Coccinelle Usage (version 1.0.4)

July 16, 2017

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 the changes between the original and transformed code, using the program `diff`. `--sp-file` is optional in this and the following cases.
- `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.
- ⬙: an option that is frequently used, often for better understanding the effect of a semantic patch.
- ◇: 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.

2 Selecting and parsing the semantic patch

2.1 Standalone options

- **--parse-cocci** ⟨**file**⟩ Parse a semantic patch file and print out some information about it.

- **--debug-parse-cocci** Print some information about the definition of virtual rules and the bindings of virtual identifiers. This is particularly useful when using iteration, as it prints out this information for each iteration.

2.2 The semantic patch

- **--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. These options are optional. If they are not used, the single file whose name ends in `.cocci` is assoumed to be the name of the file containing the semantic patch.

- **--sp "semantic patch string"** Specify a semantic match as a command-line argument. See the section "Command-line semantic match" in the manual.

2.3 Isomorphisms

- **--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.

- **--iso-limit** ⟨**int**⟩ Limit the depth of application of isomorphisms to the specified integer.

- **--no-iso-limit** Put no limit on the number of times that isomorphisms can be applied. This is the default.

- **--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.

2.4 Display options

- **--show-cocci** Show the semantic patch that is being processed before expanding isomorphisms.

- **--show-SP** Show the semantic patch that is being processed after expanding isomorphisms.

- **--show-ctl-text** Show the representation of the semantic patch in CTL.

- **--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**.

◇ **--ctl-show-mcodekind** Show transformation information within the CTL representation of the semantic patch. This option implicitly sets **--show-ctl-text**.

◇ **--show-ctl-tex** Create a LaTeX files showing the representation of the semantic patch in CTL.

3 Selecting and parsing the C files

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`.

◇ **--control-flow-to-file** ⟨**file**⟩, **--control-flow-to-file** ⟨**file**⟩:⟨**function**⟩ Like **--control-flow** but just puts the dot output in a file in the *current* directory. For PATH/file.c, this produces file:xxx.dot for each (selected) function xxx in PATH/file.c.

◇ **--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.

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** [⟨**file**⟩] 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.

The optional file name option is the name of the file in which to find the index. It has been reported that the viewer seascope can be used to generate an appropriate index. If no file name is specified, the default is .id-utils.index.

◆ **--use-coccigrep** Use a version of grep implemented in Coccinelle to check that selected files are relevant to the semantic patch. This option is only relevant to the case of working on a complete directory, when parallelism is requested (max and index options). Otherwise it is the default, except when multiple files are requested to be treated as a single unit. In that case grep is used.

Note that coccigrep or grep is used even if glimpse or id-utils is selected, to account for imprecision in the index (glimpse at least does not distinguish between underline and space, leading to false positives).

◇ **--selected-only** Just show what files will be selected for processing.

◆ **--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`". The result is put in the file `/tmp/file.res`. If writing a file in /tmp with a non-fresh name is a concern, then do not use this option.

--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. If the argument -()-expected-score-file is provided, then that file is used for the result. Otherwise, the result goes in "tests/SCORE_expected.sexp". **Warning:** It is intended that not all of the test cases provided with Coccinelle actually pass.

◇ **--test-spacing** Like --testall, but ensures that the spacing is the same as in the .res file. If the argument -()-expected-spacing-score-file is provided, then that file is used for the result. Otherwise, the result goes in "tests/SCORE_spacing_expected.sexp".

--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

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 accomodate such errors. This implicitly sets **--parse-error-msg.**

◇ **--parse-handler** ⟨**file**⟩ Loads the file containing the OCaml code in charge of parse error reporting. This function should have arguments 1) the line number containing the error, 2) the sequence of tokens, the starting and ending line of the function containing the error, and array containing the lines of the file containing the error, and the pass of the parser on which the error occurs. This functyion should then be passed to the fuction `Parse_c.set_parse_error_function`.

◇ **--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.

--cache-prefix Specify the directory in which to store preparsed versions of the C files. This sets **--use-cache**

--cache-limit Specify the maximum number of preparsed C files to store. The cache is cleared of all files with names ending in .ast-raw and .depend-raw on reaching this limit. Only effective if **--cache-prefix** is used as well. This is most useful when iteration is used to process a file multiple times within a single run of Coccinelle.

--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.

- **--noif0-passing** 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.

- **--defined** *s* The string *s* is a comma-separated list of constants that should be considered to be defined, with respect to uses of `#ifdef` and `#ifndef` in C code. No spaces should appear in *s*. Multiple **--defined** arguments can be provided and the list of strings accumulates. For the provided strings any `#else`s of `#ifdef`s are ignored and any `#ifndef`s are ignored, unless the argument **--noif0-passing** is also given, in which case **--defined** has no effect. Note that occurrences of `#define` in the C code have no effect on the list of defined constants.

- **--undefined** *s* Analogous to **--defined** except that the strings represent constants that should be considered to be undefined.

--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. **--recursive-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 to specify multipls include paths.

- **--include-headers-for-types** Header files are parsed to collect type information, but are not involved in the subsequent matching and transformation process.

- **--include** ⟨**file**⟩ This option give the name of a file to consider as being included in each processed file. The file is added to the end of the file's list of included files. The complete path name should be given; the **-I** options are not taken into account to find the file. This option can be used several times to include multiple files.

- **--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**

- **--c++** Make an extremely minimal effort to parse C++ code. Currently, this is limited to allowing identifiers to contain "::", tilde, and template invocations. Consider testing your code first with spatch --type-c to see if there are any type annotations in the code you are interested in processing. If not, then it was probably not parsed.

- **--ibm** Make a effort to parse IBM C code. Currently decimal declarations are supported.

4 Application of the semantic patch to the C code

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.

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**.

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 satifaction 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.

◇ **--ocaml-regexps** Use the regular expressions provided by the OCaml `Str` library. This is the default if the PCRE library is not available. Otherwise PCRE regular expressions are used by default.

--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.

◇ **--reverse** Inverts the semantic patch before applying it. A potential use case is backporting changes to previous versions. If a semantic patch represents an API change, then the reverse undoes the API change. Note that inverting a semantic patch is not always possible. In particular, the composition of a semantic patch with its inverse is not guaranteed to be an empty patch.

5 Generation of the result

Normally, the only output is the differences between the original code and the transformed code obtained using the program `diff` with the unified format option. If stars are used in column 0 rather than `-` and `+`, then the `-` lines in the output are the lines that matched the stars.

◆ **--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.

◆ **--indent** *n* The number of spaces to indent, if no other information is available. If this information is not provided, then the default indentation is a tab. This option is thus particularly relevent to projects that don't use tabs.

◇ **-o** ⟨**file**⟩ This causes the transformed code to be placed in the file `file`. The difference between the original code and the transformed code is still printed to the standard output using `diff` with the unified format option. This option only makes sense when `-` and `+` are used.

◇ **--in-place** Modify the input file to contain the transformed code. The difference between the original code and the transformed code is still printed to the standard output using `diff` with the unified format option. By default, the input file is overwritten when using this option, with no backup. This option only makes sense when `-` and `+` are used.

◇ **--backup-suffix** *s* The suffix *s* 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 the result of modifying the code in a .cocci-res file. The difference between the original code and the transformed code is still printed to the standard output using `diff` with the unified format option. This option only makes sense when `-` and `+` are used.

◇ **--no-show-diff** Normally, the difference between the original and transformed code is printed on the standard output. This option causes this not to be done.

◇ **-U** Set number of context lines to be provided by `diff`.

◇ **--patch** ⟨**path**⟩ The prefix of the pathname of the directory or file name that should 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.

6 Other options

6.1 Version information

◆ **--version** The version of Coccinelle is printed on the standard output. No other options are allowed.

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

6.2 Help

◆ **--h, --shorthelp** The most useful commands.

◆ **--help, --help, --longhelp** A complete listing of the available commands.

6.3 Controlling the execution of Coccinelle

◆ **--timeout** ⟨**int**⟩ The maximum time in seconds for processing a single file. A timeout of 0 is no timeout.

◇ **--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.

6.4 Parallelism

◆ **--jobs** ⟨**int**⟩ Run the specified number of jobs in parallel. Can be abbreviated as **-j**. This option is not compatible with the use of an `initialize` or `finalize` rule in the semantic patch. This option furthermore creates a temporary directory in the directory from which spatch is executed that has the name of the semantic patch (without its extension) and that contains stdout and stderr files generated by the various processes. When the semantic patch completes, the contents of these files are printed to standard output and standard error, respectively, and the directory is removed.

--chunksize ⟨**int**⟩ The specified number of files are dispatched as a single unit of parallelism. This option is only interesting with the options **--all-includes** or **--recursive-includes**, when combined with the option **--include-headers-for-types**. In this case, parsed header files are cached. It is only the files that are treated within a single chunk that can benefit from this cache, due to the lack of shared memory in ocaml.

6.5 External analyses

--external-analysis-file Loads in the contents of a database produced by some external analysis tool. Each entry contains the analysis result of a particular source location. Currently, such a database is a .csv file providing integer bounds or an integer set for some subset of the source locations that references an integer memory location. This database can be inspected with coccilib functions, e.g. to control the pattern match process.

6.6 Miscellaneous

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

--pad, --xxx, --l1