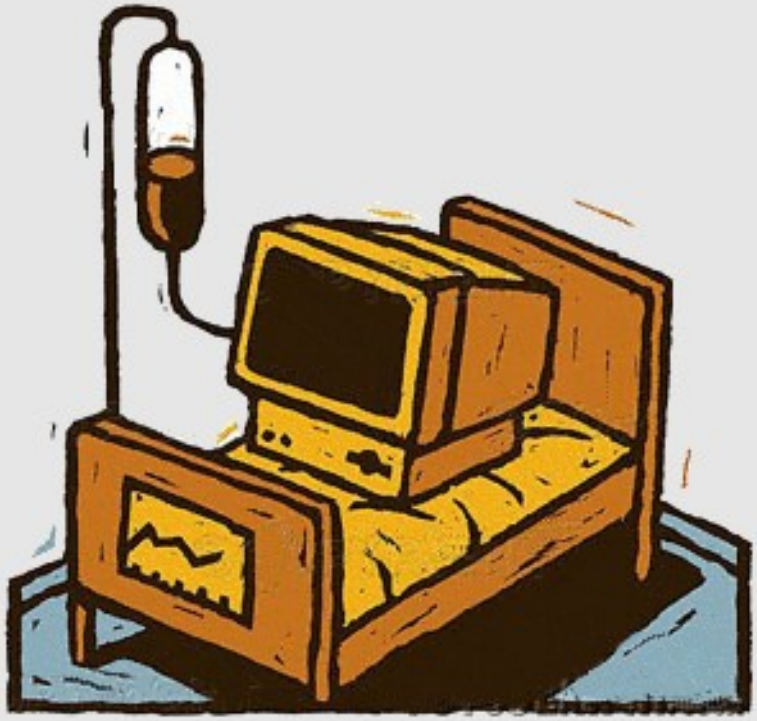
STAYING AWAKE ON FRIDAY AFTERNOON

Interactive Session on MINIX 3

ASCI GNARP workshop
March 17, 2006 – Rockanje

Jorrit N. Herder
Dept. of Computer Science
Vrije Universiteit Amsterdam



QUIZ

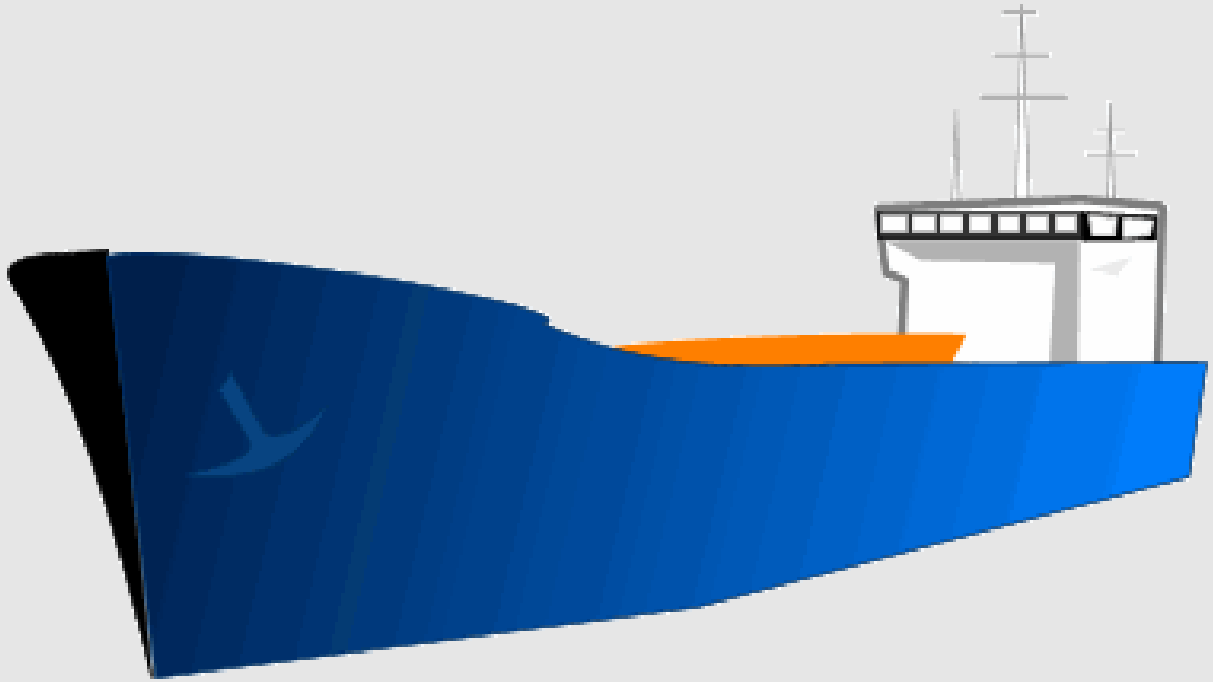
- What operating systems do you use?
- Which one is more dependable?
- Why?
- Would you install my nifty kernel module?
- How about device drivers?
- What other threats do you see?
- How much performance would you be willing to sacrifice?

WHERE DID IT GO WRONG?

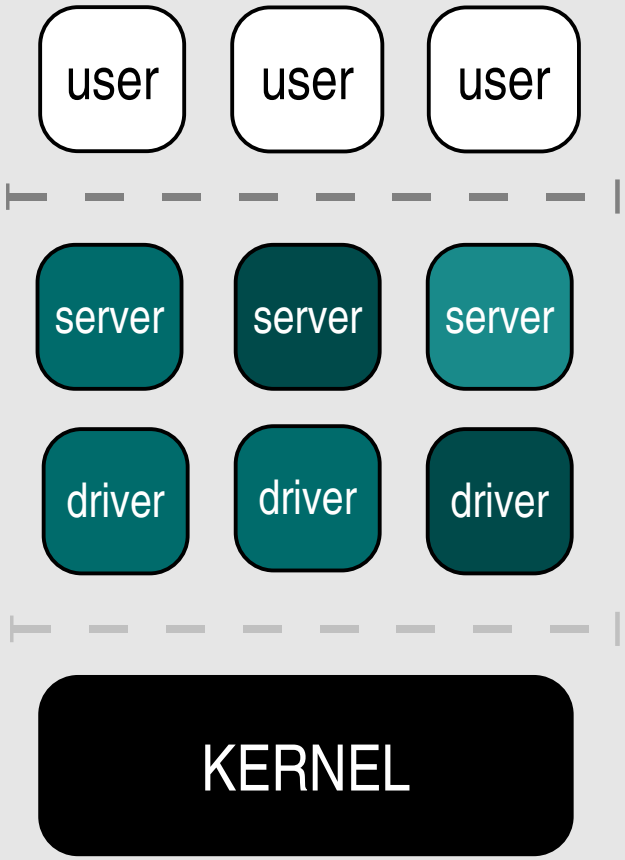
- Historically, design was guided by performance
- Fundamental design flaws in monolithic kernels
  - All code runs at highest privilege level (breaches POLA)
  - No proper fault isolation (any bug can be fatal)
  - Huge amount of code *in* kernel (1-20 bugs per 1000 LOC)
  - Untrusted, 3rd party code in kernel (70% driver bugs)
  - Complex and hard to maintain (causes bugs)
- Trade-off between performance and dependability
- New, modular design can solve above problems

DRAWING EXERCISE

- Volunteer requested to draw the intersection of a ship



DESIGN FOR DEPENDABILITY



- MINIX 3
  - Reliability
  - Availability
  - Security

CHARACTERISTICS OF MINIX 3

- Minimal kernel to support user-mode operating system
  - Stable kernel (~4000 LoC) reduces number of fatal bugs
- User-mode modules are physically isolated by MMU
  - Memory access must be explicitly granted by other party
- Privileges of each components are strongly restricted
  - Kernel policies for IPC, kernel calls, I/O, memory, scheduling
- Servers and drivers are carefully monitored
  - Failures can be detected and often automatically repaired
- TCB is reduced by over two orders of magnitude
  - Minimal set of servers comprises about 20,000 LoC

HOW ABOUT SELF HEALING?

- Special server periodically does health check
- Death is immediately detected
- Policy driven healing process, for example:

  - Look up policy of sick driver
  - Write diagnosis to log
  - Send e-mail to administrator
  - Binary exponential backoff
  - Replace driver with fresh copy

MINIX 3: PERFORMANCE

- Time from boot monitor to login is under 5 sec
- Full build of minimal system within 4 sec
- Overhead for typical applications about 6%
- File system and disk I/O overhead about 9%
- Disk throughput with DMA up to 70 MB/sec
- Fast Ethernet easily runs at full speed
- Ethernet recovery every 4 sec costs 8%

SUMMARY & CONCLUSION

- Fundamental problems with monolithic systems
- Our approach to OS dependability: MINIX 3
  - Full compartmentalization
  - Principle of least authority
  - Fault tolerance and recovery
- Practical for wide-scale adoption
  - Lightweight approach
  - Backwards compatability
- Try it yourself: www.minix3.org

TIME FOR DISCUSSION

- The MINIX 3 team
  - Jorrit Herder
  - Ben Gras
  - Philip Homburg
  - Herbert Bos
  - Andy Tanenbaum
- More information
  - Web: www.minix3.org
  - News: comp.os.minix
  - Mail: jnherder@cs.vu.nl