Chapter 5

Profiling

GHC comes with a time and space profiling system, so that you can answer questions like "why is my program so slow?", or "why is my program using so much memory?".

Profiling a program is a three-step process:

1. Re-compile your program for profiling with the `-prof` option, and probably one of the options for adding automatic annotations: `-fprof-auto` is the most common[1].

   If you are using external packages with `cabal`, you may need to reinstall these packages with profiling support; typically this is done with `cabal install -p` *`package`* `--reinstall`.

2. Having compiled the program for profiling, you now need to run it to generate the profile. For example, a simple time profile can be generated by running the program with `+RTS -p`, which generates a file named *`prog`*`.prof` where *`prog`* is the name of your program (without the `.exe` extension, if you are on Windows).

   There are many different kinds of profile that can be generated, selected by different RTS options. We will be describing the various kinds of profile throughout the rest of this chapter. Some profiles require further processing using additional tools after running the program.

3. Examine the generated profiling information, use the information to optimise your program, and repeat as necessary.

5.1 Cost centres and cost-centre stacks

GHC's profiling system assigns *costs* to *cost centres*. A cost is simply the time or space (memory) required to evaluate an expression. Cost centres are program annotations around expressions; all costs incurred by the annotated expression are assigned to the enclosing cost centre. Furthermore, GHC will remember the stack of enclosing cost centres for any given expression at run-time and generate a call-tree of cost attributions.

Let's take a look at an example:

```
main = print (fib 30)
fib n = if n < 2 then 1 else fib (n-1) + fib (n-2)
```

Compile and run this program as follows:

```
$ ghc -prof -fprof-auto -rtsopts Main.hs
$ ./Main +RTS -p
121393
$
```

When a GHC-compiled program is run with the `-p` RTS option, it generates a file called *`prog`*`.prof`. In this case, the file will contain something like this:

[1] `-fprof-auto` was known as `-auto-all` prior to GHC 7.4.1.

```
        Wed Oct 12 16:14 2011 Time and Allocation Profiling Report  (Final)

           Main +RTS -p -RTS

        total time  =        0.68 secs   (34 ticks @ 20 ms)
        total alloc = 204,677,844 bytes  (excludes profiling overheads)

COST CENTRE MODULE  %time %alloc

fib         Main    100.0  100.0


                                                  individual     inherited
COST CENTRE MODULE                  no.     entries  %time %alloc   %time %alloc

MAIN        MAIN                    102           0    0.0    0.0   100.0  100.0
 CAF        GHC.IO.Handle.FD        128           0    0.0    0.0     0.0    0.0
 CAF        GHC.IO.Encoding.Iconv   120           0    0.0    0.0     0.0    0.0
 CAF        GHC.Conc.Signal         110           0    0.0    0.0     0.0    0.0
 CAF        Main                    108           0    0.0    0.0   100.0  100.0
  main      Main                    204           1    0.0    0.0   100.0  100.0
   fib      Main                    205     2692537  100.0  100.0   100.0  100.0
```

The first part of the file gives the program name and options, and the total time and total memory allocation measured during the run of the program (note that the total memory allocation figure isn't the same as the amount of *live* memory needed by the program at any one time; the latter can be determined using heap profiling, which we will describe later in Section 5.4).

The second part of the file is a break-down by cost centre of the most costly functions in the program. In this case, there was only one significant function in the program, namely `fib`, and it was responsible for 100% of both the time and allocation costs of the program.

The third and final section of the file gives a profile break-down by cost-centre stack. This is roughly a call-tree profile of the program. In the example above, it is clear that the costly call to `fib` came from `main`.

The time and allocation incurred by a given part of the program is displayed in two ways: "individual", which are the costs incurred by the code covered by this cost centre stack alone, and "inherited", which includes the costs incurred by all the children of this node.

The usefulness of cost-centre stacks is better demonstrated by modifying the example slightly:

```
main = print (f 30 + g 30)
  where
    f n  = fib n
    g n  = fib (n `div` 2)

fib n = if n < 2 then 1 else fib (n-1) + fib (n-2)
```

Compile and run this program as before, and take a look at the new profiling results:

```
COST CENTRE MODULE                  no.     entries  %time %alloc   %time %alloc

MAIN        MAIN                    102           0    0.0    0.0   100.0  100.0
 CAF        GHC.IO.Handle.FD        128           0    0.0    0.0     0.0    0.0
 CAF        GHC.IO.Encoding.Iconv   120           0    0.0    0.0     0.0    0.0
 CAF        GHC.Conc.Signal         110           0    0.0    0.0     0.0    0.0
 CAF        Main                    108           0    0.0    0.0   100.0  100.0
  main      Main                    204           1    0.0    0.0   100.0  100.0
   main.g   Main                    207           1    0.0    0.0     0.0    0.1
    fib     Main                    208        1973    0.0    0.1     0.0    0.1
   main.f   Main                    205           1    0.0    0.0   100.0   99.9
    fib     Main                    206     2692537  100.0   99.9   100.0   99.9
```

Now although we had two calls to `fib` in the program, it is immediately clear that it was the call from `f` which took all the time. The functions `f` and `g` which are defined in the `where` clause in `main` are given their own cost centres, `main.f` and `main.g` respectively.

The actual meaning of the various columns in the output is:

entries The number of times this particular point in the call tree was entered.

individual %time The percentage of the total run time of the program spent at this point in the call tree.

individual %alloc The percentage of the total memory allocations (excluding profiling overheads) of the program made by this call.

inherited %time The percentage of the total run time of the program spent below this point in the call tree.

inherited %alloc The percentage of the total memory allocations (excluding profiling overheads) of the program made by this call and all of its sub-calls.

In addition you can use the `-P` RTS option to get the following additional information:

`ticks` The raw number of time "ticks" which were attributed to this cost-centre; from this, we get the `%time` figure mentioned above.

`bytes` Number of bytes allocated in the heap while in this cost-centre; again, this is the raw number from which we get the `%alloc` figure mentioned above.

What about recursive functions, and mutually recursive groups of functions? Where are the costs attributed? Well, although GHC does keep information about which groups of functions called each other recursively, this information isn't displayed in the basic time and allocation profile, instead the call-graph is flattened into a tree as follows: a call to a function that occurs elsewhere on the current stack does not push another entry on the stack, instead the costs for this call are aggregated into the caller[2].

5.1.1 Inserting cost centres by hand

Cost centres are just program annotations. When you say `-fprof-auto` to the compiler, it automatically inserts a cost centre annotation around every binding not marked INLINE in your program, but you are entirely free to add cost centre annotations yourself.

The syntax of a cost centre annotation is

```
{-# SCC "name" #-} <expression>
```

where `"name"` is an arbitrary string, that will become the name of your cost centre as it appears in the profiling output, and `<expression>` is any Haskell expression. An `SCC` annotation extends as far to the right as possible when parsing. (SCC stands for "Set Cost Centre"). The double quotes can be omitted if `name` is a Haskell identifier, for example:

```
{-# SCC my_function #-} <expression>
```

Here is an example of a program with a couple of SCCs:

```
main :: IO ()
main = do let xs = [1..1000000]
          let ys = [1..2000000]
          print $ {-# SCC last_xs #-} last xs
          print $ {-# SCC last_init_xs #-} last $ init xs
          print $ {-# SCC last_ys #-} last ys
          print $ {-# SCC last_init_ys #-}last $ init ys
```

which gives this profile when run:

[2] Note that this policy has changed slightly in GHC 7.4.1 relative to earlier versions, and may yet change further, feedback is welcome.

```
COST CENTRE     MODULE                  no.     entries  %time %alloc   %time %alloc

MAIN            MAIN                    102           0    0.0    0.0   100.0  100.0
 CAF            GHC.IO.Handle.FD        130           0    0.0    0.0     0.0    0.0
 CAF            GHC.IO.Encoding.Iconv   122           0    0.0    0.0     0.0    0.0
 CAF            GHC.Conc.Signal         111           0    0.0    0.0     0.0    0.0
 CAF            Main                    108           0    0.0    0.0   100.0  100.0
  main          Main                    204           1    0.0    0.0   100.0  100.0
   last_init_ys Main                    210           1   25.0   27.4    25.0   27.4
   main.ys      Main                    209           1   25.0   39.2    25.0   39.2
   last_ys      Main                    208           1   12.5    0.0    12.5    0.0
   last_init_xs Main                    207           1   12.5   13.7    12.5   13.7
   main.xs      Main                    206           1   18.8   19.6    18.8   19.6
   last_xs      Main                    205           1    6.2    0.0     6.2    0.0
```

5.1.2 Rules for attributing costs

While running a program with profiling turned on, GHC maintains a cost-centre stack behind the scenes, and attributes any costs (memory allocation and time) to whatever the current cost-centre stack is at the time the cost is incurred.

The mechanism is simple: whenever the program evaluates an expression with an SCC annotation, `{-#SCC c -#} E`, the cost centre `c` is pushed on the current stack, and the entry count for this stack is incremented by one. The stack also sometimes has to be saved and restored; in particular when the program creates a *thunk* (a lazy suspension), the current cost-centre stack is stored in the thunk, and restored when the thunk is evaluated. In this way, the cost-centre stack is independent of the actual evaluation order used by GHC at runtime.

At a function call, GHC takes the stack stored in the function being called (which for a top-level function will be empty), and *appends* it to the current stack, ignoring any prefix that is identical to a prefix of the current stack.

We mentioned earlier that lazy computations, i.e. thunks, capture the current stack when they are created, and restore this stack when they are evaluated. What about top-level thunks? They are "created" when the program is compiled, so what stack should we give them? The technical name for a top-level thunk is a CAF ("Constant Applicative Form"). GHC assigns every CAF in a module a stack consisting of the single cost centre `M.CAF`, where `M` is the name of the module. It is also possible to give each CAF a different stack, using the option `-fprof-cafs`. This is especially useful when compiling with `-ffull-laziness` (as is default with `-O` and higher), as constants in function bodies will be lifted to the top-level and become CAFs. You will probably need to consult the Core (`-ddump-simpl`) in order to determine what these CAFs correspond to.

5.2 Compiler options for profiling

`-prof`: To make use of the profiling system *all* modules must be compiled and linked with the `-prof` option. Any `SCC` annotations you've put in your source will spring to life.

Without a `-prof` option, your `SCC`s are ignored; so you can compile `SCC`-laden code without changing it.

There are a few other profiling-related compilation options. Use them *in addition to* `-prof`. These do not have to be used consistently for all modules in a program.

`-fprof-auto`: *All* bindings not marked INLINE, whether exported or not, top level or nested, will be given automatic `SCC` annotations. Functions marked INLINE must be given a cost centre manually.

`-fprof-auto-top`: GHC will automatically add `SCC` annotations for all top-level bindings not marked INLINE. If you want a cost centre on an INLINE function, you have to add it manually.

`-fprof-auto-exported`: GHC will automatically add `SCC` annotations for all exported functions not marked INLINE. If you want a cost centre on an INLINE function, you have to add it manually.

`-fprof-auto-calls`: Adds an automatic `SCC` annotation to all *call sites*. This is particularly useful when using profiling for the purposes of generating stack traces; see the function `traceStack` in the module `Debug.Trace`, or the `-xc` RTS flag (Section 4.17.7) for more details.

`-fprof-cafs`: The costs of all CAFs in a module are usually attributed to one "big" CAF cost-centre. With this option, all CAFs get their own cost-centre. An "if all else fails" option...

`-fno-prof-auto`: Disables any previous `-fprof-auto`, `-fprof-auto-top`, or `-fprof-auto-exported` options.

`-fno-prof-cafs`: Disables any previous `-fprof-cafs` option.

`-fno-prof-count-entries`: Tells GHC not to collect information about how often functions are entered at runtime (the "entries" column of the time profile), for this module. This tends to make the profiled code run faster, and hence closer to the speed of the unprofiled code, because GHC is able to optimise more aggressively if it doesn't have to maintain correct entry counts. This option can be useful if you aren't interested in the entry counts (for example, if you only intend to do heap profiling).

5.3 Time and allocation profiling

To generate a time and allocation profile, give one of the following RTS options to the compiled program when you run it (RTS options should be enclosed between `+RTS...-RTS` as usual):

`-p` or `-P` or `-pa`: The `-p` option produces a standard *time profile* report. It is written into the file *`program`*`.prof`.

The `-P` option produces a more detailed report containing the actual time and allocation data as well. (Not used much.)

The `-pa` option produces the most detailed report containing all cost centres in addition to the actual time and allocation data.

`-V`*`secs`* Sets the interval that the RTS clock ticks at, which is also the sampling interval of the time and allocation profile. The default is 0.02 seconds.

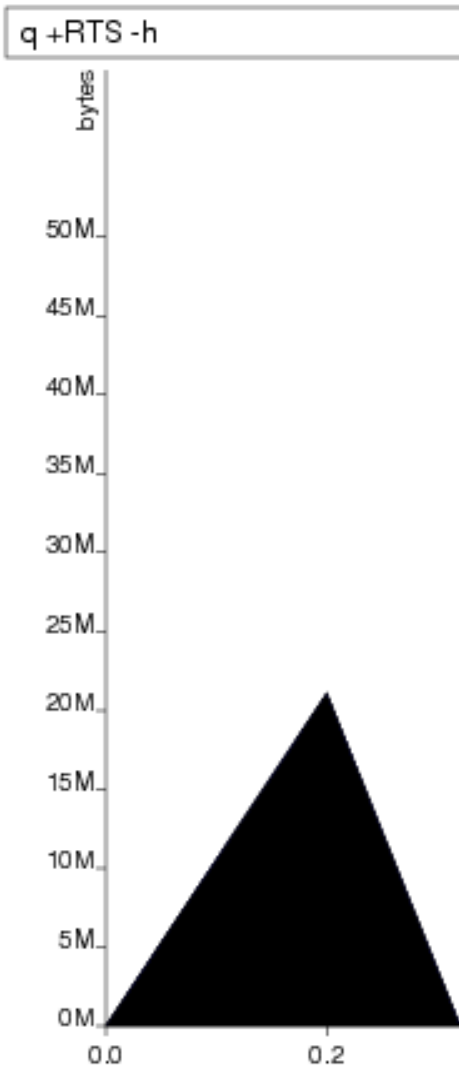
`-xc` This option causes the runtime to print out the current cost-centre stack whenever an exception is raised. This can be particularly useful for debugging the location of exceptions, such as the notorious `Prelude.head:empty list` error. See Section 4.17.7.

5.4 Profiling memory usage

In addition to profiling the time and allocation behaviour of your program, you can also generate a graph of its memory usage over time. This is useful for detecting the causes of *space leaks*, when your program holds on to more memory at run-time that it needs to. Space leaks lead to slower execution due to heavy garbage collector activity, and may even cause the program to run out of memory altogether.

To generate a heap profile from your program:

1. Compile the program for profiling (Section 5.2).
2. Run it with one of the heap profiling options described below (eg. `-h` for a basic producer profile). This generates the file *`prog`*`.hp`.
3. Run **hp2ps** to produce a Postscript file, *`prog`*`.ps`. The **hp2ps** utility is described in detail in Section 5.5.
4. Display the heap profile using a postscript viewer such as Ghostview, or print it out on a Postscript-capable printer.



For example, here is a heap profile produced for the program given above in Section 5.1.1:
You might also want to take a look at hp2any, a more advanced suite of tools (not distributed with GHC) for displaying heap profiles.

5.4.1 RTS options for heap profiling

There are several different kinds of heap profile that can be generated. All the different profile types yield a graph of live heap against time, but they differ in how the live heap is broken down into bands. The following RTS options select which break-down to use:

-hc (can be shortened to `-h`). Breaks down the graph by the cost-centre stack which produced the data.

-hm Break down the live heap by the module containing the code which produced the data.

-hd Breaks down the graph by *closure description*. For actual data, the description is just the constructor name, for other closures it is a compiler-generated string identifying the closure.

-hy Breaks down the graph by *type*. For closures which have function type or unknown/polymorphic type, the string will represent an approximation to the actual type.

-hr Break down the graph by *retainer set*. Retainer profiling is described in more detail below (Section 5.4.2).

-hb Break down the graph by *biography*. Biographical profiling is described in more detail below (Section 5.4.3).

In addition, the profile can be restricted to heap data which satisfies certain criteria - for example, you might want to display a profile by type but only for data produced by a certain module, or a profile by retainer for a certain type of data. Restrictions are specified as follows:

-hc*name*,... Restrict the profile to closures produced by cost-centre stacks with one of the specified cost centres at the top.

-hC*name*,... Restrict the profile to closures produced by cost-centre stacks with one of the specified cost centres anywhere in the stack.

-hm*module*,... Restrict the profile to closures produced by the specified modules.

-hd*desc*,... Restrict the profile to closures with the specified description strings.

-hy*type*,... Restrict the profile to closures with the specified types.

-hr*cc*,... Restrict the profile to closures with retainer sets containing cost-centre stacks with one of the specified cost centres at the top.

-hb*bio*,... Restrict the profile to closures with one of the specified biographies, where *bio* is one of `lag`, `drag`, `void`, or `use`.

For example, the following options will generate a retainer profile restricted to `Branch` and `Leaf` constructors:

```
prog +RTS -hr -hdBranch,Leaf
```

There can only be one "break-down" option (eg. `-hr` in the example above), but there is no limit on the number of further restrictions that may be applied. All the options may be combined, with one exception: GHC doesn't currently support mixing the `-hr` and `-hb` options.

There are three more options which relate to heap profiling:

-i*secs*: Set the profiling (sampling) interval to *secs* seconds (the default is 0.1 second). Fractions are allowed: for example `-i0.2` will get 5 samples per second. This only affects heap profiling; time profiles are always sampled with the frequency of the RTS clock. See Section 5.3 for changing that.

-xt Include the memory occupied by threads in a heap profile. Each thread takes up a small area for its thread state in addition to the space allocated for its stack (stacks normally start small and then grow as necessary).

This includes the main thread, so using `-xt` is a good way to see how much stack space the program is using.

Memory occupied by threads and their stacks is labelled as "TSO" and "STACK" respectively when displaying the profile by closure description or type description.

-L*num* Sets the maximum length of a cost-centre stack name in a heap profile. Defaults to 25.

5.4.2 Retainer Profiling

Retainer profiling is designed to help answer questions like "why is this data being retained?". We start by defining what we mean by a retainer:

> A retainer is either the system stack, an unevaluated closure (thunk), or an explicitly mutable object.

In particular, constructors are *not* retainers.

An object B retains object A if (i) B is a retainer object and (ii) object A can be reached by recursively following pointers starting from object B, but not meeting any other retainer objects on the way. Each live object is retained by one or more retainer objects, collectively called its retainer set, or its *retainer set*, or its *retainers*.

When retainer profiling is requested by giving the program the `-hr` option, a graph is generated which is broken down by retainer set. A retainer set is displayed as a set of cost-centre stacks; because this is usually too large to fit on the profile graph, each retainer set is numbered and shown abbreviated on the graph along with its number, and the full list of retainer sets is dumped into the file *prog*`.prof`.

Retainer profiling requires multiple passes over the live heap in order to discover the full retainer set for each object, which can be quite slow. So we set a limit on the maximum size of a retainer set, where all retainer sets larger than the maximum retainer set size are replaced by the special set `MANY`. The maximum set size defaults to 8 and can be altered with the `-R` RTS option:

`-R`*`size`* Restrict the number of elements in a retainer set to *`size`* (default 8).

5.4.2.1 Hints for using retainer profiling

The definition of retainers is designed to reflect a common cause of space leaks: a large structure is retained by an unevaluated computation, and will be released once the computation is forced. A good example is looking up a value in a finite map, where unless the lookup is forced in a timely manner the unevaluated lookup will cause the whole mapping to be retained. These kind of space leaks can often be eliminated by forcing the relevant computations to be performed eagerly, using `seq` or strictness annotations on data constructor fields.

Often a particular data structure is being retained by a chain of unevaluated closures, only the nearest of which will be reported by retainer profiling - for example A retains B, B retains C, and C retains a large structure. There might be a large number of Bs but only a single A, so A is really the one we're interested in eliminating. However, retainer profiling will in this case report B as the retainer of the large structure. To move further up the chain of retainers, we can ask for another retainer profile but this time restrict the profile to B objects, so we get a profile of the retainers of B:

```
prog +RTS -hr -hcB
```

This trick isn't foolproof, because there might be other B closures in the heap which aren't the retainers we are interested in, but we've found this to be a useful technique in most cases.

5.4.3 Biographical Profiling

A typical heap object may be in one of the following four states at each point in its lifetime:

- The *lag* stage, which is the time between creation and the first use of the object,
- the *use* stage, which lasts from the first use until the last use of the object, and
- The *drag* stage, which lasts from the final use until the last reference to the object is dropped.
- An object which is never used is said to be in the *void* state for its whole lifetime.

A biographical heap profile displays the portion of the live heap in each of the four states listed above. Usually the most interesting states are the void and drag states: live heap in these states is more likely to be wasted space than heap in the lag or use states.

It is also possible to break down the heap in one or more of these states by a different criteria, by restricting a profile by biography. For example, to show the portion of the heap in the drag or void state by producer:

```
prog +RTS -hc -hbdrag,void
```

Once you know the producer or the type of the heap in the drag or void states, the next step is usually to find the retainer(s):

```
prog +RTS -hr -hccc...
```

NOTE: this two stage process is required because GHC cannot currently profile using both biographical and retainer information simultaneously.

5.4.4 Actual memory residency

How does the heap residency reported by the heap profiler relate to the actual memory residency of your program when you run it? You might see a large discrepancy between the residency reported by the heap profiler, and the residency reported by tools on your system (eg. `ps` or `top` on Unix, or the Task Manager on Windows). There are several reasons for this:

- There is an overhead of profiling itself, which is subtracted from the residency figures by the profiler. This overhead goes away when compiling without profiling support, of course. The space overhead is currently 2 extra words per heap object, which probably results in about a 30% overhead.
- Garbage collection requires more memory than the actual residency. The factor depends on the kind of garbage collection algorithm in use: a major GC in the standard generation copying collector will usually require 3L bytes of memory, where L is the amount of live data. This is because by default (see the `+RTS -F` option) we allow the old generation to grow to twice its size (2L) before collecting it, and we require additionally L bytes to copy the live data into. When using compacting collection (see the `+RTS -c` option), this is reduced to 2L, and can further be reduced by tweaking the `-F` option. Also add the size of the allocation area (currently a fixed 512Kb).
- The stack isn't counted in the heap profile by default. See the `+RTS -xt` option.
- The program text itself, the C stack, any non-heap data (eg. data allocated by foreign libraries, and data allocated by the RTS), and `mmap()`'d memory are not counted in the heap profile.

5.5 hp2ps--heap profile to PostScript

Usage:

```
hp2ps [flags] [<file>[.hp]]
```

The program **hp2ps** converts a heap profile as produced by the `-h<break-down>` runtime option into a PostScript graph of the heap profile. By convention, the file to be processed by **hp2ps** has a `.hp` extension. The PostScript output is written to `<file>@.ps`. If `<file>` is omitted entirely, then the program behaves as a filter.

hp2ps is distributed in `ghc/utils/hp2ps` in a GHC source distribution. It was originally developed by Dave Wakeling as part of the HBC/LML heap profiler.

The flags are:

`-d` In order to make graphs more readable, **hp2ps** sorts the shaded bands for each identifier. The default sort ordering is for the bands with the largest area to be stacked on top of the smaller ones. The `-d` option causes rougher bands (those representing series of values with the largest standard deviations) to be stacked on top of smoother ones.

`-b` Normally, **hp2ps** puts the title of the graph in a small box at the top of the page. However, if the JOB string is too long to fit in a small box (more than 35 characters), then **hp2ps** will choose to use a big box instead. The `-b` option forces **hp2ps** to use a big box.

`-e<float>[in|mm|pt]` Generate encapsulated PostScript suitable for inclusion in LaTeX documents. Usually, the PostScript graph is drawn in landscape mode in an area 9 inches wide by 6 inches high, and **hp2ps** arranges for this area to be approximately centred on a sheet of a4 paper. This format is convenient of studying the graph in detail, but it is unsuitable for inclusion in LaTeX documents. The `-e` option causes the graph to be drawn in portrait mode, with float specifying the width in inches, millimetres or points (the default). The resulting PostScript file conforms to the Encapsulated PostScript (EPS) convention, and it can be included in a LaTeX document using Rokicki's dvi-to-PostScript converter **dvips**.

`-g` Create output suitable for the **gs** PostScript previewer (or similar). In this case the graph is printed in portrait mode without scaling. The output is unsuitable for a laser printer.

`-l` Normally a profile is limited to 20 bands with additional identifiers being grouped into an `OTHER` band. The `-l` flag removes this 20 band and limit, producing as many bands as necessary. No key is produced as it won't fit!. It is useful for creation time profiles with many bands.

-m<int> Normally a profile is limited to 20 bands with additional identifiers being grouped into an `OTHER` band. The `-m` flag specifies an alternative band limit (the maximum is 20).

`-m0` requests the band limit to be removed. As many bands as necessary are produced. However no key is produced as it won't fit! It is useful for displaying creation time profiles with many bands.

-p Use previous parameters. By default, the PostScript graph is automatically scaled both horizontally and vertically so that it fills the page. However, when preparing a series of graphs for use in a presentation, it is often useful to draw a new graph using the same scale, shading and ordering as a previous one. The `-p` flag causes the graph to be drawn using the parameters determined by a previous run of **hp2ps** on `file`. These are extracted from `file@.aux`.

-s Use a small box for the title.

-t<float> Normally trace elements which sum to a total of less than 1% of the profile are removed from the profile. The `-t` option allows this percentage to be modified (maximum 5%).

`-t0` requests no trace elements to be removed from the profile, ensuring that all the data will be displayed.

-c Generate colour output.

-y Ignore marks.

-? Print out usage information.

5.5.1 Manipulating the hp file

(Notes kindly offered by Jan-Willem Maessen.)

The `FOO.hp` file produced when you ask for the heap profile of a program `FOO` is a text file with a particularly simple structure. Here's a representative example, with much of the actual data omitted:

```
JOB "FOO -hC"
DATE "Thu Dec 26 18:17 2002"
SAMPLE_UNIT "seconds"
VALUE_UNIT "bytes"
BEGIN_SAMPLE 0.00
END_SAMPLE 0.00
BEGIN_SAMPLE 15.07
  ... sample data ...
END_SAMPLE 15.07
BEGIN_SAMPLE 30.23
  ... sample data ...
END_SAMPLE 30.23
... etc.
BEGIN_SAMPLE 11695.47
END_SAMPLE 11695.47
```

The first four lines (`JOB`, `DATE`, `SAMPLE_UNIT`, `VALUE_UNIT`) form a header. Each block of lines starting with `BEGIN_SAMPLE` and ending with `END_SAMPLE` forms a single sample (you can think of this as a vertical slice of your heap profile). The hp2ps utility should accept any input with a properly-formatted header followed by a series of *complete* samples.

5.5.2 Zooming in on regions of your profile

You can look at particular regions of your profile simply by loading a copy of the `.hp` file into a text editor and deleting the unwanted samples. The resulting `.hp` file can be run through **hp2ps** and viewed or printed.

5.5.3 Viewing the heap profile of a running program

The `.hp` file is generated incrementally as your program runs. In principle, running **hp2ps** on the incomplete file should produce a snapshot of your program's heap usage. However, the last sample in the file may be incomplete, causing **hp2ps** to fail. If you are using a machine with UNIX utilities installed, it's not too hard to work around this problem (though the resulting command line looks rather Byzantine):

```
head -`fgrep -n END_SAMPLE FOO.hp | tail -1 | cut -d : -f 1` FOO.hp \
  | hp2ps > FOO.ps
```

The command **fgrep -n END_SAMPLE FOO.hp** finds the end of every complete sample in `FOO.hp`, and labels each sample with its ending line number. We then select the line number of the last complete sample using **tail** and **cut**. This is used as a parameter to **head**; the result is as if we deleted the final incomplete sample from `FOO.hp`. This results in a properly-formatted .hp file which we feed directly to **hp2ps**.

5.5.4 Viewing a heap profile in real time

The **gv** and **ghostview** programs have a "watch file" option can be used to view an up-to-date heap profile of your program as it runs. Simply generate an incremental heap profile as described in the previous section. Run **gv** on your profile:

```
gv -watch -seascape FOO.ps
```

If you forget the `-watch` flag you can still select "Watch file" from the "State" menu. Now each time you generate a new profile `FOO.ps` the view will update automatically.

This can all be encapsulated in a little script:

```
#!/bin/sh
head -`fgrep -n END_SAMPLE FOO.hp | tail -1 | cut -d : -f 1` FOO.hp \
  | hp2ps > FOO.ps
gv -watch -seascape FOO.ps &
while [ 1 ] ; do
  sleep 10 # We generate a new profile every 10 seconds.
  head -`fgrep -n END_SAMPLE FOO.hp | tail -1 | cut -d : -f 1` FOO.hp \
    | hp2ps > FOO.ps
done
```

Occasionally **gv** will choke as it tries to read an incomplete copy of `FOO.ps` (because **hp2ps** is still running as an update occurs). A slightly more complicated script works around this problem, by using the fact that sending a SIGHUP to gv will cause it to re-read its input file:

```
#!/bin/sh
head -`fgrep -n END_SAMPLE FOO.hp | tail -1 | cut -d : -f 1` FOO.hp \
  | hp2ps > FOO.ps
gv FOO.ps &
gvpsnum=$!
while [ 1 ] ; do
  sleep 10
  head -`fgrep -n END_SAMPLE FOO.hp | tail -1 | cut -d : -f 1` FOO.hp \
    | hp2ps > FOO.ps
  kill -HUP $gvpsnum
done
```

5.6 Profiling Parallel and Concurrent Programs

Combining `-threaded` and `-prof` is perfectly fine, and indeed it is possible to profile a program running on multiple processors with the `+RTS -N` option.[3]

[3] This feature was added in GHC 7.4.1.

Some caveats apply, however. In the current implementation, a profiled program is likely to scale much less well than the unprofiled program, because the profiling implementation uses some shared data structures which require locking in the runtime system. Furthermore, the memory allocation statistics collected by the profiled program are stored in shared memory but *not* locked (for speed), which means that these figures might be inaccurate for parallel programs.

We strongly recommend that you use `-fno-prof-count-entries` when compiling a program to be profiled on multiple cores, because the entry counts are also stored in shared memory, and continuously updating them on multiple cores is extremely slow.

We also recommend using ThreadScope for profiling parallel programs; it offers a GUI for visualising parallel execution, and is complementary to the time and space profiling features provided with GHC.

5.7 Observing Code Coverage

Code coverage tools allow a programmer to determine what parts of their code have been actually executed, and which parts have never actually been invoked. GHC has an option for generating instrumented code that records code coverage as part of the Haskell Program Coverage (HPC) toolkit, which is included with GHC. HPC tools can be used to render the generated code coverage information into human understandable format.

Correctly instrumented code provides coverage information of two kinds: source coverage and boolean-control coverage. Source coverage is the extent to which every part of the program was used, measured at three different levels: declarations (both top-level and local), alternatives (among several equations or case branches) and expressions (at every level). Boolean coverage is the extent to which each of the values True and False is obtained in every syntactic boolean context (ie. guard, condition, qualifier).

HPC displays both kinds of information in two primary ways: textual reports with summary statistics (`hpc report`) and sources with color mark-up (`hpc markup`). For boolean coverage, there are four possible outcomes for each guard, condition or qualifier: both True and False values occur; only True; only False; never evaluated. In hpc-markup output, highlighting with a yellow background indicates a part of the program that was never evaluated; a green background indicates an always-True expression and a red background indicates an always-False one.

5.7.1 A small example: Reciprocation

For an example we have a program, called `Recip.hs`, which computes exact decimal representations of reciprocals, with recurring parts indicated in brackets.

```
reciprocal :: Int -> (String, Int)
reciprocal n | n > 1 = ('0' : '.' : digits, recur)
             | otherwise = error
            "attempting to compute reciprocal of number <= 1"
  where
  (digits, recur) = divide n 1 []
divide :: Int -> Int -> [Int] -> (String, Int)
divide n c cs | c `elem` cs = ([], position c cs)
              | r == 0      = (show q, 0)
              | r /= 0      = (show q ++ digits, recur)
  where
  (q, r) = (c*10) `quotRem` n
  (digits, recur) = divide n r (c:cs)

position :: Int -> [Int] -> Int
position n (x:xs) | n==x      = 1
                  | otherwise = 1 + position n xs

showRecip :: Int -> String
showRecip n =
  "1/" ++ show n ++ " = " ++
  if r==0 then d else take p d ++ "(" ++ drop p d ++ ")"
  where
  p = length d - r
```

```
  (d, r) = reciprocal n

main = do
  number <- readLn
  putStrLn (showRecip number)
  main
```

HPC instrumentation is enabled with the -fhpc flag:

```
$ ghc -fhpc Recip.hs
```

GHC creates a subdirectory `.hpc` in the current directory, and puts HPC index (`.mix`) files in there, one for each module compiled. You don't need to worry about these files: they contain information needed by the `hpc` tool to generate the coverage data for compiled modules after the program is run.

```
$ ./Recip
1/3
= 0.(3)
```

Running the program generates a file with the `.tix` suffix, in this case `Recip.tix`, which contains the coverage data for this run of the program. The program may be run multiple times (e.g. with different test data), and the coverage data from the separate runs is accumulated in the `.tix` file. To reset the coverage data and start again, just remove the `.tix` file.

Having run the program, we can generate a textual summary of coverage:

```
$ hpc report Recip
 80% expressions used (81/101)
 12% boolean coverage (1/8)
      14% guards (1/7), 3 always True,
                        1 always False,
                        2 unevaluated
       0% 'if' conditions (0/1), 1 always False
     100% qualifiers (0/0)
 55% alternatives used (5/9)
100% local declarations used (9/9)
100% top-level declarations used (5/5)
```

We can also generate a marked-up version of the source.

```
$ hpc markup Recip
writing Recip.hs.html
```

This generates one file per Haskell module, and 4 index files, hpc_index.html, hpc_index_alt.html, hpc_index_exp.html, hpc_index_fun.h

5.7.2 Options for instrumenting code for coverage

`-fhpc` Enable code coverage for the current module or modules being compiled.

Modules compiled with this option can be freely mixed with modules compiled without it; indeed, most libraries will typically be compiled without `-fhpc`. When the program is run, coverage data will only be generated for those modules that were compiled with `-fhpc`, and the `hpc` tool will only show information about those modules.

5.7.3 The hpc toolkit

The hpc command has several sub-commands:

```
$ hpc
Usage: hpc COMMAND ...

Commands:
  help        Display help for hpc or a single command
Reporting Coverage:
  report      Output textual report about program coverage
  markup      Markup Haskell source with program coverage
Processing Coverage files:
  sum         Sum multiple .tix files in a single .tix file
  combine     Combine two .tix files in a single .tix file
  map         Map a function over a single .tix file
Coverage Overlays:
  overlay     Generate a .tix file from an overlay file
  draft       Generate draft overlay that provides 100% coverage
Others:
  show        Show .tix file in readable, verbose format
  version     Display version for hpc
```

In general, these options act on a `.tix` file after an instrumented binary has generated it.

The hpc tool assumes you are in the top-level directory of the location where you built your application, and the `.tix` file is in the same top-level directory. You can use the flag `--srcdir` to use `hpc` for any other directory, and use `--srcdir` multiple times to analyse programs compiled from difference locations, as is typical for packages.

We now explain in more details the major modes of hpc.

5.7.3.1 hpc report

`hpc report` gives a textual report of coverage. By default, all modules and packages are considered in generating report, unless include or exclude are used. The report is a summary unless the `--per-module` flag is used. The `--xml-output` option allows for tools to use hpc to glean coverage.

```
$ hpc help report
Usage: hpc report [OPTION] .. <TIX_FILE> [<MODULE> [<MODULE> ..]]

Options:

    --per-module                  show module level detail
    --decl-list                   show unused decls
    --exclude=[PACKAGE:][MODULE]  exclude MODULE and/or PACKAGE
    --include=[PACKAGE:][MODULE]  include MODULE and/or PACKAGE
    --srcdir=DIR                  path to source directory of .hs files
                                  multi-use of srcdir possible
    --hpcdir=DIR                  append sub-directory that contains .mix files
                                  default .hpc [rarely used]
    --reset-hpcdirs               empty the list of hpcdir's
                                  [rarely used]
    --xml-output                  show output in XML
```

5.7.3.2 hpc markup

`hpc markup` marks up source files into colored html.

```
$ hpc help markup
Usage: hpc markup [OPTION] .. <TIX_FILE> [<MODULE> [<MODULE> ..]]

Options:
```

```
    --exclude=[PACKAGE:][MODULE]  exclude MODULE and/or PACKAGE
    --include=[PACKAGE:][MODULE]  include MODULE and/or PACKAGE
    --srcdir=DIR                  path to source directory of .hs files
                                  multi-use of srcdir possible
    --hpcdir=DIR                  append sub-directory that contains .mix files
                                  default .hpc [rarely used]
    --reset-hpcdirs               empty the list of hpcdir's
                                  [rarely used]
    --fun-entry-count             show top-level function entry counts
    --highlight-covered           highlight covered code, rather that code gaps
    --destdir=DIR                 path to write output to
```

5.7.3.3 hpc sum

`hpc sum` adds together any number of `.tix` files into a single `.tix` file. `hpc sum` does not change the original `.tix` file; it generates a new `.tix` file.

```
$ hpc help sum
Usage: hpc sum [OPTION] .. <TIX_FILE> [<TIX_FILE> [<TIX_FILE> ..]]
Sum multiple .tix files in a single .tix file

Options:

    --exclude=[PACKAGE:][MODULE]  exclude MODULE and/or PACKAGE
    --include=[PACKAGE:][MODULE]  include MODULE and/or PACKAGE
    --output=FILE                 output FILE
    --union                       use the union of the module namespace (default is ↩
        intersection)
```

5.7.3.4 hpc combine

`hpc combine` is the swiss army knife of `hpc`. It can be used to take the difference between `.tix` files, to subtract one `.tix` file from another, or to add two `.tix` files. hpc combine does not change the original `.tix` file; it generates a new `.tix` file.

```
$ hpc help combine
Usage: hpc combine [OPTION] .. <TIX_FILE> <TIX_FILE>
Combine two .tix files in a single .tix file

Options:

    --exclude=[PACKAGE:][MODULE]  exclude MODULE and/or PACKAGE
    --include=[PACKAGE:][MODULE]  include MODULE and/or PACKAGE
    --output=FILE                 output FILE
    --function=FUNCTION           combine .tix files with join function, default = ADD
                                  FUNCTION = ADD | DIFF | SUB
    --union                       use the union of the module namespace (default is ↩
        intersection)
```

5.7.3.5 hpc map

hpc map inverts or zeros a `.tix` file. hpc map does not change the original `.tix` file; it generates a new `.tix` file.

```
$ hpc help map
Usage: hpc map [OPTION] .. <TIX_FILE>
Map a function over a single .tix file

Options:
```

```
    --exclude=[PACKAGE:][MODULE]  exclude MODULE and/or PACKAGE
    --include=[PACKAGE:][MODULE]  include MODULE and/or PACKAGE
    --output=FILE                 output FILE
    --function=FUNCTION           apply function to .tix files, default = ID
                                  FUNCTION = ID | INV | ZERO
    --union                       use the union of the module namespace (default is  ↩
        intersection)
```

5.7.3.6 hpc overlay and hpc draft

Overlays are an experimental feature of HPC, a textual description of coverage. hpc draft is used to generate a draft overlay from a .tix file, and hpc overlay generates a .tix files from an overlay.

```
% hpc help overlay
Usage: hpc overlay [OPTION] .. <OVERLAY_FILE> [<OVERLAY_FILE> [...]]

Options:

    --srcdir=DIR   path to source directory of .hs files
                   multi-use of srcdir possible
    --hpcdir=DIR                  append sub-directory that contains .mix files
                                  default .hpc [rarely used]
    --reset-hpcdirs               empty the list of hpcdir's
                                  [rarely used]
    --output=FILE  output FILE
% hpc help draft
Usage: hpc draft [OPTION] .. <TIX_FILE>

Options:

    --exclude=[PACKAGE:][MODULE]  exclude MODULE and/or PACKAGE
    --include=[PACKAGE:][MODULE]  include MODULE and/or PACKAGE
    --srcdir=DIR                  path to source directory of .hs files
                                  multi-use of srcdir possible
    --hpcdir=DIR                  append sub-directory that contains .mix files
                                  default .hpc [rarely used]
    --reset-hpcdirs               empty the list of hpcdir's
                                  [rarely used]
    --output=FILE                 output FILE
```

5.7.4 Caveats and Shortcomings of Haskell Program Coverage

HPC does not attempt to lock the `.tix` file, so multiple concurrently running binaries in the same directory will exhibit a race condition. There is no way to change the name of the `.tix` file generated, apart from renaming the binary. HPC does not work with GHCi.

5.8 Using "ticky-ticky" profiling (for implementors)

Because ticky-ticky profiling requires a certain familiarity with GHC internals, we have moved the documentation to the GHC developers wiki. Take a look at its overview of the profiling options, which includeds a link to the ticky-ticky profiling page.