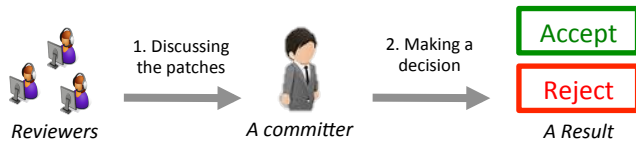


Background

1. Reviewers discuss the quality of patches.
2. A committer decides whether or not to accept patches.



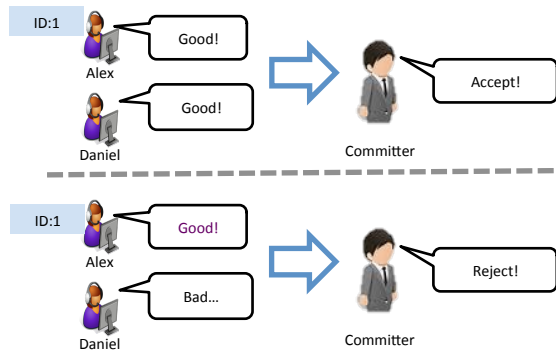
15% - 31% patches of all reviews have not reached a consensus among reviewers and a committer[1].

MOTIVATE

- A reviewer does not always give correct feedbacks.
- We think that the reviewer is likely to make a code review process ineffective.

Study Design

RQ1. How often does a reviewer fail to reach a consensus among reviewers and a committer?



The Level of Agreement (LOA) $0 \leq LOA \leq 1$

- $LOA(\text{Alex}) = 1/2 = 0.5$
- $LOA(\text{Daniel}) = 2/2 = 1.0$

RQ2. What is the impact of a reviewer with a low level of agreement in a code review?

Reviewing Time: The time in days from the first patch submission to the final review conclusion.

Discussion Length: The number of comments which reviewers post into a review.

[1] T.Hirao, A.Ihara, Y.Ueda, P.Phannachitta, K. Matsumoto, "The Impact of A Reviewer's Low Level of Agreement in a Code Review Process", The International Conference on Open Source Systems 2016 (OSS'16), in proceedings.

This work has been conducted as part of our research under the Program for Advancing Strategic International Networks to Accelerate the Circulation of Talented Researchers.

Experimental Datasets

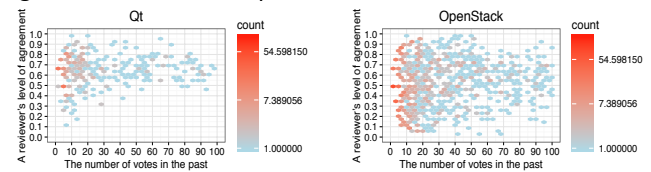
We filter out review reports with no votes or only bot's vote.

OSS Projects	Qt	openstack
# of Review Reports	55,523	56,038

Case Study

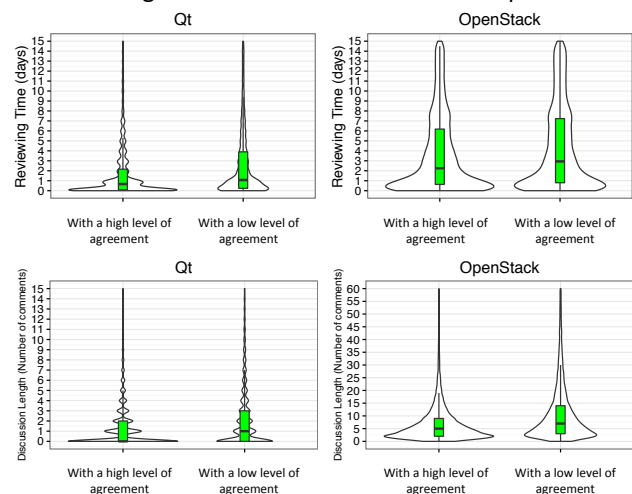
RQ1 Result

More experienced reviewers are likely to have a higher level of agreement than less experienced reviewers.



RQ2 Results

Reviewers with the lower level of agreement are more likely to take a longer time in review and discussion process.



Future Work

- To investigate why a reviewer fails to reach a consensus based on the contents of the discussion.
- To understand an actual meaning (i.e., positive or negative) of a reviewer's feedback using NLP.

GOAL

We would like to recommend a reviewer who gives correct feedbacks automatically.