

# Étalonnage de la sûreté de fonctionnement de systèmes d'exploitation

**Karama KANOUN**  
**LAAS-CNRS**



Club SEE “Systèmes Informatiques de Confiance”  
Atelier du 20 octobre 2005

Logiciels sur étagère (libres ou commerciaux) : quelle confiance leur accorder ?

# Software Systems Dependability Evaluation

- Information on software behavior
  - Field data
  - Data from development
  - Controlled experiments

Ad hoc

Standard

Dependability benchmarking

Evaluation of dependability measures / features  
in a non-ambiguous way → comparison

**Properties**

**Representativeness, repeatability, portability, acceptable cost/effort**

# Propriétés d'un étalon de SdF

## ➤ Répétitivité

À partir de plusieurs exécutions d'un prototype, on obtient des résultats statistiquement équivalents

## ➤ Reproductibilité

À partir de plusieurs implémentations de la même spécification, on obtient des résultats statistiquement équivalents

## ➤ Représentativité

Mesures et profils d'exécution acceptés par les utilisateurs

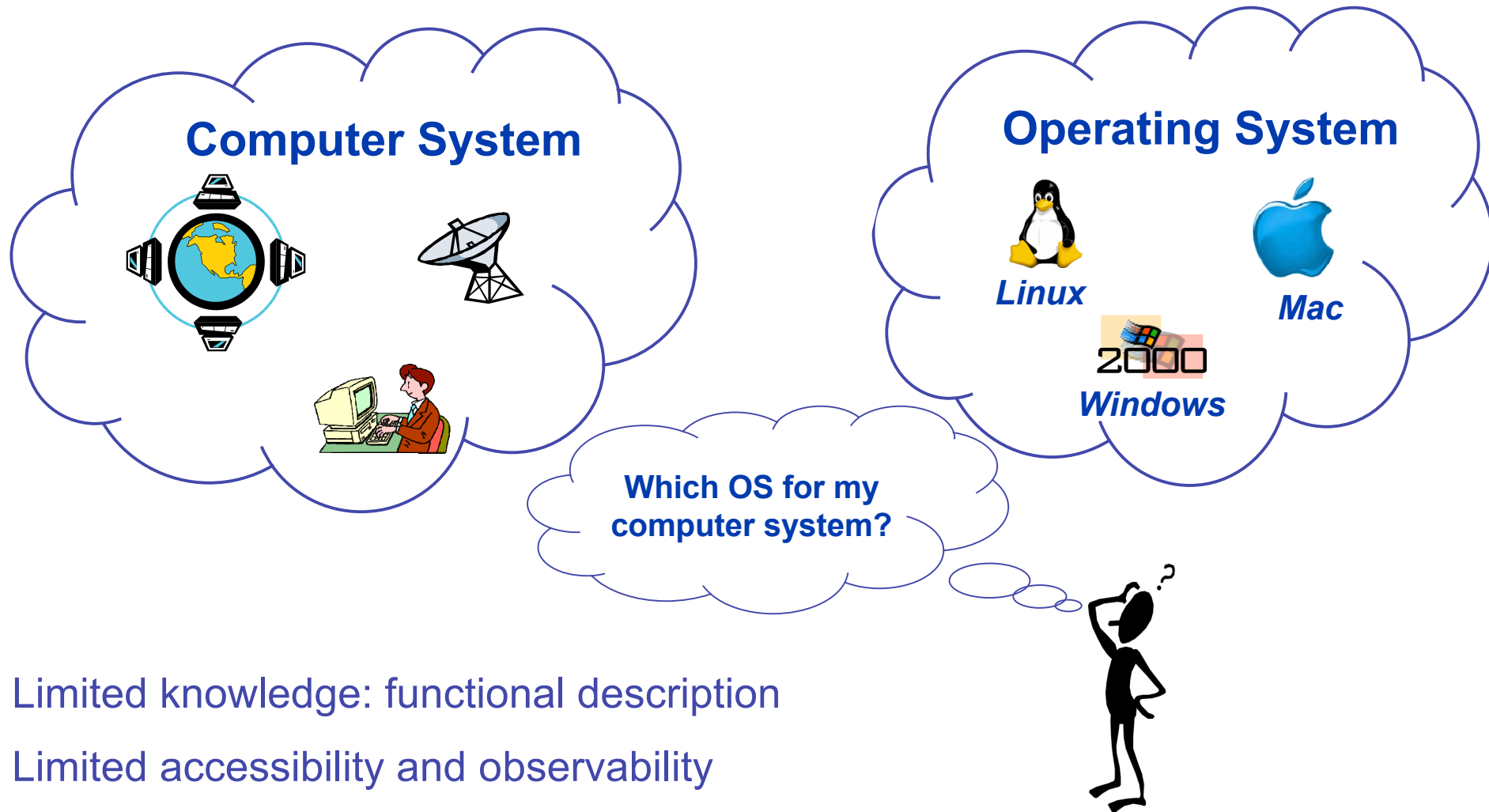
## ➤ Portabilité

Applicabilité à différents systèmes de la même catégorie

## ➤ Coût/effort acceptables

Efforts nécessaires pour développer et exécuter l'étalon

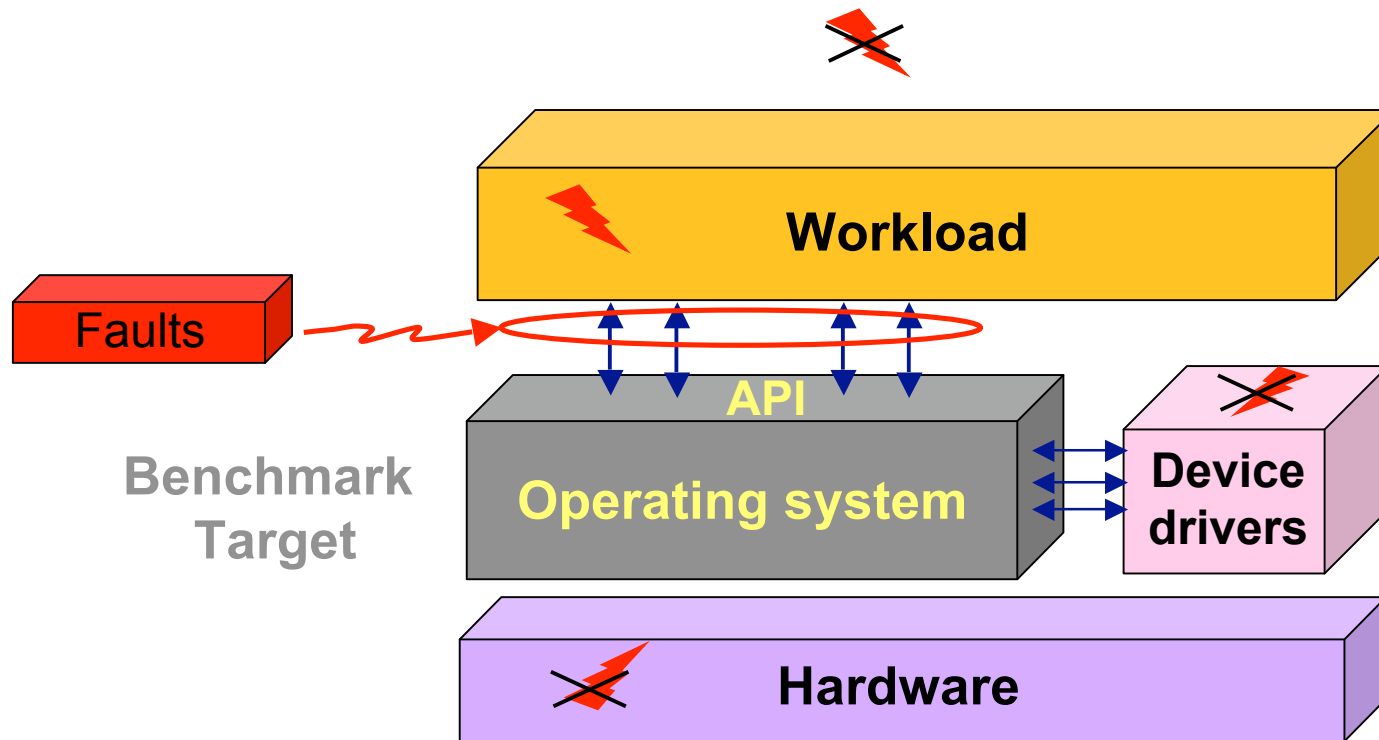
# COTS Benchmarking: User Point of View



- Limited knowledge: functional description
- Limited accessibility and observability
- Limited intrusiveness and interference

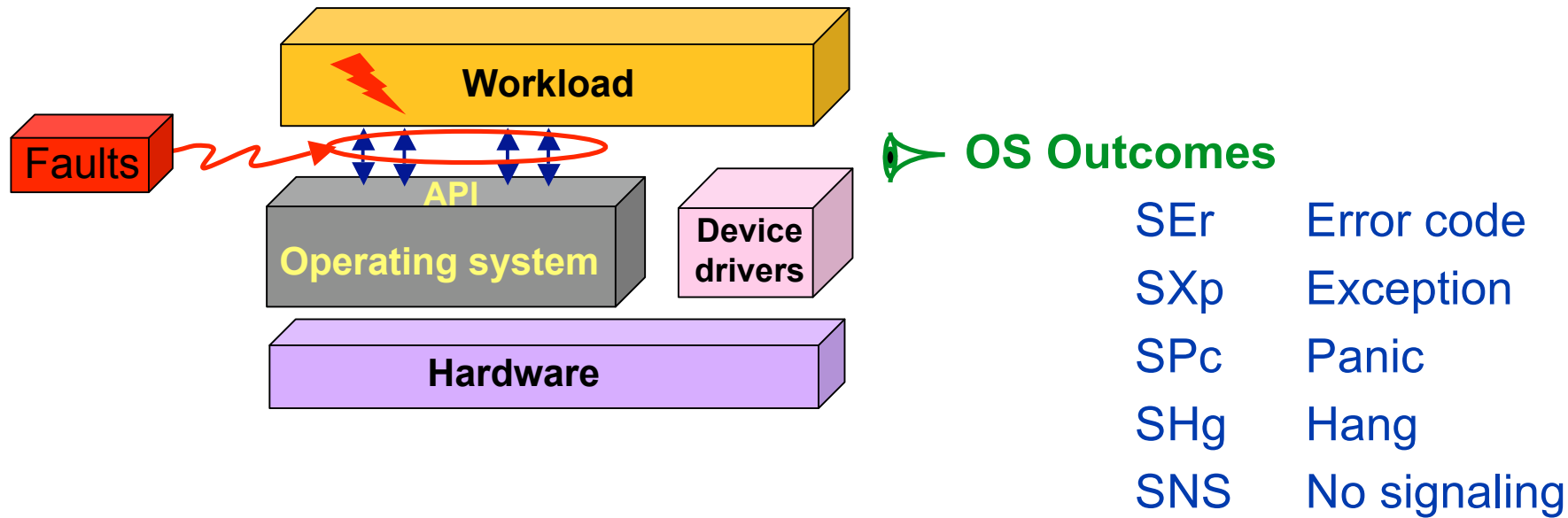
⇒ **Black-box approach** ⇒ **robustness benchmark**

# Benchmarking wrt class of faults?



Wrt application erroneous behavior

# OS Benchmark & Measures

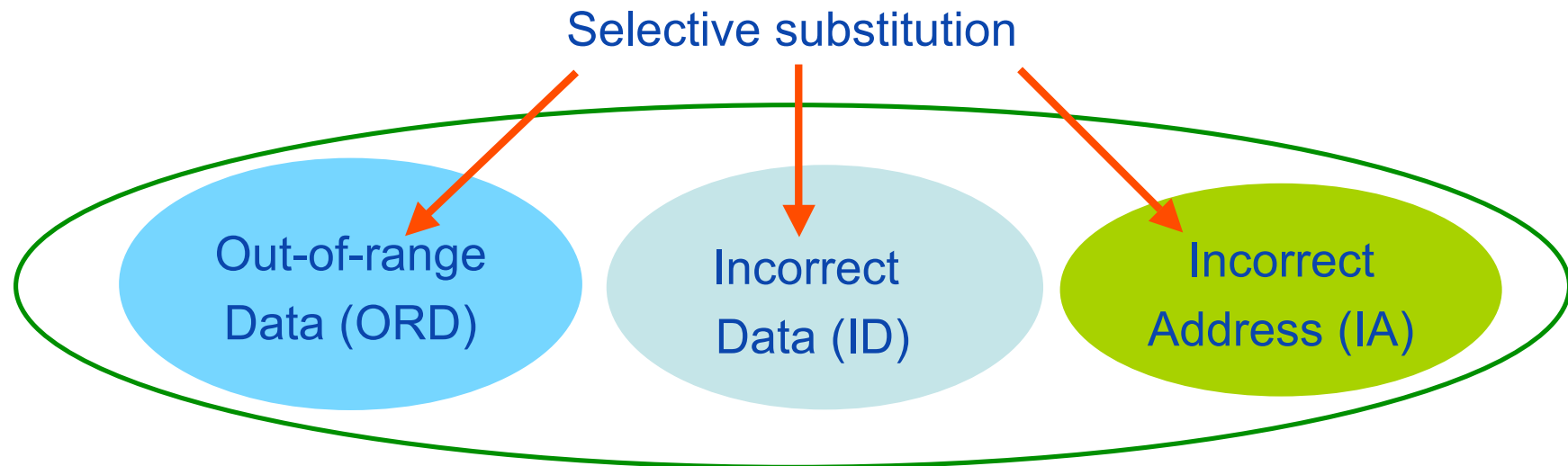


## Measures

- POS: OS Robustness [%SEr %SXP %SPc %SHg %SNS] )
- Texec: OS reaction time in the presence of faults
- Tres: OS Restart time after fault insertion

# Execution Profile

- Workload
  - TPC-C Client, Java Virtual Machine, **PostMark**
- Faultload
  - Corruption of parameters of all system calls



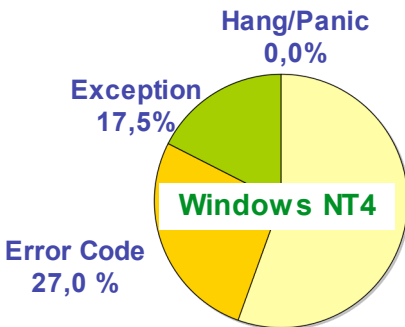
## Dependability Benchmarks with PostMark WL

	# system calls	# experiments
Windows NT 4	25	418
Windows 2000	25 + 1 + 1	433
Windows XP	25 + 1	424
Windows NT 4 Server	25	418
Windows 2000 Server	25 + 1 + 1	433
Windows 2003 Server	25 + 1 + 1	433

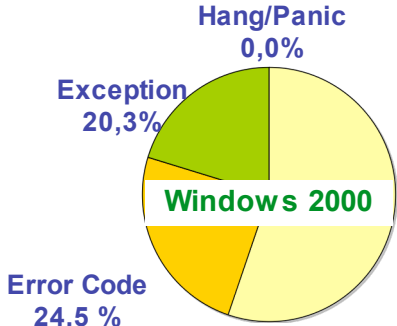
Linux 2.2.26	15 + 1	206
Linux 2.4.5	15 + 1	206
Linux 2.4.26	15 + 1	206
Linux 2.6.6	15 + 2	228

# Robustness (WL = PostMark)

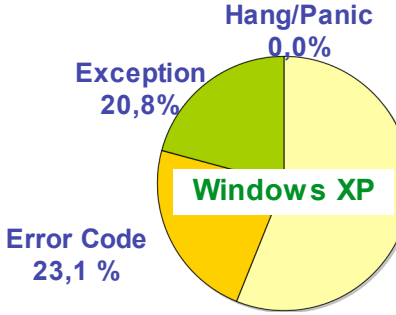
## Windows



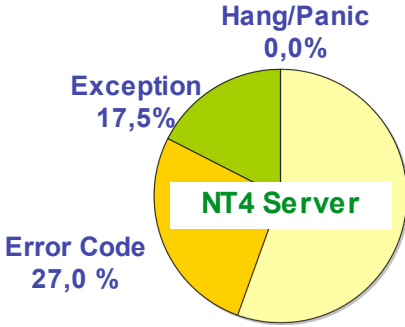
No Signaling  
55,5%



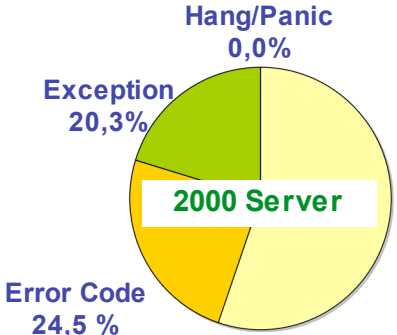
No Signaling  
55,2%



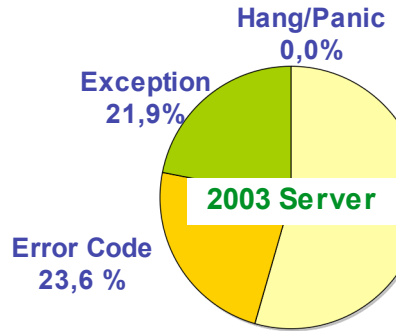
No Signaling  
56,1%



No Signaling  
55,5%

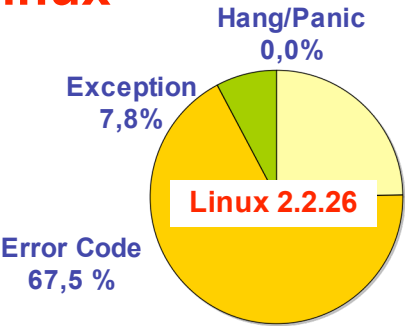


No Signaling  
55,2%

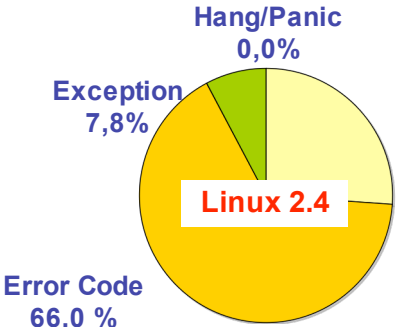


No Signaling  
54,5%

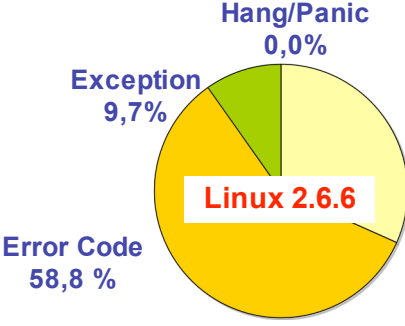
## Linux



No Signaling  
24,8%



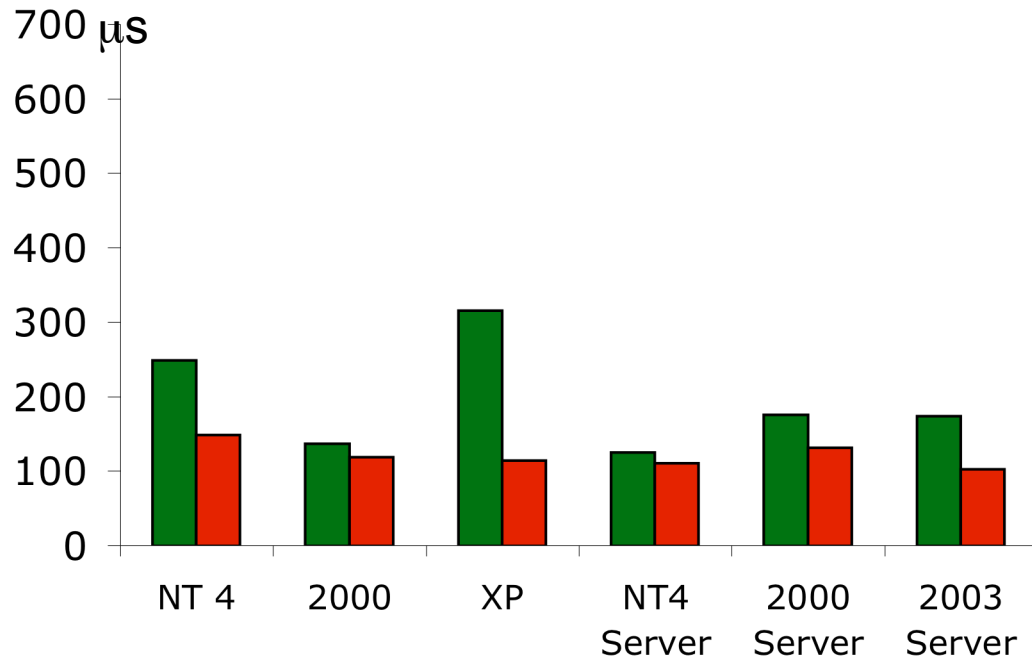
No Signaling  
26,2%



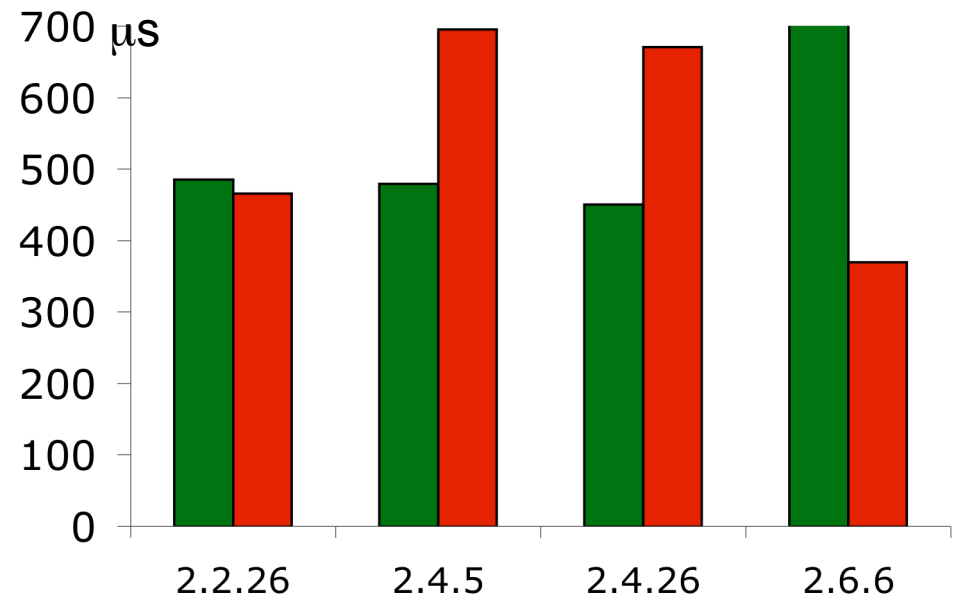
No Signaling  
31,6%

# OS Reaction Time (WL = PostMark)

## Windows



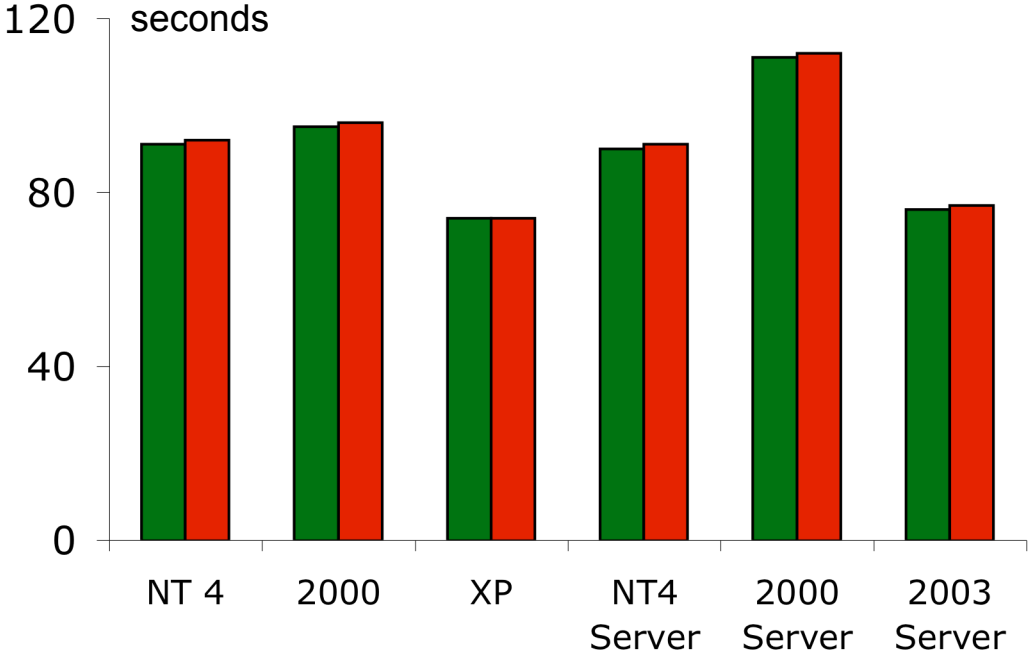
## Linux



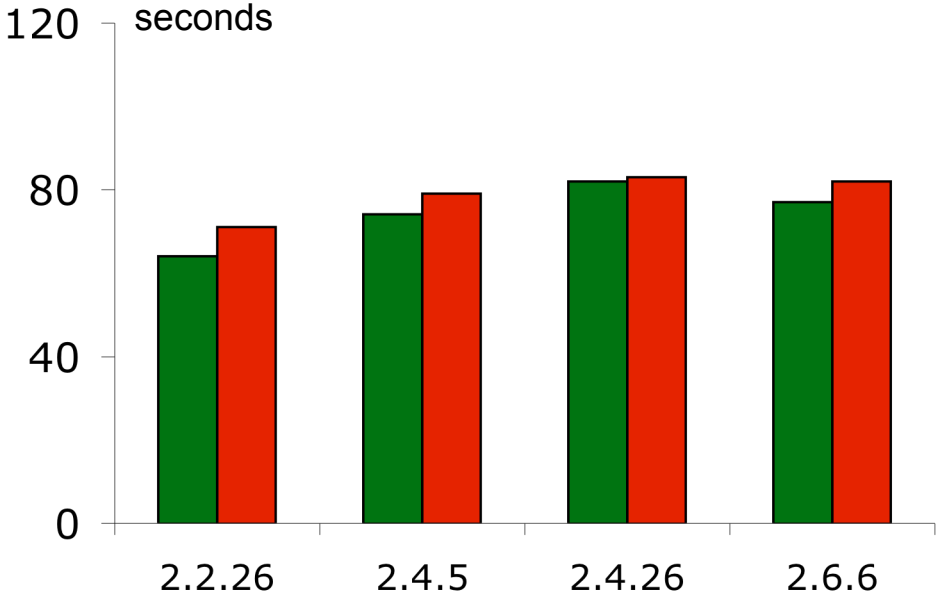
- In the presence of faults
- Without parameter corruption

# Restart Time (WL = PostMark)

## Windows



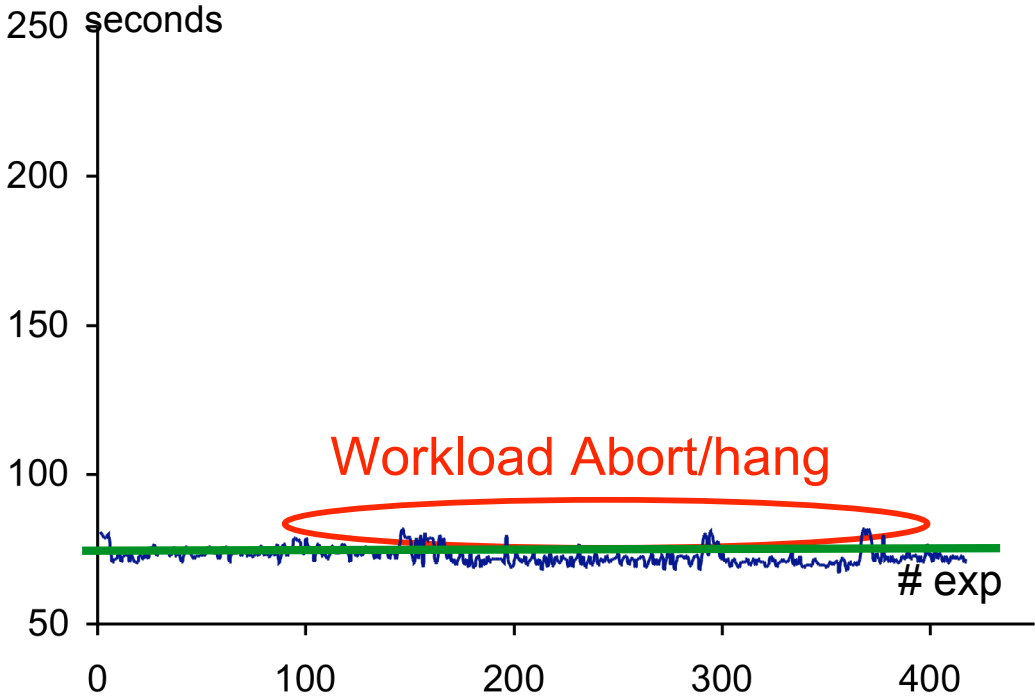
## Linux



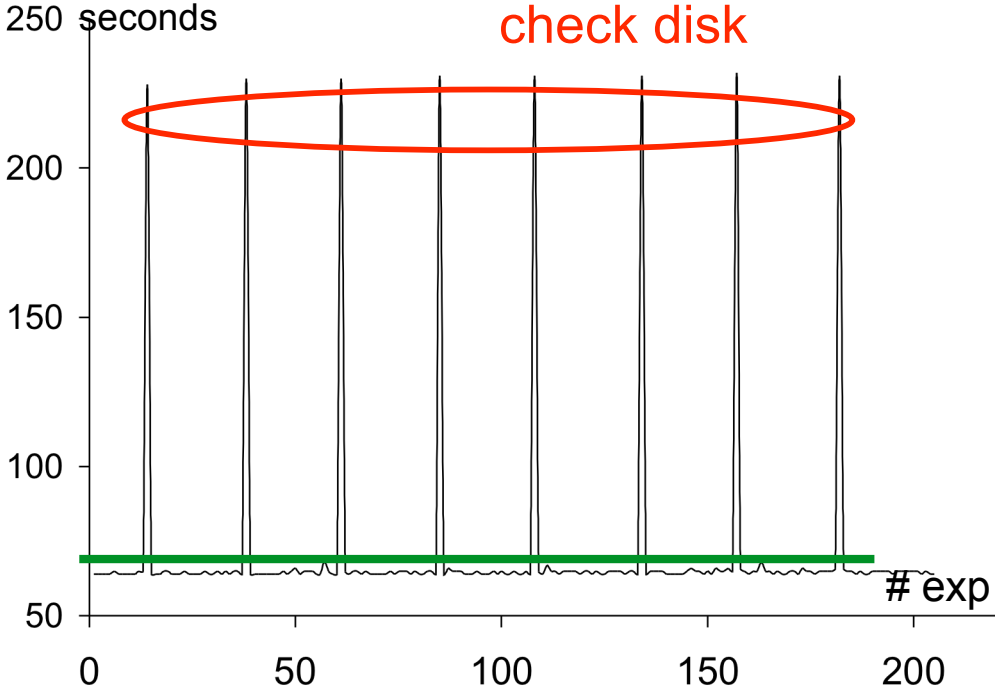
- In the presence of faults
- Without parameter corruption

# Restart Time (WL = PostMark)

## Windows XP



## Linux 2.2.26



# Validation des propriétés

## ➤ Répétitivité

- ✓ Chaque étalon a été ré-exécuté trois fois
  - Mesures de robustesse exactement identiques
  - Variation de temps de réaction (< 4% pour TPC-C client)
  - Variation de temps de redémarrage (< 3% pour TPC-C client)

## ➤ Reproductibilité

- ✓ Par construction
- ✓ Ensemble de fautes
  - Appels système à corrompre
  - Valeurs de substitution

# Validation - Représentativité

## ➤ Mesures

Mesures de grande importance pour les concepteurs des systèmes informatiques

- Mesures de robustesse et de performance en présence de fautes
- Peuvent être raffinées pour mieux comprendre le comportement du système

## ➤ Activités

Très utilisées dans la vie réelle (JVM) et dans le domaine d'étalonnage de performance (TPC-C et Postmark)

## ➤ Fautes

- ✓ Comparaison des résultats de substitution sélective et de substitution systématique par inversion de bit
- ✓ Etude de sensibilité des résultats par rapport à la technique de corruption des paramètres

## Validation - Sensitivity Analyses wrt Faultload

	Parameter corruption type			# experiments	
	Incorrect Data	Incorrect Address	Out-of-range Data	Windows NT4	Linux
Faultload 0	x	x	x	418	206
Faultload 1		x	x	331	135
Faultload 2			x	77	55

- Equivalence of versions of the same family
- Same comparison results between the two families

Additional analysis: incorrect data = out-of-range data in the context

## Validation - Coût/effort

### ➤ Durée d'implémentation d'un étalon

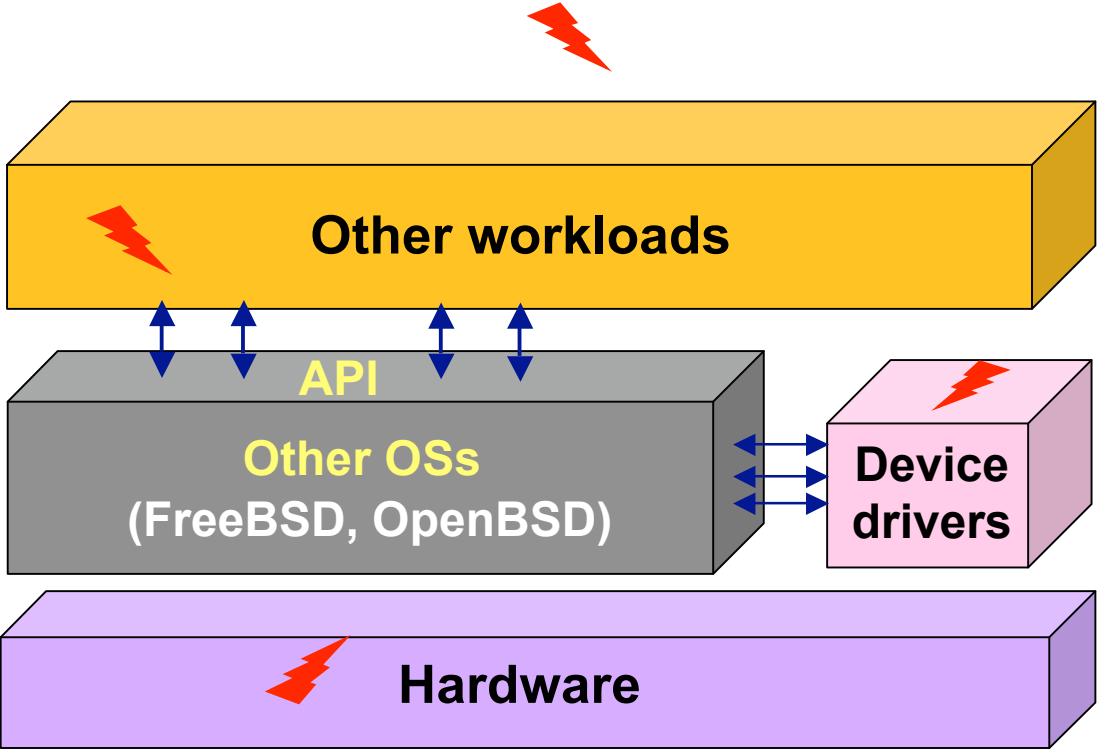
Pour chaque famille d'OSs :

- ✓ 1 à 3 jours pour l'installation de l'activité
- ✓ 2 semaines pour le développement du contrôleur, des mécanismes de corruption de paramètres et d'observation
- ✓ 1 semaine pour la définition et l'implémentation de l'ensemble de valeurs de corruption

### ➤ Durée d'expérimentation

	Windows	Linux
TPC-C client	2 jours	1 jour
Postmark	2 jours	1 jour
JVM	4 jours	2 jours

# Extensions



## Projet européen DBench (<http://www.laas.fr/DBench/>)

- General-purpose operating systems
  - Robustness and timing measures, TPC-C Client, faulty application
- Real-Time kernels in onboard space systems
  - Predictability of the kernel response time, faulty application
- Engine control applications in automotive systems
  - Impact of application failures on system safety, transient hardware faults
- On-line transaction processing (OLTP) environments
  - TPC-C-based, operator, software & hardware faults
    - TPC-C like measures, DBench-OLTP
    - Measures based on experimentation & modelling, TPC-C-Depend
  - Web-servers, SPEC-based, operator, software & hardware faults