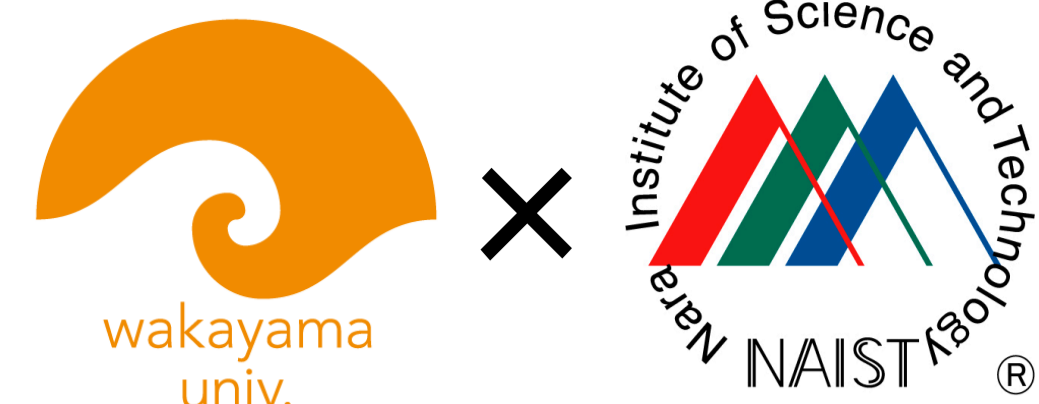


# A Dataset of High Impact Bugs: Manually-Classified Issue Reports

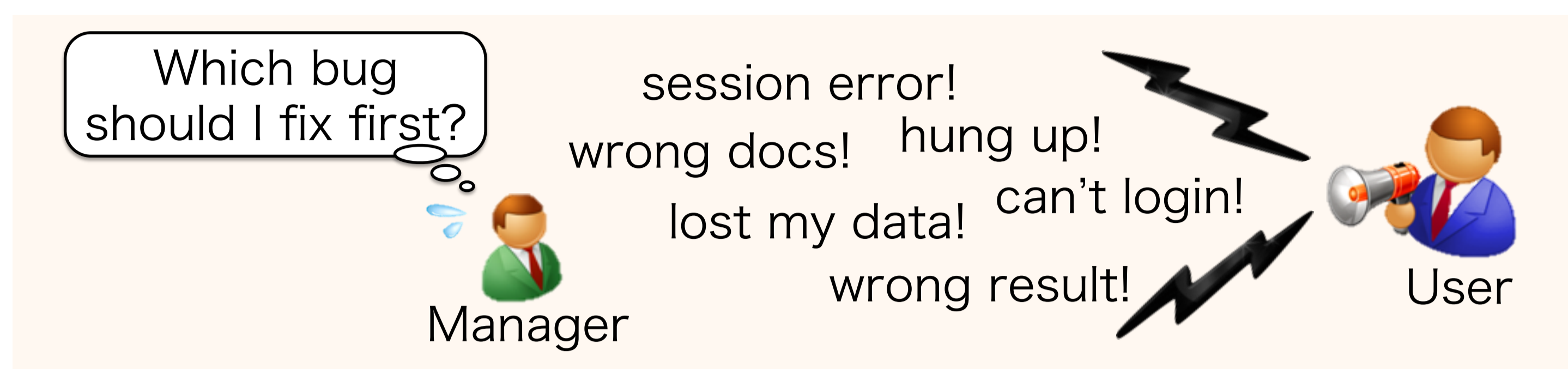


Masao Ohira<sup>1</sup>, Yutaro Kashiwa<sup>1</sup>, Yosuke Yamatani<sup>1</sup>, Hayato Yoshiyuki<sup>1</sup>, Yoshiya Maeda<sup>1</sup>, Nachai Limsettho<sup>2</sup>, Keisuke Fujino<sup>2</sup>, Hideaki Hata<sup>2</sup>, Akinori Ihara<sup>2</sup> and Kenichi Matsumoto<sup>2</sup>

<sup>1</sup>Wakayama University, Japan  
<sup>2</sup>Nara Institute of Science and Technology, Japan

## Background

A great number of bugs are found not only before but also after releasing products. A manager must prioritize reported bugs to be fixed. Many studies proposed methods to support bug-triaging. However, in these studies, each bug is equally treated without considering its impact on the bug management process and software products.



## High impact bugs

### Process bug

A process bug is a bug that affects a bug management process.

#### ■ Blocking bug [Valdivia'14 etc.]

A bug that blocks other bugs from being fixed



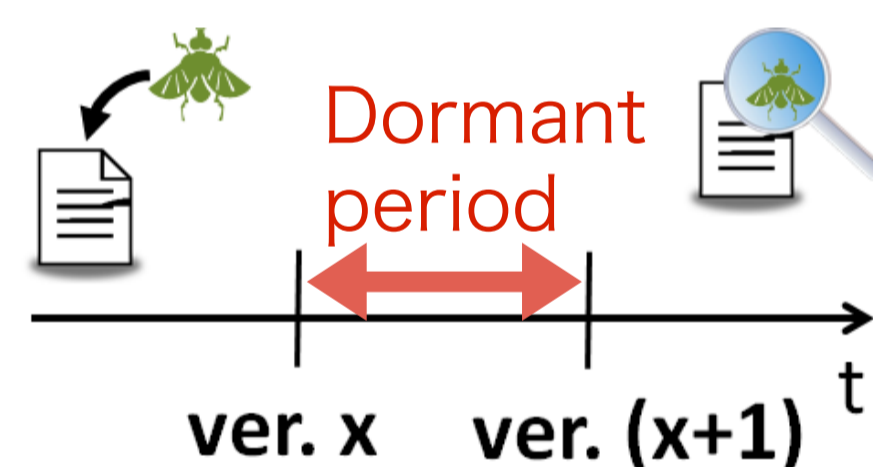
#### ■ Surprising bug [Shihab'11]

A bug that appears in unexpected timing and location



#### ■ Dormant bug [Chen'14]

A bug that was introduced in one version of a system but was not reported until the next immediate version

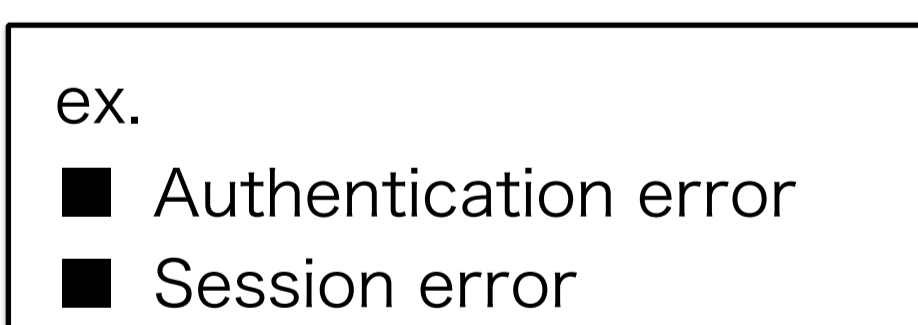


### Product bug

A product bug is a bug that affects the quality of software products.

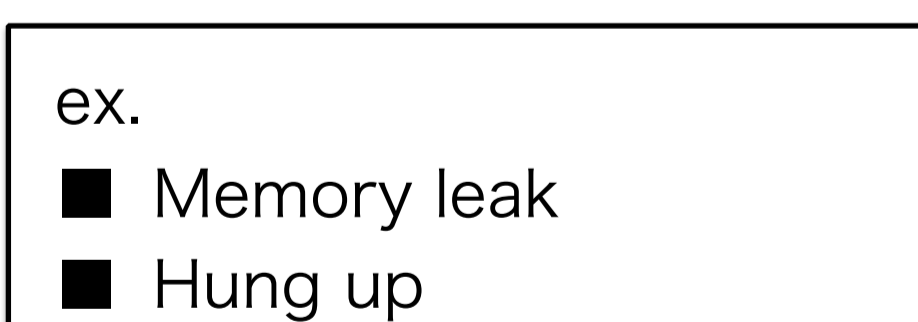
#### ■ Security bug [Gegick'10 etc.]

A bug that has security risks



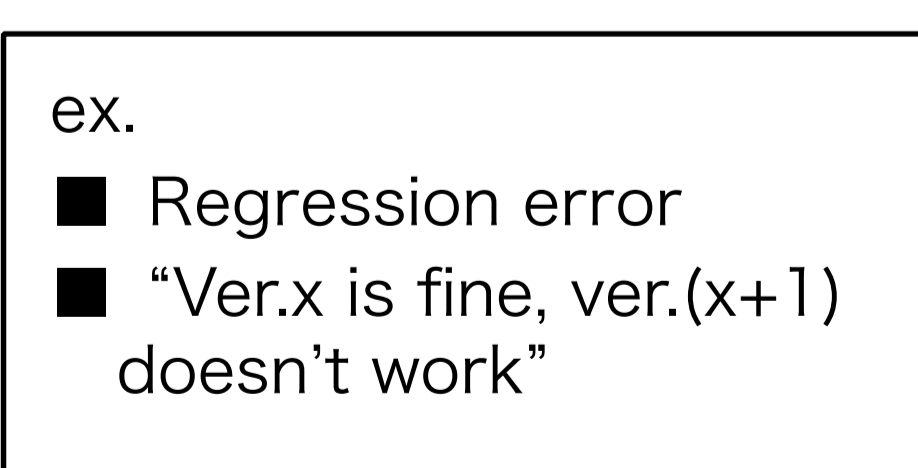
#### ■ Performance bug [Nistor'13]

A bug that makes system's performance degrade



#### ■ Breakage bug [Shihab'11]

A functional bug that was made by code modifications to add new features or to fix bugs



### Reference

[Valdivia'14] H. Valdivia Garcia and E. Shihab, "Characterizing and predicting blocking bugs in open source projects," in Proc. of MSR '14, 2014, pp. 72–81.  
[Chen'14] T.-H. Chen, M. Nagappan, E. Shihab, and A. E. Hassan, "An empirical study of dormant bugs," in Proc. of MSR '14, 2014, pp. 82–91.  
[Shihab'11] E. Shihab, A. Mockus, Y. Kamei, B. Adams, and A. E. Hassan, "High- impact defects: A study of breakage and surprise defects," in Proc. of ESEC/FSE '11, 2011, pp. 300–310.  
[Gegick'10] M. Gegick, P. Rotella, and T. Xie, "Identifying security bug reports via text mining: An industrial case study," in Proc. of MSR '10, 2010, pp. 11–20.  
[Nistor'13] A. Nistor, T. Jiang, and L. Tan, "Discovering, reporting, and fixing performance bugs," in Proc. of MSR '13, 2013, pp. 237–246.

## Dataset

Issue report data has been collected from the Apache Ambari, Camel, Derby, and Wicket projects which JIRA is used for managing reported issues.

	All the issues in Nov. 20 2014	Our Dataset	
		BUG	IMPROVEMENT
Ambari	8,389	871	129
Camel	8,063	580	420
Derby	6,772	734	26
Wicket	5,769	663	337

## Manual Classification: Results

