Popularity contest 15th birthday

B. Allombert

Debian

18/01/2013

1. **Introduction**

2. **The birthday**

3. **How it works**

4. **Recent changes**

5. **Issues**

What is popularity-contest

Package configuration

┤ Configuring popularity-contest ├

The system may anonymously supply the distribution developers with
statistics about the most used packages on this system. This information
influences decisions such as which packages should go on the first
distribution CD.

If you choose to participate, the automatic submission script will run once
every week, sending statistics to the distribution developers. The
collected statistics can be viewed on http://popcon.debian.org/.

This choice can be later modified by running "dpkg-reconfigure
popularity-contest".

Participate in the package usage survey?

<Yes>   <No>

The beginning...

Popularity
contest 15th
birthday

B. Allombert

Introduction

The birthday

How it works

Recent
changes

Issues

```
popularity-contest (0.00) unstable; urgency=low

  * Initial test release.

 -- Avery Pennarun <apenwarr@debian.org>
    Sat, 24 Oct 1998 22:33:58 -0400
```

Popularity
contest 15th
birthday

B. Allombert

Introduction

The birthday

How it works

Recent
changes

Issues

```
popularity-contest (1.0-1) unstable; urgency=low
  ...
 -- Avery Pennarun <apenwarr@debian.org>
    Wed, 23 Dec 1998 14:28:37 -0500

popularity-contest (0.1-2) frozen unstable; urgency
  ...
 -- Avery Pennarun <apenwarr@debian.org>
    Wed, 23 Dec 1998 14:28:37 -0500

popularity-contest (0.1) frozen unstable; urgency=l
  ...
 -- Avery Pennarun <apenwarr@debian.org>
    Sun, 20 Dec 1998 16:11:02 -0500
```

Popularity
contest 15th
birthday

B. Allombert

Introduction

The birthday

How it works

Recent
changes

Issues

```
popularity-contest (1.3-1) unstable; urgency=high
 ...
 -- Avery Pennarun <apenwarr@debian.org>
    Thu,  7 Mar 2002 12:39:49 -0500

popularity-contest (1.2) unstable; urgency=low
 ...
 -- Avery Pennarun <apenwarr@debian.org>
    Thu, 21 Feb 2002 09:11:58 -0500

popularity-contest (1.1-1) unstable; urgency=low
 ...
 -- Avery Pennarun <apenwarr@debian.org>
    Tue,  6 Nov 2001 21:19:38 -0500
```
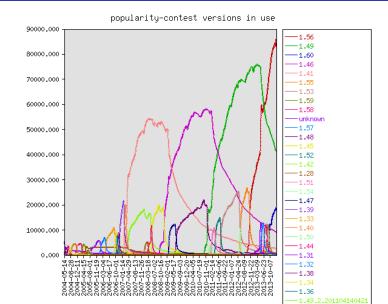
Some history highlight

popularity-contest versions in use

Trends

- 160000 submitters.
- Architectures: amd64 60%, i386 38%
- Releases: wheezy 53%, squeeze 26%, testing 10%, lenny 5%

What is reported ?

popularity-contest reports:

- the random HOSTID (in /etc/popularity-contest.conf)
- the dpkg Vendor field (dpkg-vendor –query Vendor)
- the dpkg architecture (dpkg –print-architecture)
- the version of popularity-contest
- the list of packages installed
- the ctime and atime of relevant files in installed packages

The data is sent every 7 days, either by HTTP or SMTP.

Server side process

Popularity
contest 15th
birthday

B. Allombert

Introduction

The birthday

How it works

Recent
changes

Issues

The server maintains a database of reports indexed by HOSTID.
Once a day:

- the server anonymize the emails by removing the headers.
- it stores the new reports in the database.
- it removes reports older than 20 days.
- it computes the statistics (slow).
- it generates the website.

Traditional popcon.debian.org layout

debian

Debian Popularity Contest

The popularity contest project is an attempt to map the usage of Debian packages. This site publishes the statistics gathered from report sent by users of the popularity-contest package. This package sends every week the list of packages installed and the access time of relevant files to the server via email. Every day the server anonymizes the result and publishes this survey. For more information, read the README and the FAQ.

Popcon statistics for source package [　　　　　　　　　] [Go]

All reports **Stable reports**

Statistics for the whole archive sorted by fields:

inst [gz] vote [gz] old [gz] recent [gz] no-files [gz]

Statistics by maintainers sorted by fields:

inst [gz] vote [gz] old [gz] recent [gz] no-files [gz]

Statistics by source packages (sum) sorted by fields:

inst [gz] vote [gz] old [gz] recent [gz] no-files [gz]

Statistics by source packages (max) sorted by fields:

inst [gz] vote [gz] old [gz] recent [gz] no-files [gz]

Statistics for sections sorted by fields

		inst	[gz]	vote	[gz]	old	[gz]	recent	[gz]	no-files	[gz]
main	:	inst	[gz]	vote	[gz]	old	[gz]	recent	[gz]	no-files	[gz]
contrib	:	inst	[gz]	vote	[gz]	old	[gz]	recent	[gz]	no-files	[gz]
non-free	:	inst	[gz]	vote	[gz]	old	[gz]	recent	[gz]	no-files	[gz]
non-US	:	inst	[gz]	vote	[gz]	old	[gz]	recent	[gz]	no-files	[gz]

New popcon.debian.org layout by Stéphane Blondon

Privacy improvements in Jessie

Popularity
contest 15th
birthday

B. Allombert

Introduction
The birthday
How it works
Recent
changes
Issues

- Submissions are encrypted using public-key cryptography to protect against eavesdropping.
- All timestamps reported are truncated to multiple of twelve hours.

- Report the dpkg Vendor field to track derivative distribution.

Two basic issues:

- Statistic for packages used by popcon are useless
- Some files acess times are changed by cron job, dpkg hooks, etc.
- Using the mount option "noatime" breaks vote reporting. Use "relatime" instead.

Wishlist

Popularity
contest 15th
birthday

B. Allombert

Introduction
The birthday
How it works
Recent
changes
Issues

- Allows private packages to opt out of popularity-contest by using a Dpkg field.
- Support /etc/popularity-contest.d/ for derivatives distribution.
- Make it easier for derivatives distribution. to set up a popularity-contest server.

Criticisms

Popularity
contest 15th
birthday

B. Allombert

Introduction

The birthday

How it works

Recent
changes

Issues

- Popcon measures popularity, not usefulness or quality
- It is biaised against some use cases
- vote is biaised against some packages