cran2deb: A system to automatically provide 1500+ CRAN packages as Debian binaries

Charles Blundell[1], **Dirk Eddelbuettel**[2]

1. Gatsby Computational Neuroscience Unit, UCL
2. Debian Project

* Contact author: edd@debian.org

Keywords: R, CRAN, packages, Debian

This paper introduces the cran2deb system that provides automated builds of binary Debian packages from (essentially all) available CRAN packages for the R statistical environment and language.

Part of the growing popularity of R is due to the availability of over 1500 source packages at the CRAN repositories alone. These packages provide anything from R extension via new statistical methodologies, add approaches specific to given scientific disciplines, connect R to various backends such as databases, or provide specific user-interfaces. The BioConductor repository provides another few hundred packages specifically for bioinformatics research bringing the total to almost 2000 packages.

Users, however, are frequently stymied by the system administration task of building and compiling these source packages as this may entail obtaining and installing other toolchains or libraries from unknown third parties. This is often a multi-step process with manual intervention to resolve problems: source packages may not compile, dependencies may require further dependencies (that in themselves do not compile). The whole process may, after several hours, prove entirely fruitless. Linux distributions have shown how a single package management system, for all installed software, can help: universal control over the intricate details inter-package dependencies makes the system administration task significantly less arduous. Packages are already compiled and as the exact same package is used by many people, the user can have a higher expectation of an installation working first time. It would therefore be helpful to provide binary packages that are fully integrated into an existing Linux distribution such as Debian. Users could then use the existing package management tools to install, upgrade, query or remove packages. This becomes even more useful in large installations such as departments, work groups or computing clusters.

The cran2deb system fits into the infrastructure for the Debian GNU/Linux distribution, a Linux distribution with over 20,000 packages. cran2deb utilises the Debian toolchains for package building, in particular the pbuilder program to facilitate unsupervised building of packages in pristine build environments augmented with finely-grained build-dependencies. It uses a simple database backend (currently provided by SQLite) for stateful information and logging.

This work extends prior work by Eddelbuettel et al (2007). It is however a reimplementation using R as a scripting language, provided primarily by the first author. This was part of the Google Summer of Code (GSoC) program 2008, with the second author acting the mentor for this GSoC project. It has been extended further since the GSoC 2008 program finished.

The approach taken here is bottom-up: information about individual packages is analysed and aggregated into a dependency graph—the dependency graph is not just restricted to R packages; any Debian package may be included. If necessary, the minimal set of packages needed to fulfil the dependencies for a given package are then built alongside with the target package. The set of available packages is updated by querying the CRAN mirror network using R's internal tools, and new packages are compiled at each update pulse. Current versions of available packages are then provided in a downloadable repository.

We discuss some of the lessons learned in building this service, and the next steps that are needed to make `cran2deb` part of the CRAN network.

References

Eddelbuettel D, Vernazobres D, Gebhard A and Moeller S (2007), `apt-get install cran bioc`: On automated builds of 1700 R packages for Debian. Presentation at useR! 2007, Iowa State University, Ames, Iowa, August 8-10, 2007.