Does Distributed Development Affect Software Quality? An Empirical Case Study of Windows Vista

By C. Bird, N. Nagappan, P. Devanbu, H. Gall, B. Murphy

Presented by

Tanja Werthmüller
Overview

1. What is distributed development
   - What does the paper say
2. Difficulties in distributed development
3. Assumptions in the paper
4. Hypotheses
5. Methods and Results
6. Conclusion
What is distributed development

- “A Distributed Development project is a research and development project that is done across many business work sites or locations.” (Wikipedia)

- A project can be distributed in many different ways:
  - Geographically
  - Organizationally
  - Temporally
  - ...
What does the paper say

- In the paper they were interested in the effect of **globally** distributed software development.
- The paper defines six different levels of distribution:
  - Building
  - Cafeteria
  - Campus
  - Locality/Region
  - Continent
  - World
Difficulties in Distributed Development

- Lack of communication
- Inconsistent use of tools
- Cultural differences
- Distance
Assumptions in the paper

- Completely within Microsoft
- Historical development data from the implementation of Windows Vista
- Focus on post-release failures
Hypotheses

I. „Binaries that are developed by teams of engineers that are distributed will have more post-release failures than those developed by collocated engineers.“

II. „Binaries that are distributed will be less complex, experience less code churn, and have fewer dependencies than collocated binaries.“
Experimental Analysis

- Increase of failures in distributed binaries compared to collocated binaries:
  - < 17%
  - < 9% considering team size
- Only statistically significant for split > D

→ I. could not be confirmed
Differences in Binaries

- Code changes:
  - Size of changes
  - Frequency of edits
  - ...
  - Dependencies
  - ...

- Size and complexity:
  - Number of lines
  - Number of functions
  - Depth of inheritance
  - ...

- Only differences because of team sizes

⇒ II. could not be confirmed
Validity/Limitation of the paper

- 4'000 binaries
- Source code base of over 60 MLOC
- 3'000 developers

- Only one large project
- Only Microsoft intern
Situation at Microsoft

- Communication
- Inconsistent use of tools
- Cultural differences
- Distance

- Synchronous communication
- Standardized tools
- Teams mostly in one country
- Knowing team members for years
Conclusions and Further Studies

- “It is possible to have a globally distributed development without affecting software quality.”
- “Organizational differences are much stronger indicators of quality than geography.”

- Determine, which strategies practiced by Microsoft actually helped to avoid negative impacts on software quality.
- New, not considered, characteristics how distributed and collocated binaries can differ from each other.