Faults in Linux:
10 years later
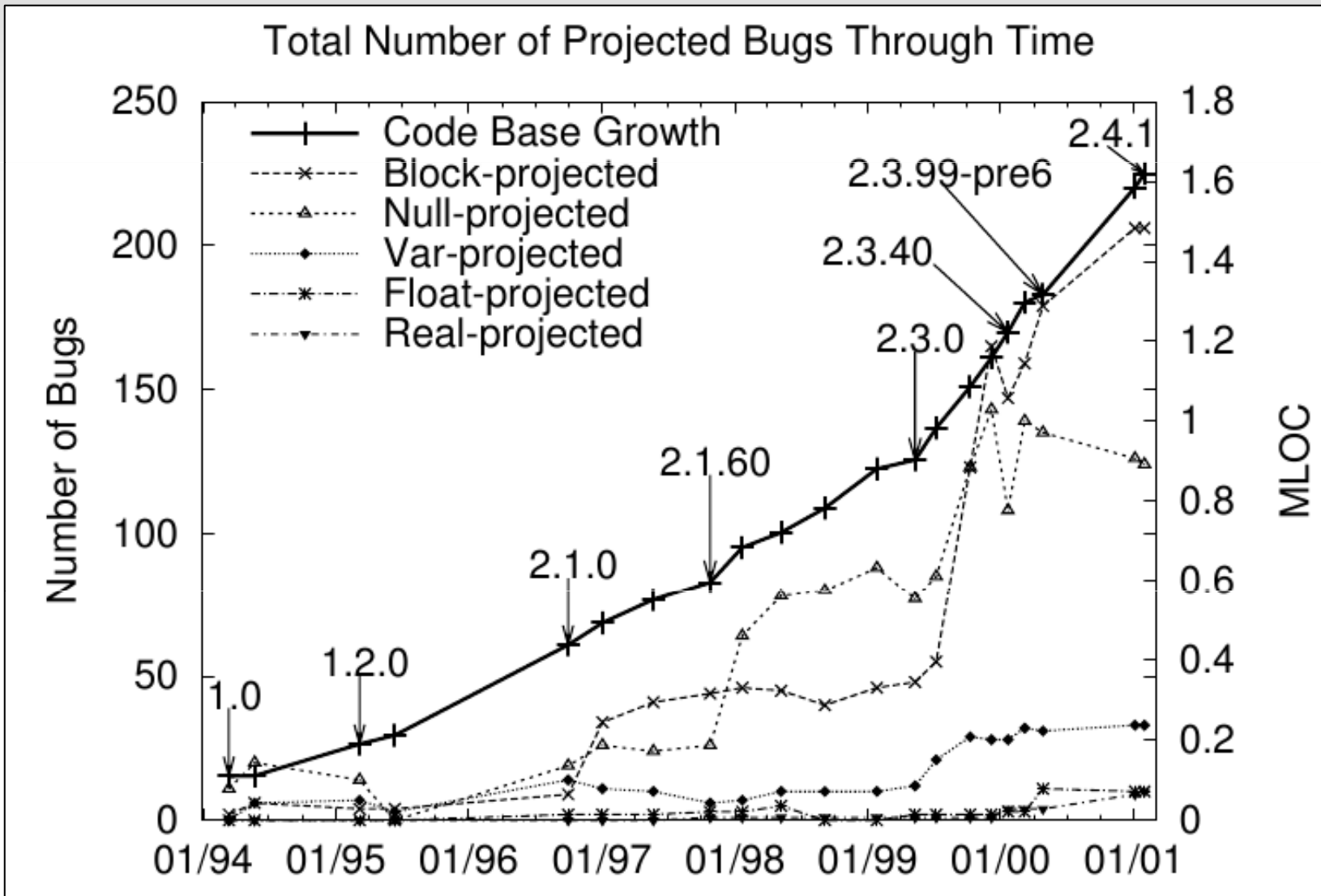
To be presented at ASPLOS 2011

Gilles Muller (INRIA/LIP6)

Nicolas Palix, Gaël Thomas, Suman Saha,
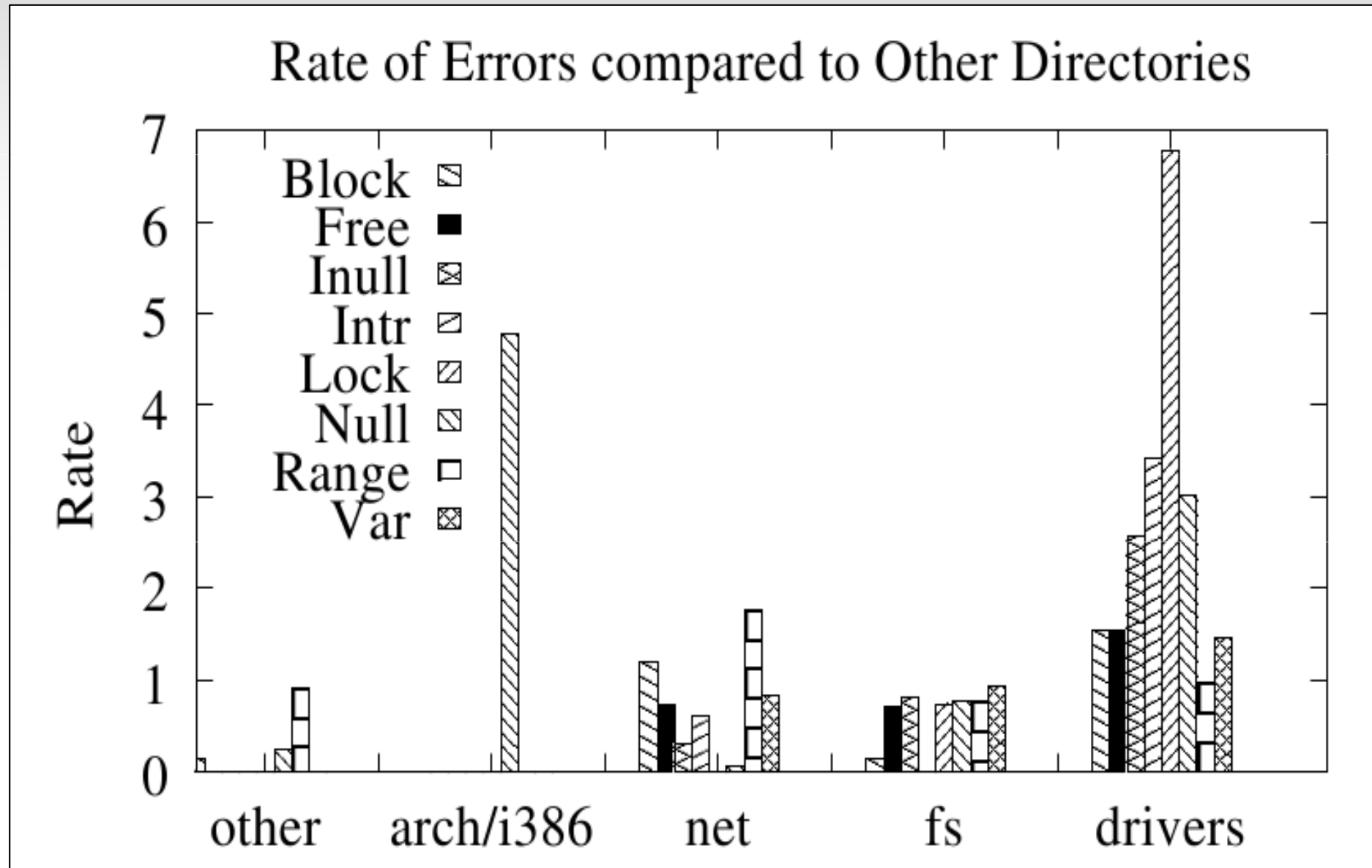Christophe Calvès, Julia Lawall

10 years ago, ...

- Chou et al. published a study on Linux at SOSP'01

- Study of Linux up to 2.4.1

- xgcc used to find faults (now Coverity)

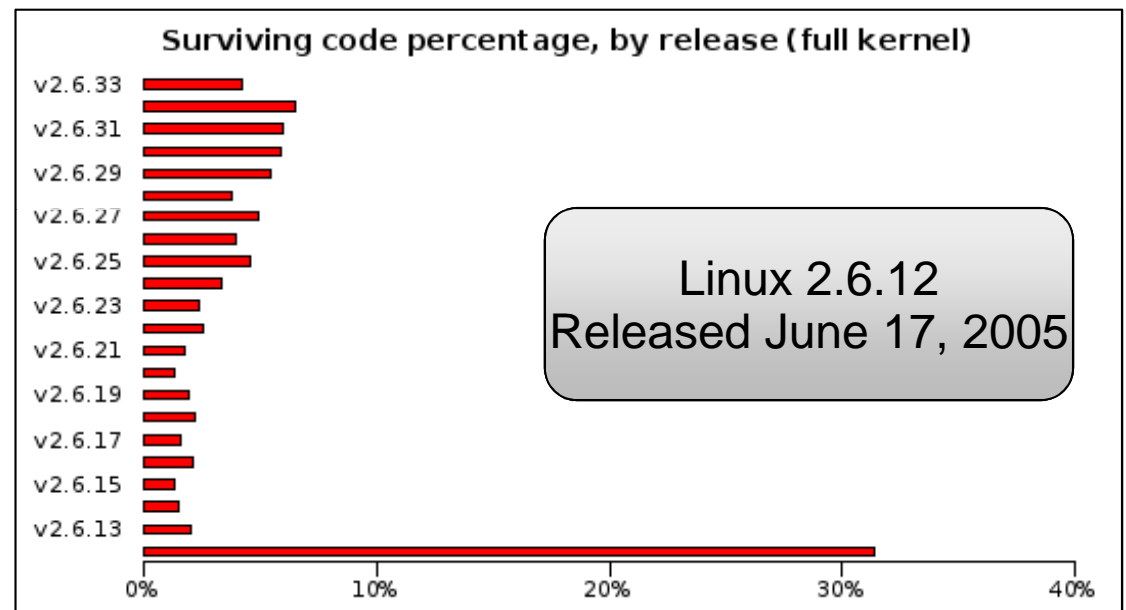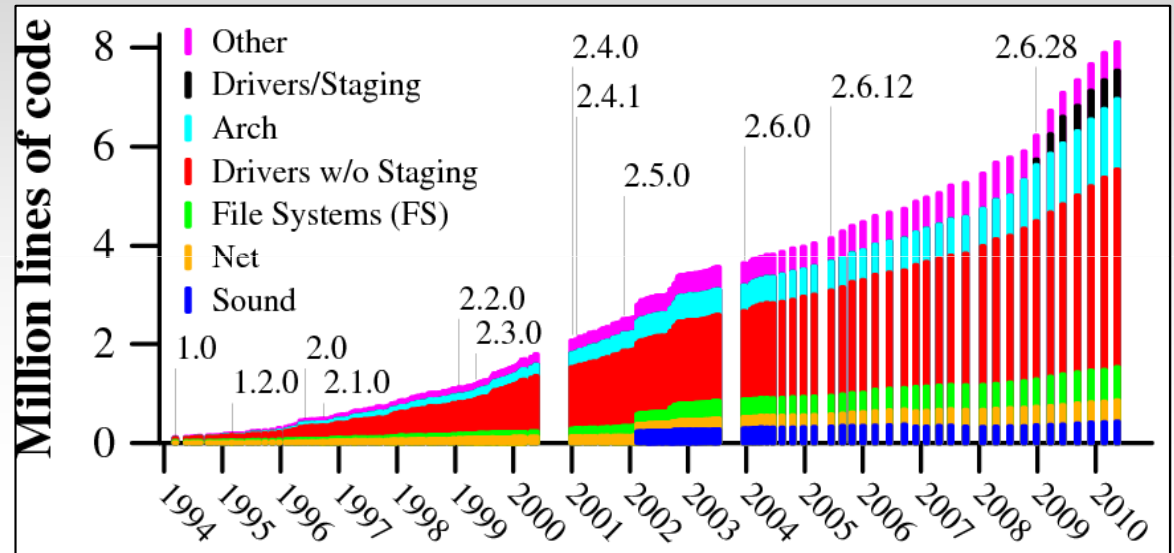- Up to 70% of code dedicated to drivers

… faults were rising! …

Total Number of Projected Bugs Through Time

...up to 7x more faults in drivers than any other directory

Rate of Errors compared to Other Directories

Today, ...

- Up to 8 MLOC
- Since v2.6.12
  - Git SCM
  - 69% of new code
- New directories
  - Sound, Staging
- New dev. model
  - 2.6.x vs 2.{4,5}

Linux 2.6.12
Released June 17, 2005

What about faults in Linux 2.6?

- Where should developers/researchers focus their attention?

- Are drivers still the Devil?
  - Where are the faults ?

- Fault kind trends?
  - Still rising, stalling, decreasing?

We Need New Data !!!

Our methodology

- Anyone, should be able to update the study
  - FLOSS tools
  - Coccinelle [Eurosys'08, POPL'09, DSN'09]
  - Herodotos [AOSD'10]

- Release fault definitions
  - Distribute Coccinelle SmPL rules

- Track faults through time
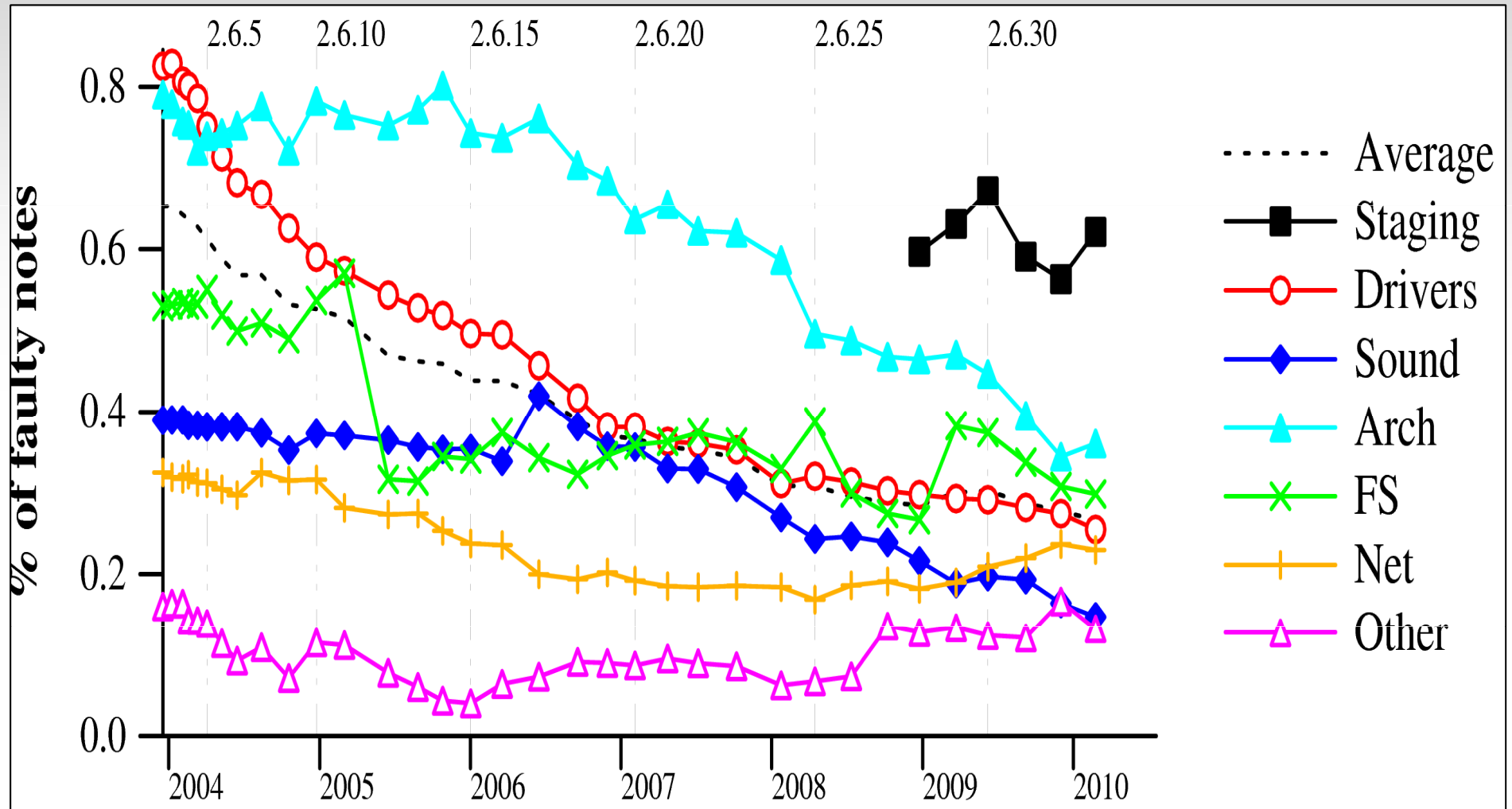 - Distribute correlation annotations

A few numbers...

- Study of Linux 2.4.1 and 34 versions of Linux 2.6
  - More than 170 MLOC analyzed
  - 697K files
  - 6.15M functions
- 83 SmPL scripts
  for notes, code pre-processing, fault finding
  - 4.44M notes
  - 4,815 correlated reports
  - 3,052 faults
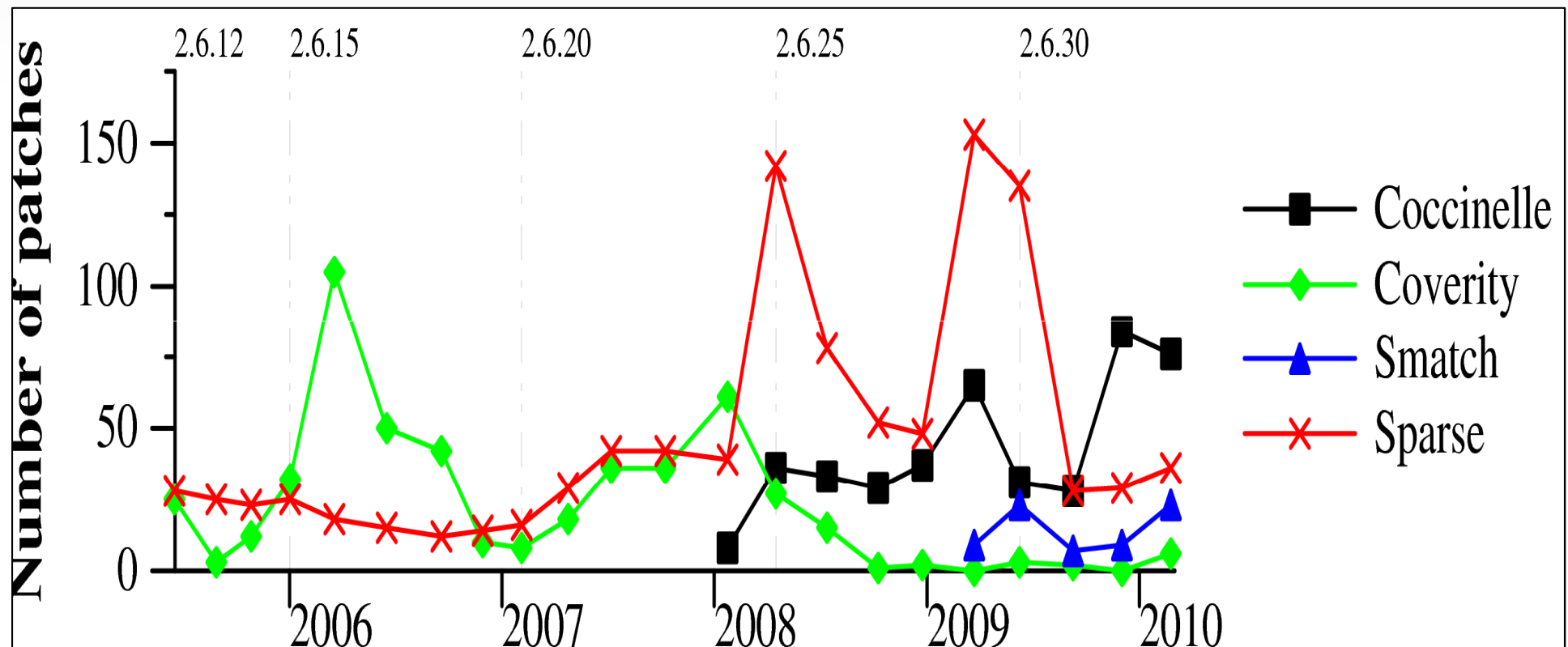
Fault in Linux 2.6
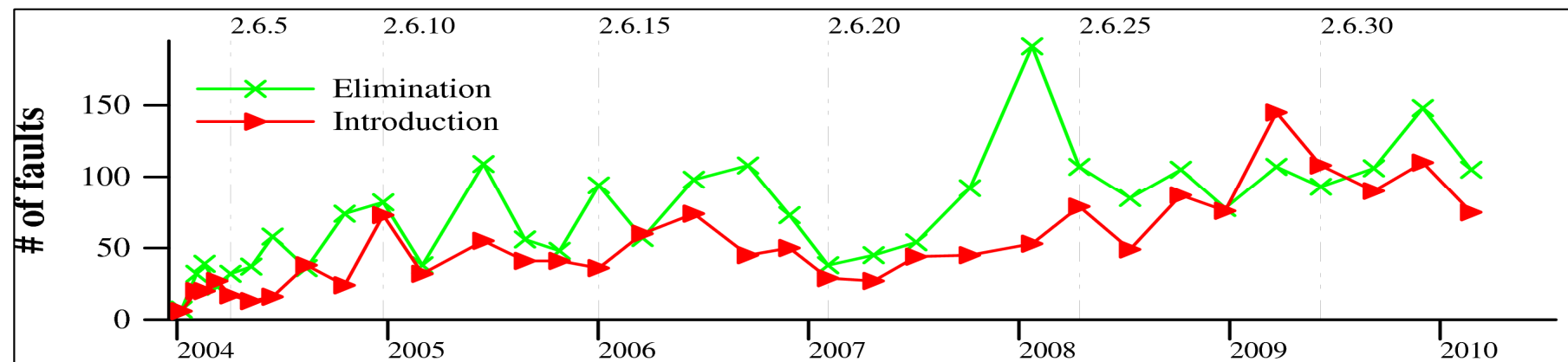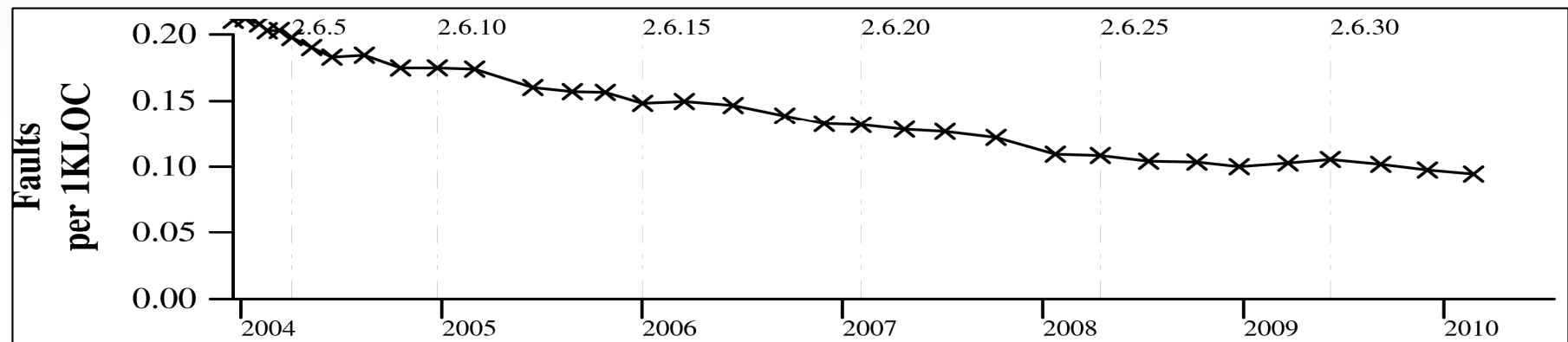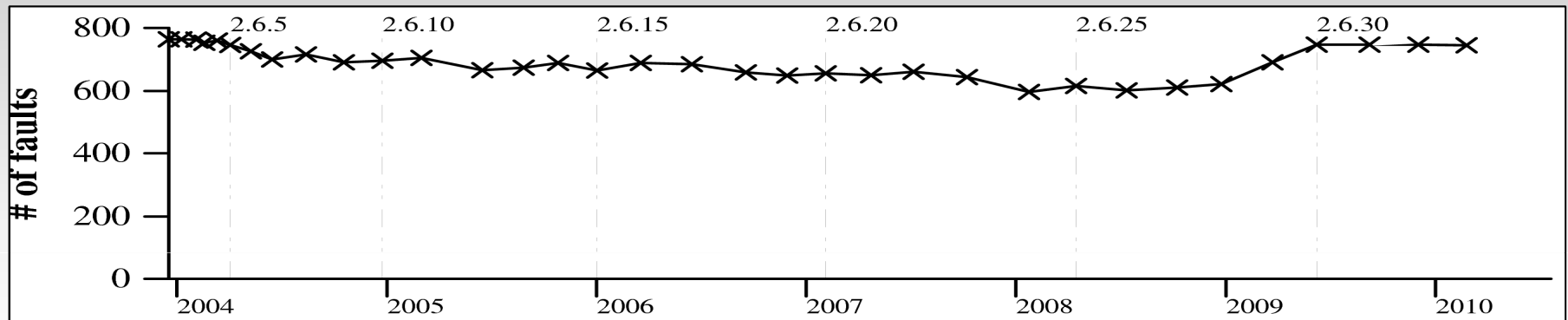
Fault rate (Per directory)

Rate = # Faults / # Notes

Tool users wanted!

- Still between 600 and 700 faults per version

- Tools are used, but not intensively

  - Not deeply integrated in the development process

Main results

Kill bugs before they hatch!!!

- Methodology
  - Based on FLOSS
  - Fault definitions in SmPL

- Linux 2.6

COCCINELLE
  - Overall quality is improving even if the code size is increasing
  - Drivers have improved but still have most of the faults
  - File-systems and arch (HAL) now have the worst fault rates

- Existing tools are under-exploited

Availability of the results

- Fault definitions and correlation annotations

http://faultlinux.lip6.fr/

- Database of reports

http://faultlinux.lip6.fr/phppgadmin/ (Browseable)

http://hal.inria.fr/inria-00509256/ [RR7357 annexes]

Availability of the tools

- Coccinelle

http://coccinelle.lip6.fr/

- Herodotos

http://coccinelle.lip6.fr/herodotos.html