

Threats to Validity

- Generated repositories may not reflect actual systems
- Lack gold-standard for measuring precision in actual systems
- Roadblocks to commit origin recovery:
 - re-writing the default merge commit message
 - multiple branches pointing to the same commit
 - implicit merges (e.g., cherry-pick, rebase, fast-forward merge) are undetectable
 - renamed branches are undetectable

Conclusion

- The most prevalent explicit merges (80%) involve pull-requests and branches.
- Most merge commits use the default message.
- Commit origins can be found effectively and efficiently using the algorithm presented.
 - average precision of 80%
 - average accuracy of 98%
- Commits can be separated based on branch origin (i.e., all commits involved in one task or concern).