

# Standard Regression Testing Does Not Work

Daniel Hansson, CEO Verifyter

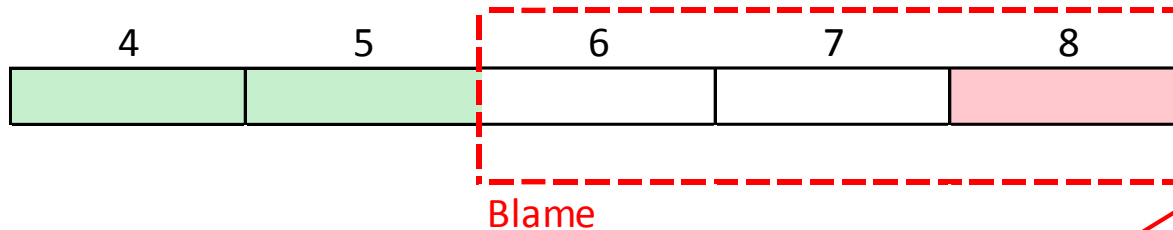
**verifyter**

# Standard Regression Testing

- Check out the latest from the version control system
- Kick off a script that runs a lot of tests and present the test results
- Email the engineer(s) who have committed new changes to the VCS since the last time the test(s) passed

# Blame – 3 Standard Options

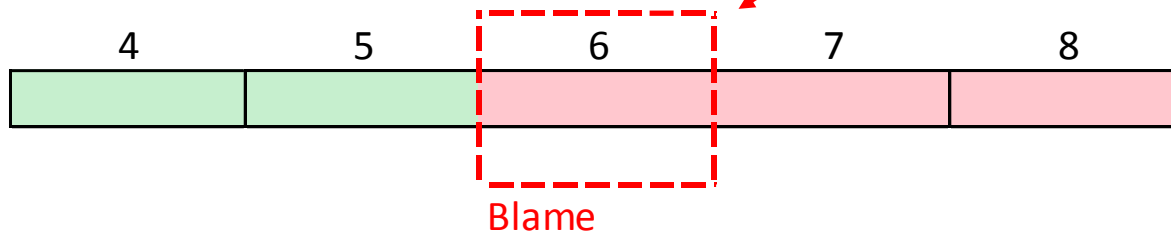
Large Test Suite  
(testing every x revision)



1. Blame all commits since last pass

2. Blame the first commit that caused the test to fail

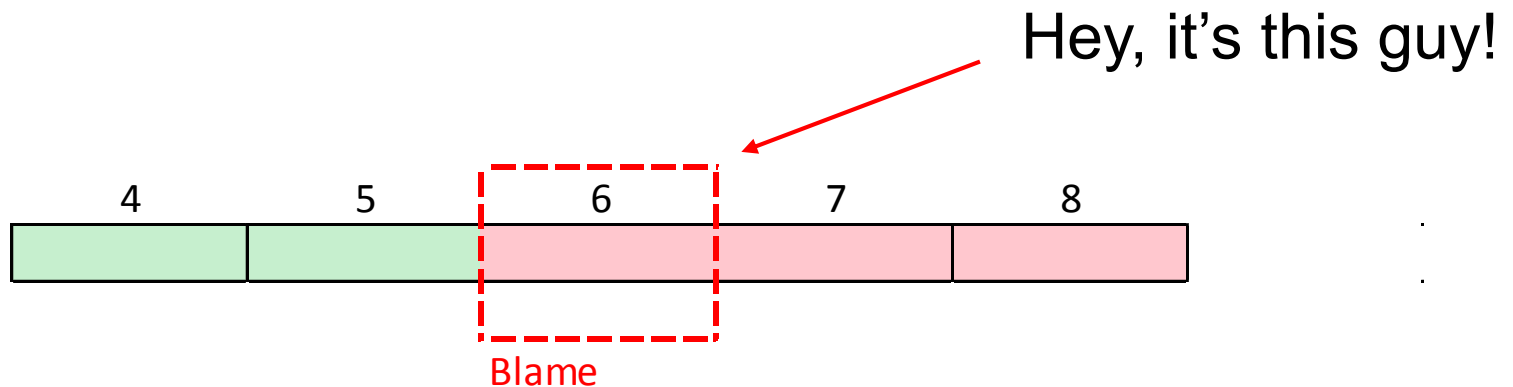
Small Test Suite  
(testing each revision)



3. Don't blame anyone.  
Manually debug the test failure

# The Standard Approach

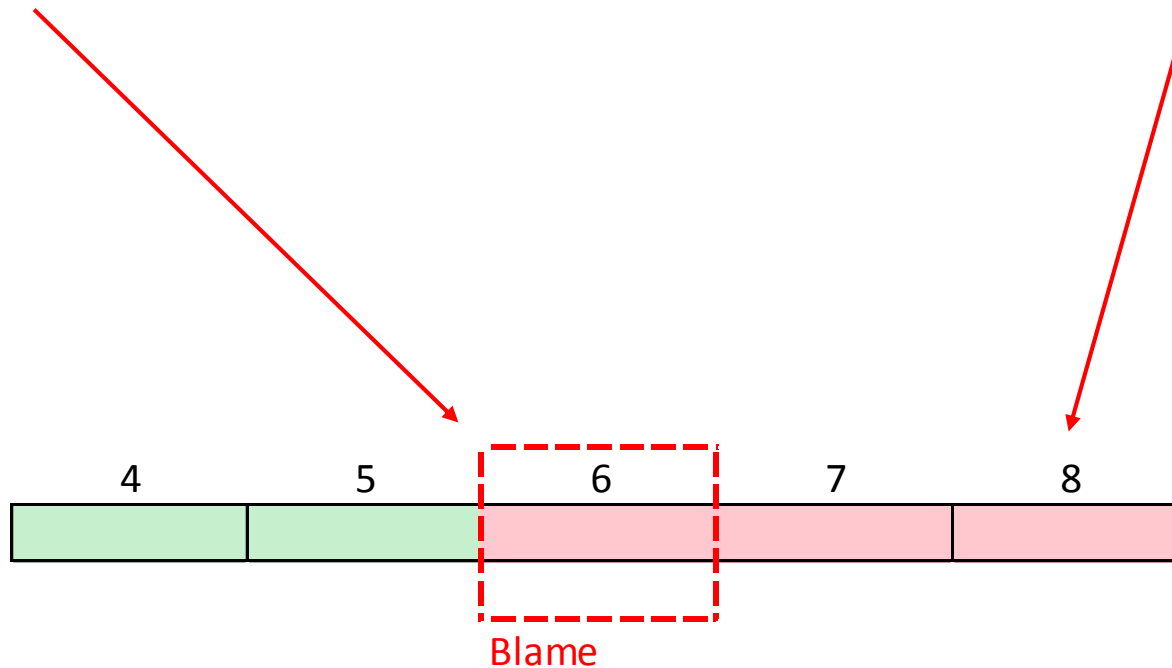
“If we just test often enough then we will know who caused each regression failure”



Unfortunately this is totally unreliable!

# What We Measured

Is that the reason the test fails on the latest?

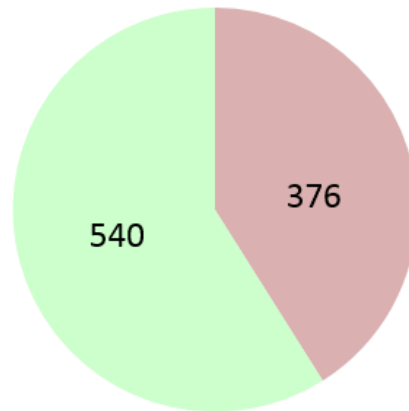


If so, undoing the change should make the test pass

# The Result

We checked 916 bug reports in ASIC projects

Bug Report Accuracy



In 41% of the cases the bug report was wrong!

Why is this simple approach so unreliable?

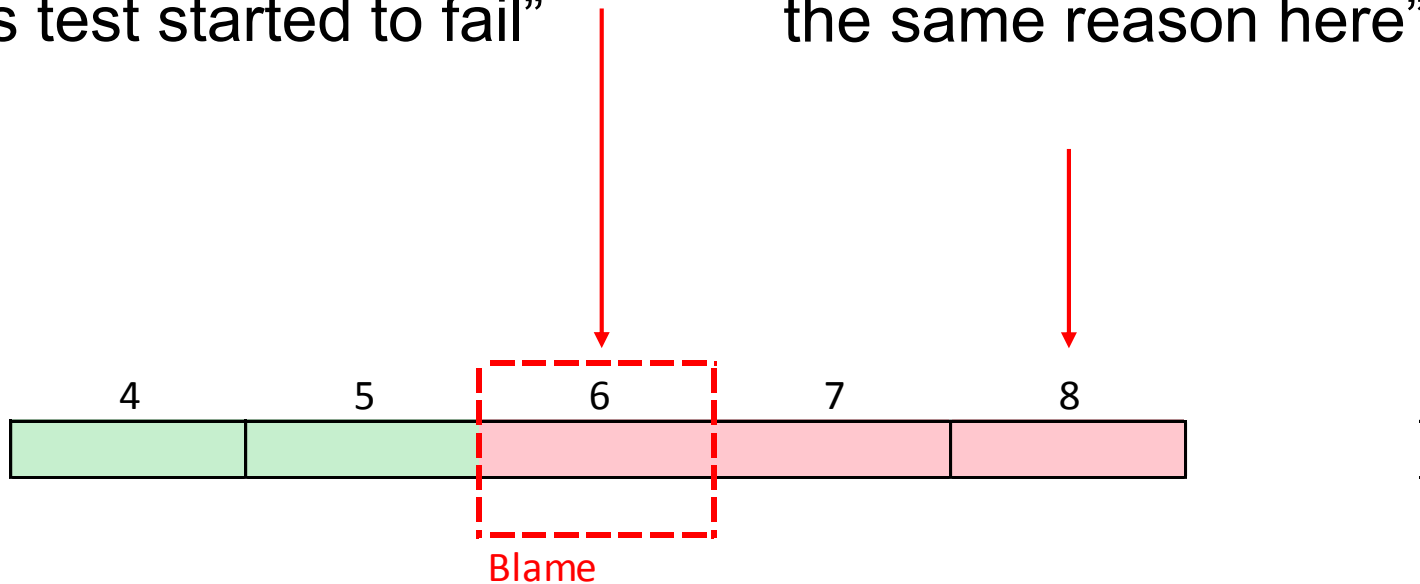
# Accurate/Inaccurate Blame

## Accurate Blame

“This commit was the reason  
this test started to fail”

## Inaccurate Blame

“The test must still fail for  
the same reason here”



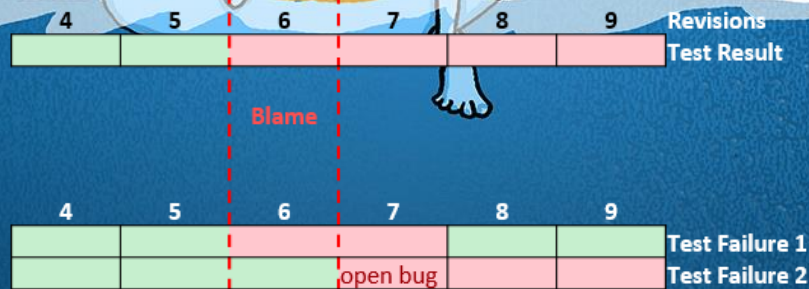
A lot of things happen after the first failure

# Many Scenarios

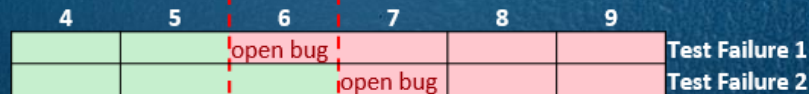
## Inaccurate Blame

41%

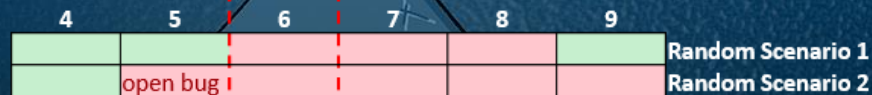
### Overlapping Failures 1



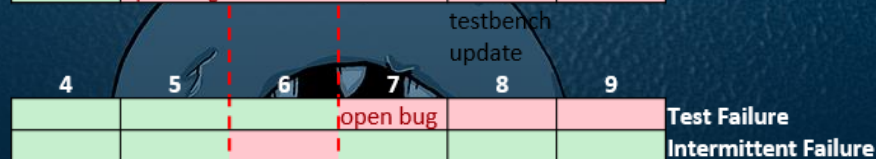
### Overlapping Failures 2



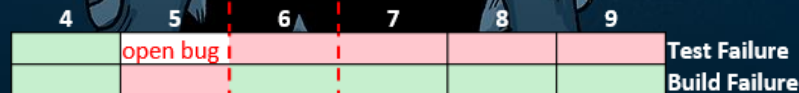
### Random Instability



### Intermittent Failures



### Build Failures





# Two Solutions

## 1. Prevent Complexity

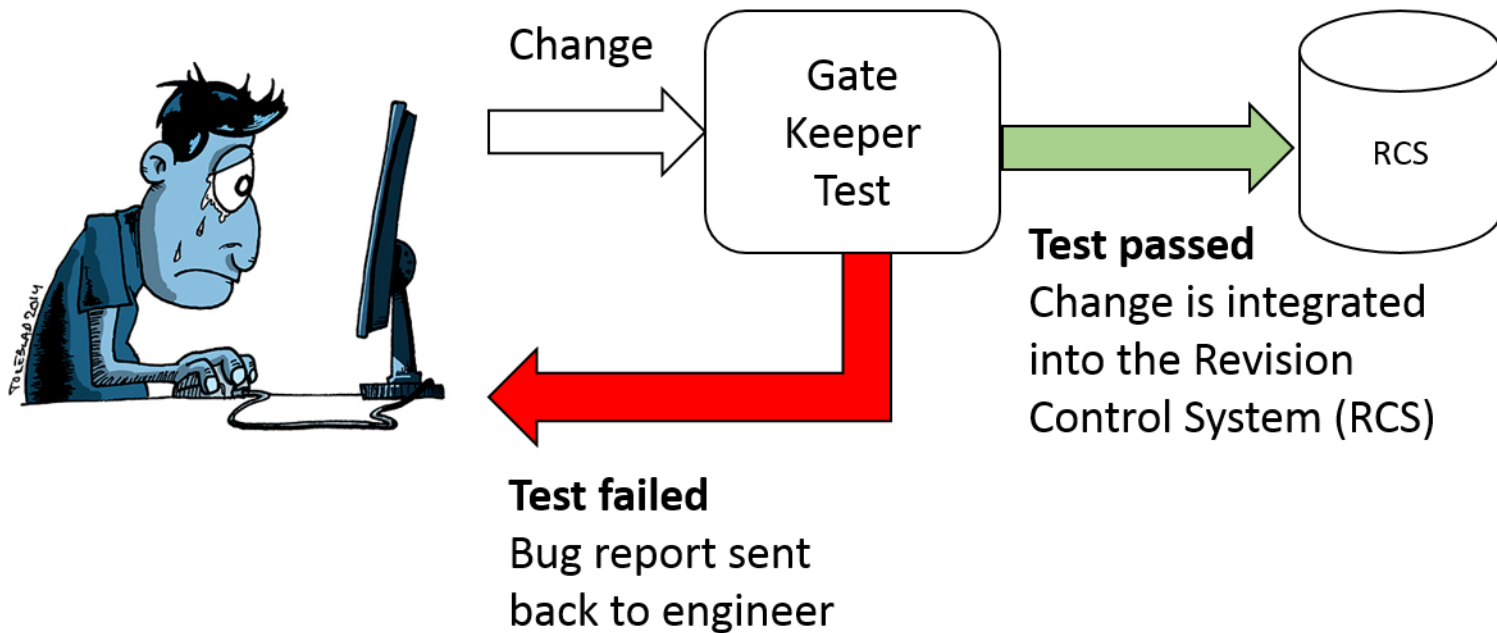
- Test one commit at a time
- Strict Continuous Integration

## 2. Handle Complexity

- You must debug correctly all complex scenarios
- PinDown, automatic debugger of regression failures

# Strict Continuous Integration

Test each commit. Only let it in if the tests pass



Perfect if complete test suite takes minutes and does not contain random tests. Very popular in the software industry.

# Test & Debug Framework

PinDown TestHub

testbed

Test Results Open Bugs Diagnose Settings

| Config   | Test         | maj 11 2012 8:51 | maj 11 2012 8:52 | maj 11 2012 8:52 |
|----------|--------------|------------------|------------------|------------------|
| config_1 | Build Result |                  |                  |                  |
| config_1 | t1           |                  | 2                |                  |
| config_1 | t2           | 1                |                  |                  |
| config_1 | t3           |                  |                  | 3                |
| config_2 | Build Result |                  |                  |                  |
| config_2 | t1           | 1                |                  |                  |
| config_2 | t2           |                  |                  |                  |
| config_2 | t3           |                  |                  |                  |



Email Bug Reports

From: 2010-12-21 To: 2012-02-14 Messages, authors and paths

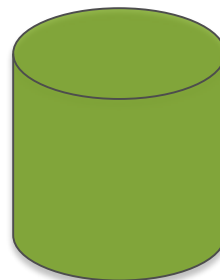
| Revision | Actions | Author | Date                           | Message   |
|----------|---------|--------|--------------------------------|---|
| 23       |         | daniel | 10:50:27, den 14 februari 2012 | Inserted tool change.                           |
| 22       |         | daniel | 10:46:42, den 14 februari 2012 | common chedon both branches                     |
| 21       |         | daniel | 13:25:44, den 21 december 2010 | Commented out redundant line. Don't think it is |
| 19       |         | daniel | 13:24:07, den 21 december 2010 | Combined generation of variables e and f as the |
| 16       |         | daniel | 13:21:41, den 21 december 2010 | Added assignment of f.                          |
| 15       |         | daniel | 13:20:52, den 21 december 2010 | Added assignment of e.                          |
| 12       |         | daniel | 13:19:22, den 21 december 2010 | Added assignment to c.                          |
| 11       |         | daniel | 13:18:50, den 21 december 2010 | Fixed bug in assignment of b. Should be 1.      |
| 8        |         | daniel | 13:17:33, den 21 december 2010 | Added assignment to b.                          |
| 7        |         | daniel | 13:17:03, den 21 december 2010 | Added init                                      |
| 5        |         | daniel | 13:15:37, den 21 december 2010 | Added repos, first versions.                    |

Inserted tool change.

Action Path Copy from path Revision

Modified /trunk/zazaam\_beta/ileB.txt

Version Control System



PinDown Results Database

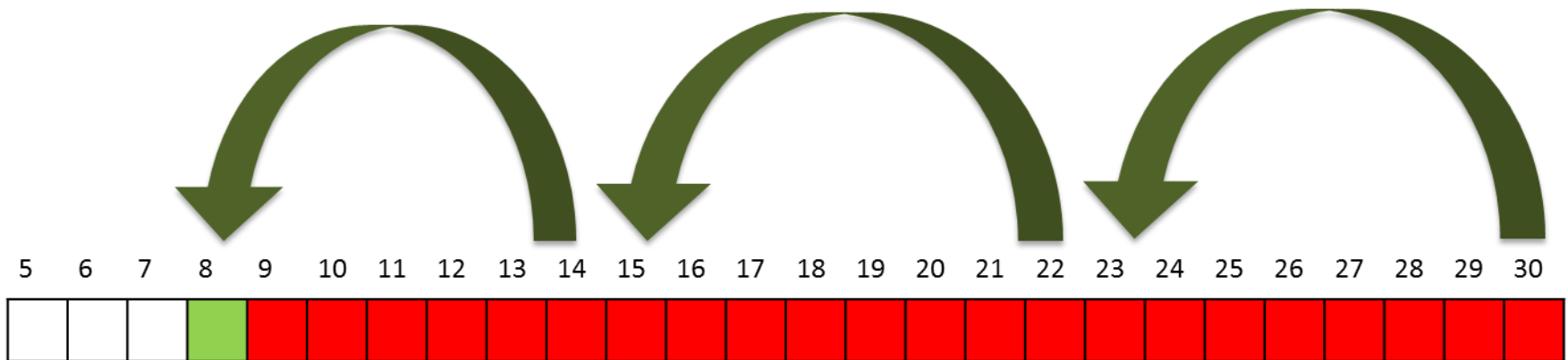


Test Executor  
(LSF Farm)



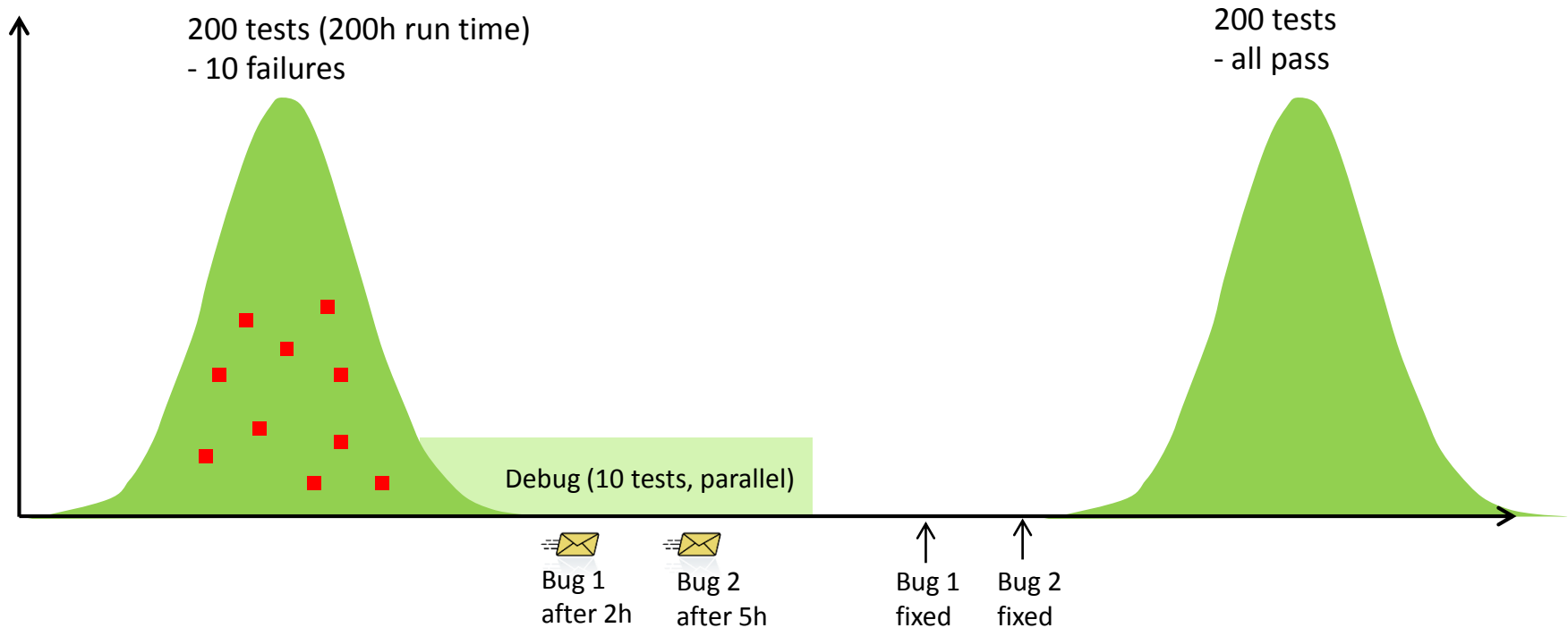
# Rerun on Old Revisions

Rerunning tests on older revisions to find bad commit



- Only takes 1 – 4 reruns of the failing test to find culprit
- Combines revisions in order to analyze complex scenarios
- This is only done for one test per bug

# Farm Usage



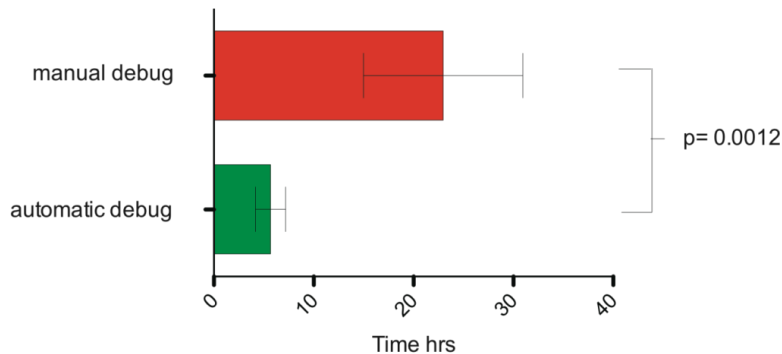
- Typical cost for re-running tests: 10 jobs on the farm
- Good investment to avoid next test run to show same state

# Benefits

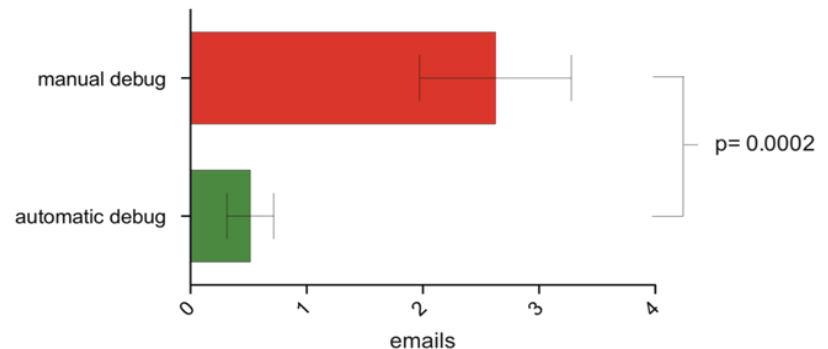
400% Faster Fixes, 5x Less Discussion, to 11% Shorter Projects



Bug Fix Time

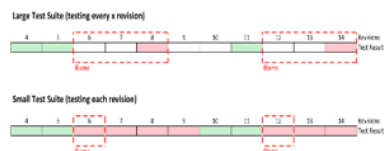


Email Discussions Regarding Bugs

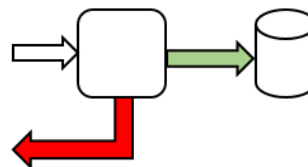


# Summary

## Standard Regression Testing



## Strict Continuous Integration



## PinDown

PinDown

| Test     | Results      | Open Stage       | Progress         | Settings         |
|----------|--------------|------------------|------------------|------------------|
| conf_1   | Test         | Aug 11 2013 8:10 | Aug 11 2013 8:10 | Aug 11 2013 8:10 |
| conf_1_1 | Build Result |                  |                  |                  |
| conf_1_1 | 10           |                  |                  |                  |
| conf_1_1 | 10           |                  |                  |                  |
| conf_1_1 | 10           |                  |                  |                  |
| conf_1_1 | Build Result |                  |                  |                  |
| conf_1_1 | 10           |                  |                  |                  |
| conf_1_1 | 10           |                  |                  |                  |
| conf_1_1 | 10           |                  |                  |                  |

**Bad Commit Known**

No

Yes

Yes

**Benefits**

Automatic Triage

No

Yes

Yes

Bugs Fixed Faster

No

Yes

Yes

Less Debate

No

Yes

Yes

Shorter Project

No

Yes

Yes

**Support**

Large Test Suites

Yes

No

Yes

Random Tests

Yes

No

Yes

**Cost**

Debug Testing

Yes (manual)

No

Yes (automatic)