Masaryk University
Faculty of Informatics

GPU-based speedup of EACirc project

Bachelor Thesis

Jiří Novotný

Brno, Spring 2015

Contents

1 CMake

The EACirc sources mainly consists of *C* and *C++* code. The code was divided into reasonably logical sections but the overall structure and concept of the project were monolithic.[1] This led to compilation of all sources into one big executable of approximately 9 MB which took some non-trivial time. This design did not allow scaling.

On top of this EACirc is developed as a cross-platform application. To provide native builds for each supported platform (Windows [1] and Linux) special makefile or an IDE specific project file were used which described how to build the application. When a change in the build was introduced, e.g. a new source file was added, the change had to be manually implemented to all makefiles to provide consistency. This workflow was not easy to maintain as the violation of these rules could cause an uncomfortable pitfall.

To solve these problems the CMake [2] tool was integrated into the project of EACirc. The tool is developed and maintained by Kitware, Inc. [3] as an open-source software. The main purpose of this tool is to provide native builds of cross-platform applications and to minimize the effort to maintain the project.

Although there are many similar tools as CMake and some of them provides better features they are not so widely supported. For instance CMake generates project files for almost every common IDE and some of those IDEs comes with a built-in support for CMake.

1.1 CMake toolset

The CMake is actually a set of several tools that are taking care of building, testing, and deploying a user's *C* or *C++* project. These tools can be installed on Linux, Windows, or MacOSX. The CMake toolkit consists of the main tool `cmake` and the supportive `ccmake` (or `cmake-gui`), `ctest`, and `cpack`.

The `cmake` tool takes a configuration file called `CMakeLists.txt` distributed with the project source files and generates the platform specific makefiles as an output. Then the user invokes a platform specific tool for building – usually `make`, `ninja`, or `MSBuild`. If the process is successful the native binaries of the project are now made.

The `ctest` tool provides a simple platform for project testing. If the build is successful the user can run some custom made tests on the binaries.

The `cpack` tool provides a cross-platform mean to deploy your application on the target system.

The remaining `ccmake` and `cmake-gui` are just more convenient ways to use a `cmake` tool since `cmake` has only a command line interface. The former provides a TUI[2] and the latter provides GUI[3].

---

1. A monolithic binary is an executable that does not need any other dependencies or resources at a runtime. In other words, the binary is independent.
2. Text-based user interface (TUI)
3. Graphical user interface (GUI)

1.2 A closer look at the `cmake` executable

The `cmake` executable is not just a dummy build-system. The process of generating a makefile is quite sophisticated. At First the user chooses the *source directory* and the *build directory*. Then (s)he invokes the `cmake` command in a *build directory* with appropriate parameters. The subsequent process consists of several phases – selection of a native build-system (in a CMake terminology referenced as a *generator*), configuration based on a user-specific input, and the own generation of a makefile.[4]

The *source directory* is simply a directory where the project sources are located and as well as the top-level `CMakeLists.txt` file which is distributed with the sources. The build directory is an empty user-created directory in which the user wants the binaries to be build.

The selection of the *generator* depends on the user's platform, on the user-installed native build-systems, and on the user's intentions. The generator used on Linux is usually `make` or `ninja`. When the user wants to generate project files to a specific IDE, he chooses the appropriate generator – e.g. Visual Studio 2013 [4] on Microsoft Windows [1]. Usually the selection of the appropriate generator is done by CMake automatically.

The subsequent phase is configuration. Here the user specifies variable options for the build that the project supports. For instance some features of the application can be switched on/off or the location of a third party dependencies can be specified. Also the different build configuration can be switched, i.e. release or debug.

If the configuration is all right then the makefile is successfully created in the *build directory*. Then the user just invokes the appropriate tool to execute the makefile and the binaries are build.

It is worth mentioning that the makefile automatically detects any changes made in the *source directory*. So the user invokes the `cmake` executable just once to generate the makefile or to change the variable options of the build. The makefile also provides a way to install the application and/or to test it.

The minimal and the most common sequence of commands to build and install a project on Linux using the CMake is as follows:

```
mkdir <build_directory>
cd <build_directory>
cmake <path_to_source_directory>
make
make install
```

Note that the `make` is chosen as a default generator. In addition the default project settings and configurations are applied. The binaries are installed to the platform specific location, i.e. on Linux it is `/usr/share/local`.

4. Note that the exact scheme of this process can differ according to which interface of CMake is used – i.e. `cmake`, `ccmake`, or `cmake-gui`.

1.3 Changes made to the EACirc repository structure

There were several changes made to the EACirc repository structure. The new folder design reflects the logical structure of the EACirc philosophy.

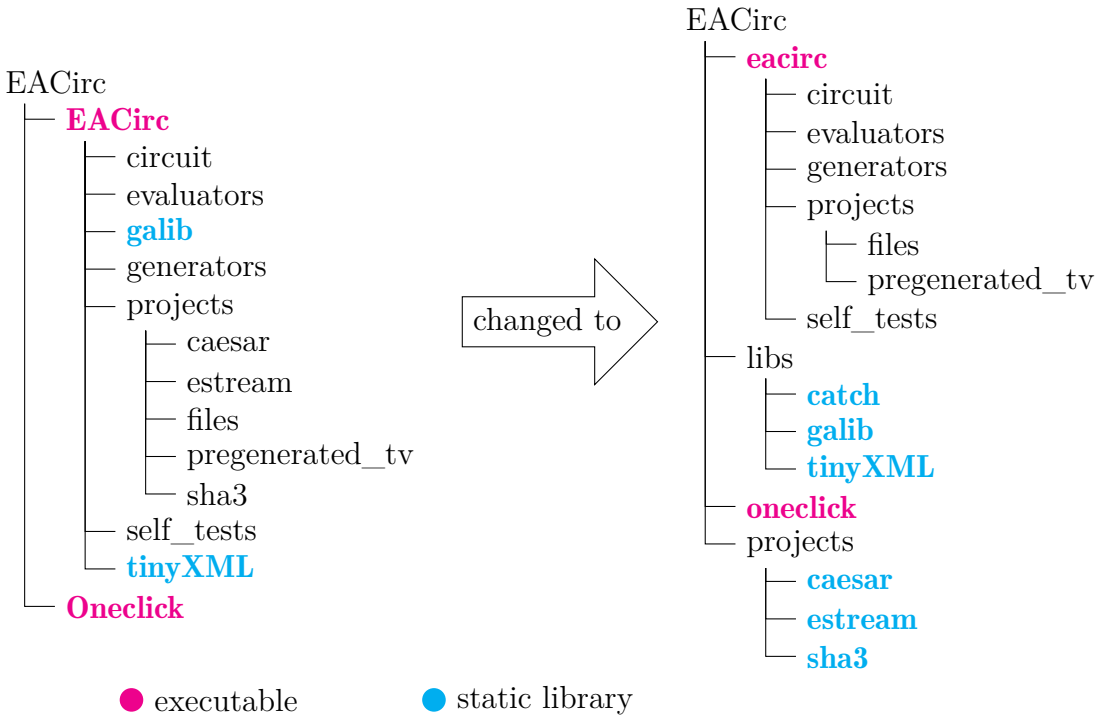

Figure 1.1: Old vs. new repository structure

The first and also the smallest change was to name all source folders with only small letters. Next the libraries from 3rd party providers `catch`, `galib`, and `tinyXML` were moved into the separate folder – the *libs* directory.

Then the so called *projects* were isolated. A *project* in EACirc terminology means a problem solving module. These *projects* are `caesar`, `estream`, `sha3`, `files` and `pregenerated_tv`. Since `files` and `pregenerated_tv` are both just small modules consisting from only one source file, it would be impractical to isolated them. Whereas the big modules `caesar`, `estream`, and `sha3` were moved to the the separate folder called the *projects* folder. Each of the isolated projects was remade to compile into a static library.[5]

The content od folders `eacirc` and `oneclick` is build into executables which are named accordingly to their corresponding folder. The *projects* which are now compiled into the static libraries are now statically linked to the `eacirc` executable representing the EACirc tool as a whole. The `oneclick` executable is a supportive tool for automated task management developed by Ľubomír Obrátil. [5]

5. There is a plan to remake the projects to modules loaded dynamically at runtime. This would require to compile them separately into the dynamic libraries.

1.4 The new build-system of EACirc

The new build-system is written on the CMake platform. This platform allows to define custom options for generating the build. Here is a descriptive list of EACirc specific options:

`BUILD_ONECLICK` enables building of Oneclick, the supportive tool for EACirc.

`BUILD_CAESAR` enables building of the Caesar project.

`BUILD_ESTREAM` enables building of the Estream project.

`BUILD_SHA3` enables building of the SHA-3 project.

`BUID_CUDA` enables to build the support for CUDA devices. This option is available only if the CUDA Toolkit [6] is installed on the build machine[6] and found by the CMake.

Since the *projects* are build into static libraries they must be linked to the `eacirc` executable at the compile time. This is done automatically when the option for the specific *project* is enabled. In the figure 1.2 are shown the dependencies of the all build targets.

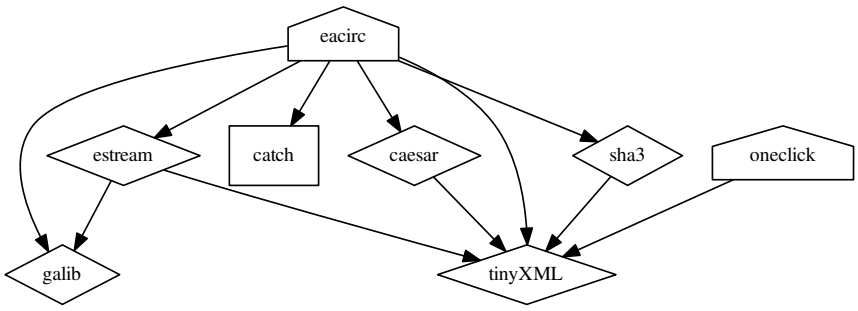

Figure 1.2: EAcirc dependency graph
The static libraries are shown in the rhombus. The executables have a house around them. The square represents an interface library.[7] The direction of the arrows represents that some build target depends on another one.

The build-system is also version aware. The current version is stored in the `eacirc/Version.h` header file. The version corresponds to git commit hash [7]. This means that for the correct build generation git tools must be properly installed on the build machine and found by CMake.[8]

The usage of CMake and the new options of building EACirc are explained in detail on the Github wiki project page under the Building EACirc section.

---

6. A build machine is a physical or a virtual machine that is used to build the project.
8. If git tools are installed and not found automatically by CMake then the path to git tools can be specified manually.

1.5 Project settings for CUDA

It is now much easier to set the project for CUDA support with CMake than with oridinaly makefiles. When the CUDA Toolkit [6] is installed and automatically found by CMake[9] then the option `BUILD_CUDA` becomes available. If this option is enabled then the `eacirc` executable is build using Nvidia [8] `nvcc` compiler and the C preprocessor macro `CUDA` is defined causing that the executable will be runnable on CUDA capable devices. When writing a code for CUDA the preprocessor macro `CUDA` can be queried.

9. If CUDA Toolkit is installed on the build machine but not found by CMake automatically then the path to CUDA Toolkit can be specified manually.

Bibliography

[1] Microsoft. (2015). Windows – microsoft windows, [Online]. Available: http://windows.microsoft.com.

[2] I. Kitware. (). Cmake, [Online]. Available: http://www.cmake.org/ (visited on 03/08/2015).

[3] ——, (). Kitware, inc. – leading edge, high-quality software, [Online]. Available: http://www.kitware.com/ (visited on 03/08/2015).

[4] Microsoft. (2015). Visual studio – microsoft developer tools, [Online]. Available: https://www.visualstudio.com/.

[5] Ľ. Obrátil, "Automated task management for eacirc and boing", type, FI MUNI.

[6] N. Corporation. (2015). Cuda toolkit, [Online]. Available: https://developer.nvidia.com/cuda-toolkit.

[7] B. S. Scott Chacon, *Pro Git*, 2nd editon. Apress, Dec. 24, 2014, ISBN: 978-1484200773.

[8] N. Corporation. (2015). Welcome to nvidia - world leader in visual computing technologies, [Online]. Available: http://www.nvidia.com.

TODO Fix the autors of online resources
TODO Fix the titles in the bibliography to dislay big letters correctly.
TODO Cite Lobo's theses about oneclick and fix the source.
TODO Cite Martin Ukrop thesis in Introduction. What is EACirc?