression, while in an array reference, *expr*₁ [*expr*₂], the nonterminal *expr*₂, because it is delimited by brackets, can be also instantiated as ..., representing an arbitrary expression. To distinguish between the various possibilities, we define three nonterminals for expressions: *expr* does not allow either top-level nests or ellipses, *nest_expr* allows a nest but not an ellipsis, and *dot_expr* allows both. The *EXPR* macro is used to express these variants in a concise way.


```

expr ::= EXPR(expr)
nest_expr ::= EXPR(nest_expr)
           | NEST(nest_expr, exp_whencode)
dot_expr ::= EXPR(dot_expr)
           | NEST(dot_expr, exp_whencode)
           | ... [exp_whencode]
EXPR(exp) ::= exp assign_op exp
           | exp++
           | exp--
           | unary_op exp
           | exp bin_op exp
           | exp ? dot_expr : exp
           | (type) exp
           | exp [dot_expr]
           | exp . id
           | exp -> id
           | exp ([PARAMSEQ(arg, exp_whencode)] )
           | id
           | metaidExp
           | metaidConst
           | const
           | (dot_expr)
           | OR(exp)
arg ::= nest_expr
       | metaidExpList
exp_whencode ::= when != expr
assign_op ::= = | -= | += | *= | /= | %=
           | &= | |= | ^= | <<= | >>=
bin_op ::= * | / | % | + | -
           | << | >> | ^ | & | |
           | < | > | <= | >= | == | != | && | ||
unary_op ::= ++ | -- | & | * | + | - | !

```

9.10 Constant, Identifiers and Types for Transformations

```

const ::= string
           | [0-9]+
           | ...
string ::= "[^"]*"
id ::= id | metaidId
typedef_ident ::= id | metaidType
type ::= ctype | metaidType
pathToIsoFile ::= <.*>
regexp ::= "[^"]*"

```

Chapter 10

spatch command line options

10.1 Introduction

This document describes the options provided by Coccinelle. The options have an impact on various phases of the semantic patch application process. These are:

1. Selecting and parsing the semantic patch.
2. Selecting and parsing the C code.
3. Application of the semantic patch to the C code.
4. Transformation.
5. Generation of the result.

One can either initiate the complete process from step 1, or to perform step 1 or step 2 individually.

Coccinelle has quite a lot of options. The most common usages are as follows, for a semantic match `foo.cocci`, a C file `foo.c`, and a directory `foodir`:

- `spatch -parse_cocci foo.cocci`: Check that the semantic patch is syntactically correct.
- `spatch -parse_c foo.c`: Check that the C file is syntactically correct. The Coccinelle C parser tries to recover during the parsing process, so if one function does not parse, it will start up again with the next one. Thus, a parse error is often not a cause for concern, unless it occurs in a function that is relevant to the semantic patch.
- `spatch -sp_file foo.cocci foo.c`: Apply the semantic patch `foo.cocci` to the file `foo.c` and print out any transformations as a diff.
- `spatch -sp_file foo.cocci foo.c -debug`: The same as the previous case, but print out some information about the matching process.
- `spatch -sp_file foo.cocci -dir foodir`: Apply the semantic patch `foo.cocci` to all of the C files in the directory `foodir`.
- `spatch -sp_file foo.cocci -dir foodir -include_headers`: Apply the semantic patch `foo.cocci` to all of the C files and header files in the directory `foodir`.

In the rest of this document, the options are annotated as follows:

- **◆**: a basic option, that is most likely of interest to all users.

- \blacklozenge : an option that is frequently used, often for better understanding the effect of a semantic patch.
- \blacklozenge : an option that is likely to be rarely used, but whose effect is still comprehensible to a user.
- An option with no annotation is likely of interest only to developers.

10.2 Selecting and parsing the semantic patch

10.2.1 Standalone options

- \blacklozenge **-parse_cocci** *<file>* Parse a semantic patch file and print out some information about it.

10.2.2 The semantic patch

- \blacklozenge **-sp_file** *<file>*, **-c** *<file>*, **-cocci_file** *<file>* Specify the name of the file containing the semantic patch. The file name should end in `.cocci`. All three options do the same thing; the last two are deprecated.

10.2.3 Isomorphisms

- \blacklozenge **-iso**, **-iso_file** Specify a file containing isomorphisms to be used in place of the standard one. Normally one should use the `using` construct within a semantic patch to specify isomorphisms to be used *in addition to* the standard ones.
- \blacklozenge **-iso_limit** *<int>* Limit the depth of application of isomorphisms to the specified integer.
- \blacklozenge **-no_iso_limit** Put no limit on the number of times that isomorphisms can be applied. This is the default.
- \blacklozenge **-disable_iso** Disable a specific isomorphism from the command line. This option can be specified multiple times.
- track_iso** Gather information about isomorphism usage.
- profile_iso** Gather information about the time required for isomorphism expansion.

10.2.4 Display options

- \blacklozenge **-show_cocci** Show the semantic patch that is being processed before expanding isomorphisms.
- \blacklozenge **-show_SP** Show the semantic patch that is being processed after expanding isomorphisms.
- \blacklozenge **-show_ctl_text** Show the representation of the semantic patch in CTL.
- \blacklozenge **-ctl_inline_let** Sometimes `let` is used to name intermediate terms CTL representation. This option causes the `let`-bound terms to be inlined at the point of their reference. This option implicitly sets **-show_ctl_text**.
- \blacklozenge **-ctl_show_mcodekind** Show transformation information within the CTL representation of the semantic patch. This option implicitly sets **-show_ctl_text**.
- \blacklozenge **-show_ctl_tex** Create a LaTeX files showing the representation of the semantic patch in CTL.

10.3 Selecting and parsing the C files

10.3.1 Standalone options

- ◆ **-parse_c** *<file/dir>* Parse a `.c` file or all of the `.c` files in a directory. This generates information about any parse errors encountered.
- ◆ **-parse_h** *<file/dir>* Parse a `.h` file or all of the `.h` files in a directory. This generates information about any parse errors encountered.
- ◆ **-parse_ch** *<file/dir>* Parse a `.c` or `.h` file or all of the `.c` or `.h` files in a directory. This generates information about any parse errors encountered.
- ◆ **-control_flow** *<file>*, **-control_flow** *<file>:<function>* Print a control-flow graph for all of the functions in a file or for a specific function in a file. This requires `dot` (<http://www.graphviz.org/>) and `gv`.
- ◆ **-type_c** *<file>* Parse a C file and pretty-print a version including type information.
- tokens_c** *<file>* Prints the tokens in a C file.
- parse_unparse** *<file>* Parse and then reconstruct a C file.
- compare_c** *<file>* *<file>*, **-compare_c_hardcoded** Compares one C file to another, or compare the file `tests/compare1.c` to the file `tests/compare2.c`.
- test_cfg_ifdef** *<file>* Do some special processing of `#ifdef` and display the resulting control-flow graph. This requires `dot` and `gv`.
- test_attributes** *<file>*, **-test_cpp** *<file>* Test the parsing of `cpp` code and attributes, respectively.

10.3.2 Selecting C files

An argument that ends in `.c` is assumed to be a C file to process. Normally, only one C file or one directory is specified. If multiple C files are specified, they are treated in parallel, *i.e.*, the first semantic patch rule is applied to all functions in all files, then the second semantic patch rule is applied to all functions in all files, etc. If a directory is specified then no files may be specified and only the rightmost directory specified is used.

- ◆ **-include_headers** This option causes header files to be processed independently. This option only makes sense if a directory is specified using **-dir**.
- ◆ **-use_glimpse** Use a glimpse index to select the files to which a semantic patch may be relevant. This option requires that a directory is specified. The index may be created using the script `coccinelle/scripts/glimpseindex_cocci.sh`. Glimpse is available at <http://webglimpse.net/>. In conjunction with the option **-patch_cocci** this option prints the regular expression that will be passed to glimpse.
- ◆ **-use_idutils** Use an `id-utils` index created using `lid` to select the files to which a semantic patch may be relevant. This option requires that a directory is specified. The index may be created using the script `coccinelle/scripts/idindex_cocci.sh`. In conjunction with the option **-patch_cocci** this option prints the regular expression that will be passed to glimpse.

◇ **-dir** Specify a directory containing C files to process. A trailing / is permitted on the directory name and has no impact on the result. By default, the include path will be set to the “include” subdirectory of this directory. A different include path can be specified using the option **-I**. **-dir** only considers the rightmost directory in the argument list. This behavior is convenient for creating a script that always works on a single directory, but allows the user of the script to override the provided directory with another one. Spatch collects the files in the directory using `find` and does not follow symbolic links.

-kbuild_info `<file>` The specified file contains information about which sets of files should be considered in parallel.

-disable_worth_trying_opt Normally, a C file is only processed if it contains some keywords that have been determined to be essential for the semantic patch to match somewhere in the file. This option disables this optimization and tries the semantic patch on all files.

-test `<file>` A shortcut for running Coccinelle on the semantic patch “`file.cocci`” and the C file “`file.c`”.

-testall A shortcut for running Coccinelle on all files in a subdirectory `tests` such that there are all of a `.cocci` file, a `.c` file, and a `.res` file, where the `.res` contains the expected result.

-test_okfailed, -test_regression_okfailed Other options for keeping track of tests that have succeeded and failed.

-compare_with_expected Compare the result of applying Coccinelle to `file.c` to the file `file.res` representing the expected result.

-expected_score_file `<file>` which score file to compare with in the `testall` run

10.3.3 Parsing C files

◇ **-show_c** Show the C code that is being processed.

◇ **-parse_error_msg** Show parsing errors in the C file.

◇ **-verbose_parsing** Show parsing errors in the C file, as well as information about attempts to accommodate such errors. This implicitly sets `-parse_error_msg`.

◇ **-type_error_msg** Show information about where the C type checker was not able to determine the type of an expression.

◇ **-int_bits** `<n>`, **-long_bits** `<n>` Provide integer size information. `n` is the number of bits in an unsigned integer or unsigned long, respectively. If only the option **-int_bits** is used, unsigned longs will be assumed to have twice as many bits as unsigned integers. If only the option **-long_bits** is used, unsigned ints will be assumed to have half as many bits as unsigned integers. This information is only used in determining the types of integer constants, according to the ANSI C standard (C89). If neither is provided, the type of an integer constant is determined by the sequence of “u” and “l” annotations following the constant. If there is none, the constant is assumed to be a signed integer. If there is only “u”, the constant is assumed to be an unsigned integer, etc.

◇ **-no_loops** Drop back edges for loops. This may make a semantic patch/match run faster, at the cost of not finding matches that wrap around loops.

-use_cache Use preparsed versions of the C files that are stored in a cache.

-debug_cpp, -debug_lexer, -debug_etdt, -debug_typedef Various options for debugging the C parser.

-filter_msg, -filter_define_error, -filter_passed_level Various options for debugging the C parser.

-only_return_is_error_exit In matching “...” in a semantic patch or when forall is specified, a rule must match all control-flow paths starting from a node matching the beginning of the rule. This is relaxed, however, for error handling code. Normally, error handling code is considered to be a conditional with only a then branch that ends in goto, break, continue, or return. If this option is set, then only a then branch ending in a return is considered to be error handling code. Usually a better strategy is to use `when strict` in the semantic patch, and then match explicitly the case where there is a conditional whose then branch ends in a return.

Macros and other preprocessor code

◆ **-macro_file <file>** Extra macro definitions to be taken into account when parsing the C files. This uses the provided macro definitions in addition to those in the default macro file.

◆ **-macro_file_builtins <file>** Builtin macro definitions to be taken into account when parsing the C files. This causes the macro definitions provided in the default macro file to be ignored and the ones in the specified file to be used instead.

◆ **-ifdef_to_if,-no_ifdef_to_if** The option **-ifdef_to_if** represents an `#ifdef` in the source code as a conditional in the control-flow graph when doing so represents valid code. **-no_ifdef_to_if** disables this feature. **-ifdef_to_if** is the default.

◆ **-use_if0_code** Normally code under `#if 0` is ignored. If this option is set then the code is considered, just like the code under any other `#ifdef`.

-noadd_typedef_root This seems to reduce the scope of a typedef declaration found in the C code.

Include files

◆ **-recursive_includes, -all_includes, -local_includes, -no_includes** These options control which include files mentioned in a C file are taken into account. **-really_all_includes** indicates that all included files mentioned in the `.c` file(s) or any included files will be processed. **-all_includes** indicates that all included files mentioned in the `.c` file(s) will be processed. **-local_includes** indicates that only included files in the current directory will be processed. **-no_includes** indicates that no included files will be processed. If the semantic patch contains type specifications on expression metavariables, then the default is **-local_includes**. Otherwise the default is **-no_includes**. At most one of these options can be specified.

◆ **-I <path>** This option specifies a directory in which to find non-local include files. This option can be used several times.

◆ **-relax_include_path** This option when combined with `-all_includes` causes the search for local include files to consider the current directory, even if the include patch specifies a subdirectory. This is really only useful for testing, eg with the option **-testall**

10.4 Application of the semantic patch to the C code

10.4.1 Feedback at the rule level during the application of the semantic patch

- ◆ **-show_bindings** Show the environments with respect to which each rule is applied and the bindings that result from each such application.
- ◆ **-show_dependencies** Show the status (matched or unmatched) of the rules on which a given rule depends. **-show_dependencies** implicitly sets **-show_bindings**, as the values of the dependencies are environment-specific.
- ◆ **-show_trying** Show the name of each program element to which each rule is applied.
- ◆ **-show_transinfo** Show information about each transformation that is performed. The node numbers that are referenced are the number of the nodes in the control-flow graph, which can be seen using the option **-control_flow** (the initial control-flow graph only) or the option **-show_flow** (the control-flow graph before and after each rule application).
- ◆ **-show_misc** Show some miscellaneous information.
- ◆ **-show_flow <file>**, **-show_flow <file>:<function>** Show the control-flow graph before and after the application of each rule.

-show_before_fixed_flow This is similar to **-show_flow**, but shows a preliminary version of the control-flow graph.

10.4.2 Feedback at the CTL level during the application of the semantic patch

- ◆ **-verbose_engine** Show a trace of the matching of atomic terms to C code.
- ◆ **-verbose_ctl_engine** Show a trace of the CTL matching process. This is unfortunately rather voluminous and not so helpful for someone who is not familiar with CTL in general and the translation of SmPL into CTL specifically. This option implicitly sets the option **-show_ctl_text**.
- ◆ **-graphical_trace** Create a pdf file containing the control flow graph annotated with the various nodes matched during the CTL matching process. Unfortunately, except for the most simple examples, the output is voluminous, and so the option is not really practical for most examples. This requires `dot` (<http://www.graphviz.org/>) and `pdftk`.
- ◆ **-gt_without_label** The same as **-graphical_trace**, but the PDF file does not contain the CTL code.
- ◆ **-partial_match** Report partial matches of the semantic patch on the C file. This can be substantially slower than normal matching.
- ◆ **-verbose_match** Report on when CTL matching is not applied to a function or other program unit because it does not contain some required atomic pattern. This can be viewed as a simpler, more efficient, but less informative version of **-partial_match**.

10.4.3 Actions during the application of the semantic patch

- ◆ **-D rulename** Run the patch considering that the virtual rule “rulename” is satisfied. Virtual rules should be declared at the beginning of the semantic patch in a comma separated list following the keyword `virtual`. Other rules can depend on the satisfaction or non satisfaction of these rules using the keyword `depends on` in the usual way.
 - ◆ **-D variable=value** Run the patch considering that the virtual identifier metavariable “variable” is bound to “value”. Any identifier metavariable can be designated as being virtual by giving it the rule name `virtual`. An example is in `demos/vm.coci`
 - ◆ **-allow_inconsistent_paths** Normally, a term that is transformed should only be accessible from other terms that are matched by the semantic patch. This option removes this constraint. Doing so, is unsafe, however, because the properties that hold along the matched path might not hold at all along the unmatched path.
 - ◆ **-disallow_nested_exps** In an expression that contains repeated nested subterms, *e.g.* of the form $f(f(x))$, a pattern can match a single expression in multiple ways, some nested inside others. This option causes the matching process to stop immediately at the outermost match. Thus, in the example $f(f(x))$, the possibility that the pattern $f(E)$, with metavariable E , matches with E as x will not be considered.
 - ◆ **-no_safe_expressions** normally, we check that an expression does not match something earlier in the disjunction. But for large disjunctions, this can result in a very big CTL formula. So this option give the user the option to say he doesn't want this feature, if that is the case.
 - ◆ **-pyoutput coccilib.output.Gtk, -pyoutput coccilib.output.Console** This controls whether Python output is sent to Gtk or to the console. **-pyoutput coccilib.output.Console** is the default. The Gtk option is currently not well supported.
- loop** When there is “...” in the semantic patch, the CTL operator **AU** is used if the current function does not contain a loop, and **AW** may be used if it does. This option causes **AW** always to be used.
- steps <int>** This limits the number of steps performed by the CTL engine to the specified number. This option is unsafe as it might cause a rule to fail due to running out of steps rather than due to not matching.
- bench <int>** This collects various information about the operations performed during the CTL matching process.
- popl, -popl_mark_all, -popl_keep_all_wits** These options use a simplified version of the SmPL language. **-popl_mark_all** and **-popl_keep_all_wits** implicitly set **-popl**.

10.5 Generation of the result

Normally, the only output is a diff printed to standard output.

- ◆ **-keep_comments** Don't remove comments adjacent to removed code.
- ◆ **-linux_spacing, -smpl_spacing** Control the spacing within the code added by the semantic patch. The option **-linux_spacing** causes `spatch` to follow the conventions of Linux, regardless of the spacing in the semantic patch. This is the default. The option **-smpl_spacing** causes `spatch` to follow the spacing given in the semantic patch, within individual lines.

- ◇ **-o** *<file>* The output file.
 - ◇ **-in_place** Modify the input file. By default, the input file is overwritten, with no backup.
 - ◇ **-backup_suffix** The suffix of the file to use in making a backup of the original file(s). This suffix should include the leading “.”, if one is desired. This option only has an effect when the option `-in_place` is also used.
 - ◇ **-out_place** Store modifications in a `.cocci_res` file.
 - ◇ **-no_show_diff** Normally, a diff between the original and transformed code is printed on the standard output. This option causes this not to be done.
 - ◇ **-U** Set number of diff context lines.
 - ◇ **-patch** *<path>* The prefix of the pathname of the directory or file name that should be dropped from the diff line in the generated patch. This is useful if you want to apply a patch only to a subdirectory of a source code tree but want to create a patch that can be applied at the root of the source code tree. An example could be `spatch -sp_file foo.cocci -dir /var/linuxes/linux-next/drivers -patch /var/linuxes/linux-next`. A trailing `/` is permitted on the directory name and has no impact on the result.
 - ◇ **-save_tmp_files** Coccinelle creates some temporary files in `/tmp` that it deletes after use. This option causes these files to be saved.
- debug_unparsing** Show some debugging information about the generation of the transformed code. This has the side-effect of deleting the transformed code.

10.6 Other options

10.6.1 Version information

- ◆ **-version** The version of Coccinelle. No other options are allowed.
- ◆ **-date** The date of the current version of Coccinelle. No other options are allowed.

10.6.2 Help

- ◆ **-h, -shorthelp** The most useful commands.
- ◆ **-help, -help, -longhelp** A complete listing of the available commands.

10.6.3 Controlling the execution of Coccinelle

- ◆ **-timeout** *<int>* The maximum time in seconds for processing a single file.
- ◆ **-max** *<int>* This option informs Coccinelle of the number of instances of Coccinelle that will be run concurrently. This option requires `-index`. It is usually used with `-dir`.

◇ **-index** *<int>* This option informs Coccinelle of which of the concurrent instances is the current one. This option requires **-max**.

◇ **-mod_distrib** When multiple instances of Coccinelle are run in parallel, normally the first instance processes the first *n* files, the second instance the second *n* files, etc. With this option, the files are distributed among the instances in a round-robin fashion.

-debugger Option for running Coccinelle from within the OCaml debugger.

-profile Gather timing information about the main Coccinelle functions.

-disable_once Print various warning messages every time some condition occurs, rather than only once.

10.6.4 Miscellaneous

◇ **-quiet** Suppress most output. This is the default.

-pad, -hrule *<dir>*, **-xxx, -ll**

Part III

Appendix

GNU Free Documentation License

GNU Free Documentation License Version 1.3, 3 November 2008

Copyright (C) 2000, 2001, 2002, 2007, 2008 Free Software Foundation, Inc. <<http://fsf.org/>> Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing

editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

The "publisher" means any person or entity that distributes copies of the Document to the public.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission. B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement. C. State on the Title page the name of the publisher of the Modified Version, as the publisher. D. Preserve all the copyright notices of the Document. E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices. F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below. G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice. H. Include an unaltered copy of this License. I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence. J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission. K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein. L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles. M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version. N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section. O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be

replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

11. RELICENSING

"Massive Multiauthor Collaboration Site" (or "MMC Site") means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A "Massive Multiauthor Collaboration" (or "MMC") contained in the site means any set of copyrightable works thus published on the MMC site.

"CC-BY-SA" means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

"Incorporate" means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is "eligible for relicensing" if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

Copyright (c) YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with...Texts." line with this: with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.