Perspectives on Bugs in the Debian Bug Tracking System

Julius Davies*, Hanyu Zhang*, Lucas Nussbaum†, Daniel M. German*

* Dept. of Computer Science, University of Victoria, Canada
† LORIA, University of Nancy, France
Introduction

Debian:
- Most important community-driven GNU/Linux distribution
- Large collection of software packages
  Debian maintainers \(\rightsimeq\) journal editors
Introduction

Debian:
- Most important community-driven GNU/Linux distribution
- Large collection of software packages
  Debian maintainers $\leadsto$ journal editors

Debian bugs:
- Usually reported against packages
  - Packages $\leadsto$ Software modules
  - Packages uploads $\leadsto$ Aggregation of commits, releases
  - But 25,000 packages in Debian

J. Davies, H. Zhang, L. Nussbaum and D. German  Perspectives on bugs in the Debian BTS
Introduction

Debian:
- Most important community-driven GNU/Linux distribution
- Large collection of software packages
  Debian maintainers ~ journal editors

Debian bugs:
- Usually reported against packages
  - Packages ~ Software modules
  - Packages uploads ~ Aggregation of commits, releases
  - But 25,000 packages in Debian
- Also used to track Debian development
  - Problems in the Debian infrastructure
  - Development procedures (intend-to-package bugs)
    → using pseudo-packages (e.g. ftp.debian.org)
Questions

1. Number of bugs associated with packages (vs pseudo-packages)?

2. Correlation between changes (uploads) and reported bugs?

3. Who is reporting the bugs?

4. How does bug frequency relate to package popularity?
Input Data and Methodology

Input data:

1. Ultimate Debian Database (see earlier presentation)
   - But only contains summary info for bugs
   - Not all the bug comments
     - Useful in some cases

2. Public dump from the Debian Bug Tracking System
Input Data and Methodology

Input data:
1. Ultimate Debian Database (see earlier presentation)
   - But only contains summary info for bugs
   - Not all the bug comments
     - Useful in some cases
2. Public dump from the Debian Bug Tracking System

Methodology:
1. Fetched all bugs between 01/2007 and 12/2009 (158,058 bugs)
2. Aggregated per month and package
   → (iceweasel, 2007-01, 70), (iceweasel, 2007-02, 80)
3. Used UDD to separate packages from pseudo-packages
Bugs affecting packages (vs pseudo-packages)

89% of bugs are reported against real packages
Correlation between uploads and bugs

Number of uploads and of reported bugs are correlated

Spearman = 0.811, p < 0.001
## Bug reporters and packages

Number of different packages each bug reporter has filed bugs for:

<table>
<thead>
<tr>
<th>Number of Packages</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>55%</td>
</tr>
<tr>
<td>2-4</td>
<td>25%</td>
</tr>
<tr>
<td>5-9</td>
<td>9%</td>
</tr>
<tr>
<td>10-19</td>
<td>5%</td>
</tr>
<tr>
<td>20-29</td>
<td>2%</td>
</tr>
<tr>
<td>30-99</td>
<td>3%</td>
</tr>
<tr>
<td>100+</td>
<td>1%</td>
</tr>
</tbody>
</table>

(18 bug reporters with $\geq 1000$ different packages)

Most bug reporters only report bugs on a small number of packages.
Conclusions

Debian BTS:
- Interesting research subject
- Bug management as an ecology
  many different software packages

Ultimate Debian Database:
- Enabled easy data-mining of different types of Debian data
  Packages, bugs, uploads, packages popularity

Open questions:
- Who is reporting bugs in Debian? Users? Debian developers?
- Relationship between Debian development and other projects:
  - Upstream projects?
  - Derivative distributions (Ubuntu)?