


Reliability Estimates For the Windows Operating System

Why is writing Software So Difficult?

Brendan Murphy
Microsoft Research Cambridge
11th April 2008

Brendan Murphy, Server Computing Summit, Berlin 2008

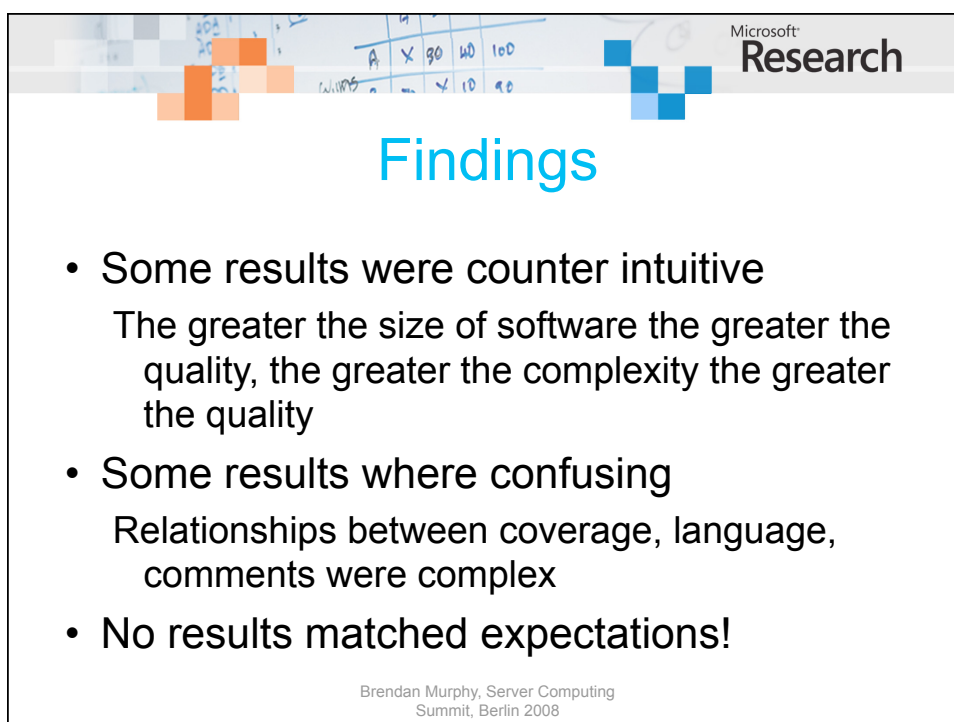
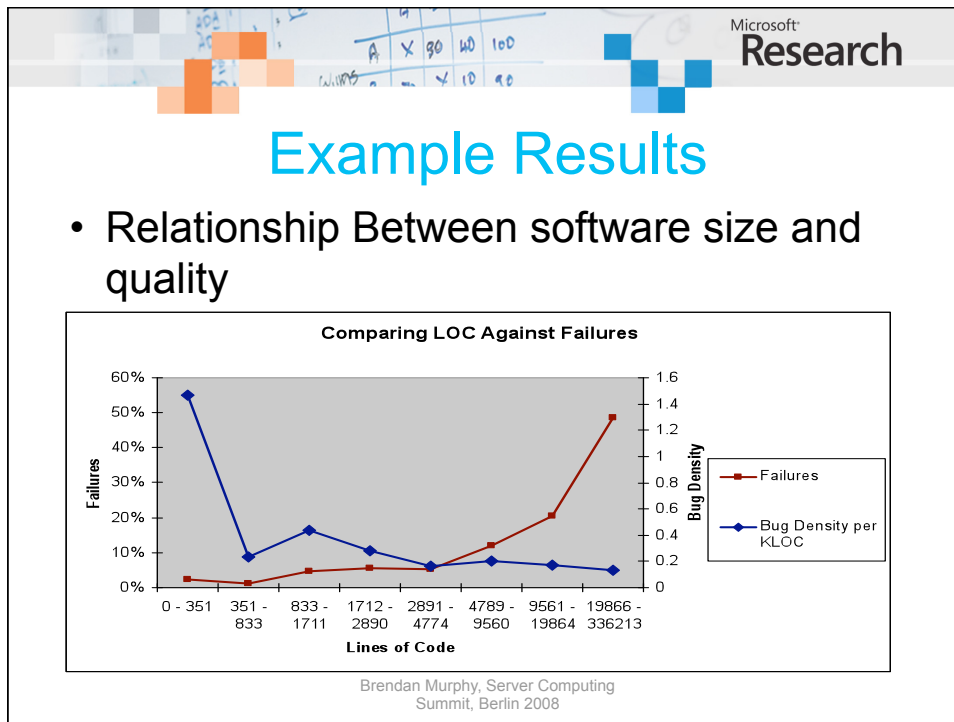



Background

Understand the relationship between Software Development and its resulting quality

- Assume that the relationship between software attributes (size, complexity, coverage, language etc) and quality are well understood
- Measure these attributes on Windows 2000
- Measure end user quality
- Quantify the impact of each attribute on end user quality

Brendan Murphy, Server Computing Summit, Berlin 2008






Presentation Outline

- Measuring Quality
- Development metrics
- Risk and Prediction models
- People Measurement
- Organizational Structure

Brendan Murphy, Server Computing
Summit, Berlin 2008



Characterizing Quality

Quality is as perceived by the end user

- Windows collects bugs through multiple processes
 - OCA, Watson, PSS, Call Centres, Partner services etc.
- Quality is also impacted by look and feel issues
- **Quality is defined as bugs fixed in the source code within six months of the product release**

Logic

- Bugs have been triaged based on business case
- It's the least worse measure

Brendan Murphy, Server Computing
Summit, Berlin 2008

Microsoft
Research

Software Churn

- Lehman and Belady identified its importance in 1960s
- Measure code rather than Binary churn
- Key attributes
 - Opportunities to churn (dependent upon development process)
 - Churn frequency
 - Amount of churn
 - Frequency of repetitive churn

Brendan Murphy, Server Computing Summit, Berlin 2008

Microsoft
Research

Relationship Identifies Between Churn and Failures

Use of Relative Code churn measures to predict System Defect Density, ICSE 2005
Nagappan, Ball (Microsoft)

Brendan Murphy, Server Computing Summit, Berlin 2008

Microsoft
Research

Additional Development Metrics

- Changes in Metrics Over the Development
 - Test Coverage
 - Bugs discovered during development
 - Code complexity
 - Cyclomatic (McCabes)
 - OO Metric
 - Dependency Information
- Architectural Layering
- Legacy Code
- Code Velocity


Brendan Murphy, Server Computing
Summit, Berlin 2008

Microsoft
Research

Developing a risk model Initial Findings

		Training data
		Win 2003
Win 2003		73%
Win XP SP1		64%
Win XP SP2		71%
Win 2003 SP1 RC		78%

Brendan Murphy, Server Computing
Summit, Berlin 2008



The missing Factor


If a team is writing software what is the most important issue?

- The process being used?
- The language?
- The development tools?
- The people?

If People which aspects

- Individual
- Team Profile
- Organizational Behaviour

Brendan Murphy, Server Computing Summit, Berlin 2008



Characterizing People

- How do you measure effort?
- Charactering development bugs is complex
- Engineering behaviour is influenced by their department as well as their own ability
- Bugs are assigned not earned
 - The best engineers often have the most bugs assigned to them!

People Data CANNOT be Interpreted

Brendan Murphy, Server Computing Summit, Berlin 2008




Team Profile

- Prior Experience
 - On Specific Software
 - Of Projects
- Mentor Prior Experience
- Project Experience
- Architectural Layering

Vista/ Win 2008 Provides A Benchmark For Future Analysis

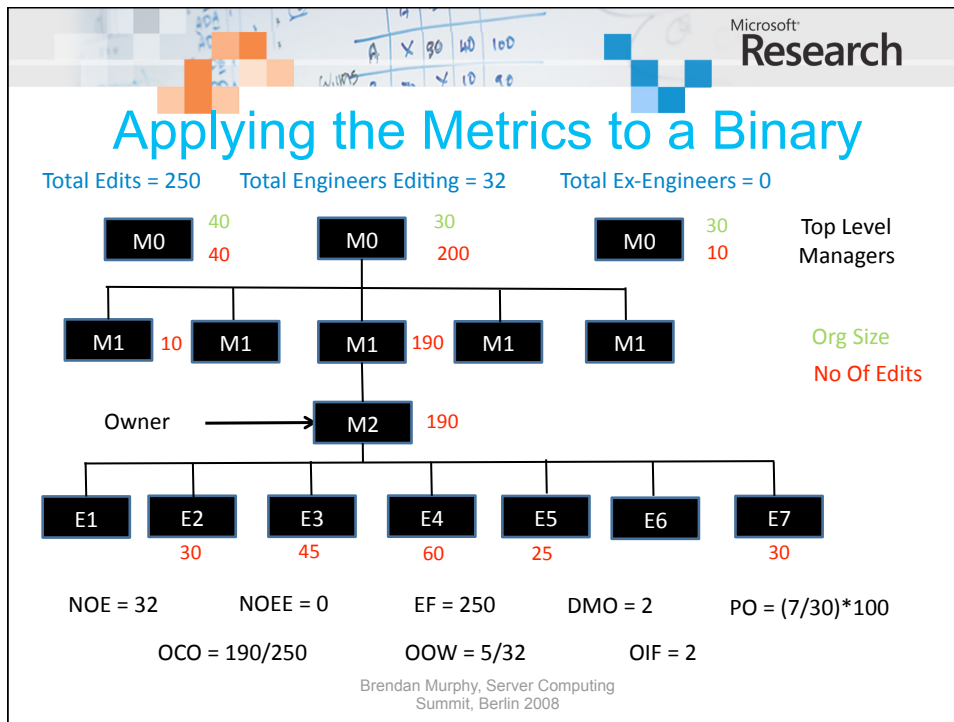
Brendan Murphy, Server Computing
Summit, Berlin 2008



Organizational Structure Metrics

- Propose eight measures that quantify organizational complexity capturing issues such as
 - Organizational distance of the developers
 - The number of developers working on a component
 - Component changes within the context of an organization
- Empirically evaluate the ability of these metrics to identify failure-prone binaries in Windows Vista

Brendan Murphy, Server Computing
Summit, Berlin 2008

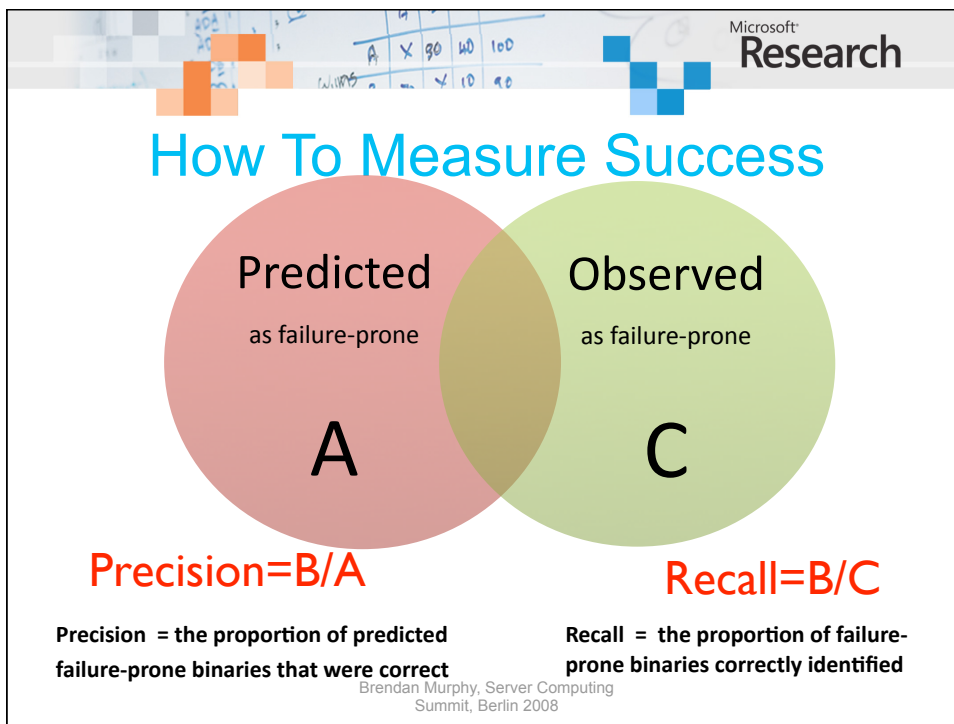


Microsoft Research

Are all these metrics required

- Step-wise regression is a robust technique to build a model
- Identify the predictor with the largest correlation with the dependent variable
- Add new predictors to the model based on their partial correlation with the predictors already in the model
- With each new set of predictors, the model is re-evaluated

Brendan Murphy, Server Computing Summit, Berlin 2008



Microsoft
Research

How Good Are Organizational Metrics

Comparison to other predictors?

Model	Precision	Recall
Organization Structure		
Churn	78.6%	79.9%
Complexity	79.3%	66.0%
Dependencies	74.4%	69.9%
Coverage	83.8%	54.4%
Pre-Release Bugs	73.8%	62.9%

Brendan Murphy, Server Computing Summit, Berlin 2008

Microsoft
Research

Threats to Validity


- Independence
 - Engineers had no knowledge of the study (no gaming)
 - Researchers have no axe to grind
- Co-incidental issues
 - Complex binaries are those that straddle multiple orgs
 - Binaries managed by single groups are simpler and their requirements are well understood
- Causal Issues
 - No major re-evaluation of binary ownership
 - Binaries with no major functional changes may have transitioned ownership between orgs
 - Legacy code

Brendan Murphy, Server Computing
Summit, Berlin 2008

Microsoft
Research

What Behaviour Indicates Risk


Brendan Murphy, Server Computing
Summit, Berlin 2008



Next Steps

- Transfer these findings into the development process.
- Replicate studies on other product releases
- Improve Failure Characterization
- Further research the impact of team experience on resulting software quality

Brendan Murphy, Server Computing
Summit, Berlin 2008



Conclusions

- Adds to our understanding of what drives software quality
- New set of metrics developed to quantify organizational complexity
- Organizational metrics are a strong predictor of software quality
 - Better than code churn metrics
 - Investigating merging the metrics together
- “Mythical Man Month” was correct all along
 - Projects need to plan their organizational structure

Brendan Murphy, Server Computing
Summit, Berlin 2008