Verification & Testing

Roderick Bloem
IAIK

Today

1. Administrative
2. Motivation

Administrative

Material & Communications

Physical lecture – no recordings

Webpage: https://www.iaik.tugraz.at/vt

Question Hours: Thursday 15:00 in the week before deadlines.

Discord: https://discord.gg/RaNW4KgGJf channel VT (activate with check mark)

Email: benedikt.maderbacher@iaik.tugraz.at
erwin.peterlin@student.tugraz.at
sebastian.puck@student.tugraz.at

Plan

DATE	TOPIC
13 Oct	Eraser & Locktree
20 Oct (VH)	Memory Debuggers
27 Oct	Symbolic Methods
03 Nov	Hoare Logic
10 Nov (BM)	Hoare Logic II
17 Nov, i11 (BM)	Deductive Program Verification
24 Nov, 1 Dec, 15 Dec	SLAM
~~22 Dec, 19 Dec, 5 Jan~~	— Christmas Holidays —
12 Jan	Java Pathfinder
19 Jan	Current Research Topics + Question Hour
26 Jan	EXAM

How to get a grade?

Lecture:

- Take the exam (main exam date: 26 Jan 2023)

Exercises:

- 4 assignments
- At least one submission → you'll get a grade
- Exercise Interviews: 24 Jan 2023

Exercises

Assignment	UE Handout	UE Question Hour	UE Deadline
A1 Eraser	13 Oct	20 Oct	27 Oct
A2 Hoare	3 Nov	10 Nov	17 Nov
A3 Dafny	17 Nov	24 Nov	9 Dec
A4 SLAM	16 Dec	17 Jan	20 Jan

Grading Scale (Exercise):

```
if (points(a1) >= 10 && points(a2) >= 10 &&
    points(a3) >= 10 && points(a4) >= 10)
{
    sum = points(a1) + points(a2) + points(a3) + points(a4)
    if (sum / 4.0 >= 87.5)
        return 1;
    if (sum / 4.0 >= 75)
        return 2;
    if (sum / 4.0 >= 62.5)
        return 3;
    if (sum / 4.0 >= 50)
       return 4;
}
return 5;
```

The Sorry State of Testing

```
Apple SSL/TSL v55741, Feb. 2014
SSLVerifySignedServerKeyExchange:
. . .
hashOut.data = hashes + SSL_MD5_DIGEST_LEN;
hashOut.length = SSL_SHA1_DIGEST_LEN;
if ((err = SSLFreeBuffer(&hashCtx)) != 0)
    goto fail;
if ((err = ReadyHash(&SSLHashSHA1, &hashCtx)) != 0)
    goto fail;
if ((err = SSLHashSHA1.update(&hashCtx, &clientRandom)) != 0)
    goto fail;
if ((err = SSLHashSHA1.update(&hashCtx, &serverRandom)) != 0)
    goto fail;
if ((err = SSLHashSHA1.update(&hashCtx, &signedParams)) != 0)
    goto fail;
    goto fail;  /* MISTAKE! THIS LINE SHOULD NOT BE HERE. err==0 */
if ((err = SSLHashSHA1.final(&hashCtx, &hashOut)) != 0)
    goto fail;

err = sslRawVerify(...);
. . .
```


http://nakedsecurity.sophos.com/2014/02/24/anatomy-of-a-goto-fail-apples-ssl-bug-explained-plus-an-unofficial-patch/

Microsoft EULA

Except for the Limited Warranty and to the maximum extent permitted by applicable law, [we] provide the Software and support services (if any) AS IS AND WITH ALL FAULTS, and hereby disclaim all other warranties and conditions, whether express, implied or statutory, including, but not limited to, any (if any) implied warranties, duties or conditions of merchantability, of fitness for a particular purpose, of reliability or availability, of accuracy or completeness of responses, of results, of workmanlike effort, of lack of viruses, and of lack of negligence, all with regard to the Software, and the provision of or failure to provide support or other services, information, software, and related content through the Software or otherwise arising out of the use of the Software.

this is an old EULA. Newer software comes with a 90-day limited warranty

Damage due to Bugs (US alone)

$20-$60 billion annually

Size of software industry: $120billion

Things Go Very Wrong

Things Go Very Wrong

Ariane 5 flight 501, 4 June 1996

Reuse of module written for Ariane 4, which is slower.
Acceleration values do not fit 16 bit integer.

1. Out-of-range value leads to unhandled exception in active and backup systems
2. Software transmits diagnostic data to main computer.
3. Main computer interprets diagnostic input as navigation data
4. Rockets starts tearing apart, triggers self destruct system

Failed system was not needed on Ariane 5.

Cost: $400m

More?

1993 – Intel Pentium floating point divide

2000 – National Cancer Institute, Panama City.

2003 – Northeast blackout

Faulty software may always be a part of the electric grid's DNA – Tom Kropp, manager, enterprise information security program, Electric Power Research Institute

Toronto, 14 August 2003 evening (Wikipedia)

Report: Software bug led to death in Uber's self-driving crash

Sensors detected Elaine Herzberg, but software reportedly decided to ignore her.

TIMOTHY B. LEE - 5/8/2018, 12:12 AM

What about Other Disciplines?

Si-o-se Pol bridge, 1600, Isfahan

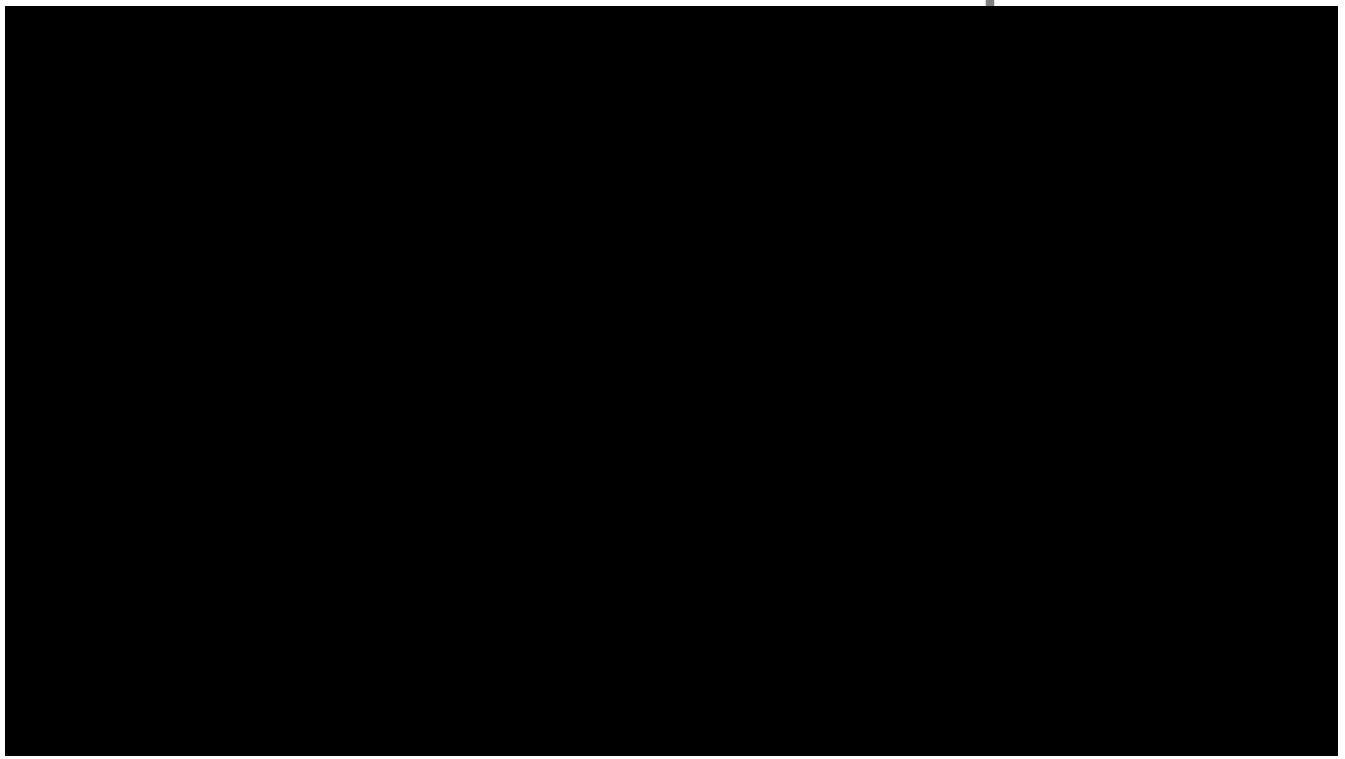
What about Other Disciplines?

"Engineering is the discipline, art, and profession that applies scientific theory to design, develop, and analyze technological solutions."

Civil engineering is an engineering discipline. Computer science is not.

Why not?

What about Other Disciplines?



Tacoma Narrows Bridge. Washington, 1940

Fragile suspension bridge (new type of design)

Aerodynamics!

What about Other Disciplines?

Tacoma Narrows Bridge. Washington, 1940

Fragile suspension bridge (new type of design). Aerodynamics!

What about Other Disciplines?

Erasmus Bridge

Rotterdam, Netherlands, 1997

Aerodynamical problem

Solved by adding extra wires

What about Other Disciplines?

Millenium Bridge

London, 2000. Cost: £18M.

`Suspension bridge´

Resonance problem

Added shock absorbers (Cost: £5M)

What about Other Disciplines?

Common theme in failures: **new design**

In computer science, every design is new!

Two contributions

1. **Mathematical rigor**: Verify, don't test!
2. **Correctness first**: Establish correctness while programming

Verification & Testing

Testing: Try out the software for many different scenarios

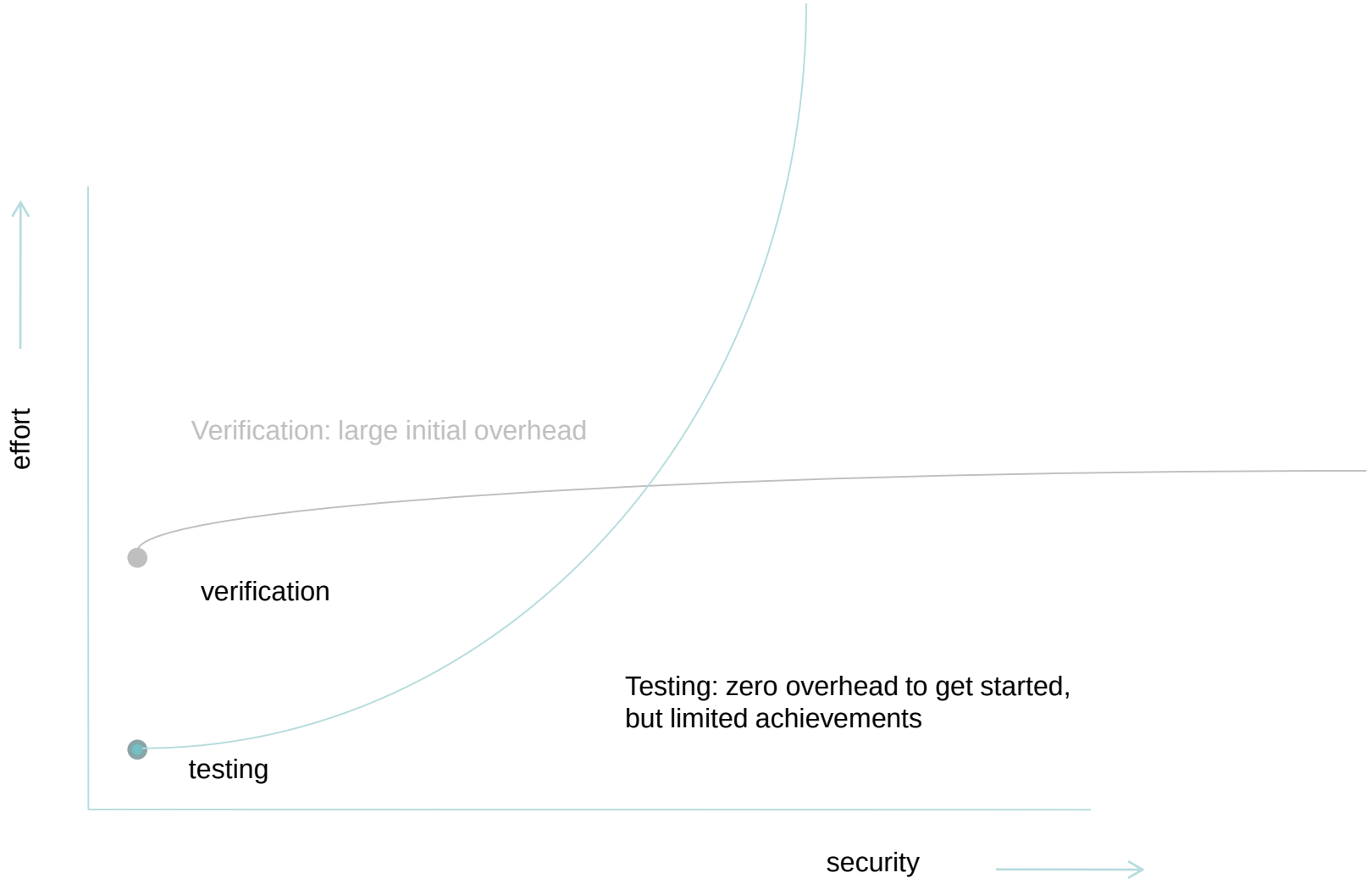
Verification: prove the correctness of software

Testing: some payoff for any size system

Verification: scaling is hard

MAIN CHALLENGE

Testing vs Verification



Hardware & Software

Hardware		Software
high	expectation of quality	low
high	cost of reimplementation	low
high	cost to update / replace	low
high	cost of failure	low?

model checking becoming standard

model checking in limited domains

Verification Example: Pentium

Floating Point Unit formally verified

Working on full formal verification

Why?

Testing uses ~6000 CPUs running 24/7

Total simulation cycles prior to tapeout < 1 minute of a 2 GHz system

No amount of dynamic validation is enough

- A single dyadic extended-precision (80-bit) FP instruction has O(10^50) possible combinations
- Exhaustive testing is impossible, even on real silicon

Age of universe: 10^26ns.

Bob Bentley, Intel, 2005

A problem has been detected and Windows has been shut down to prevent damage to your computer.

PFN_LIST_CORRUPT

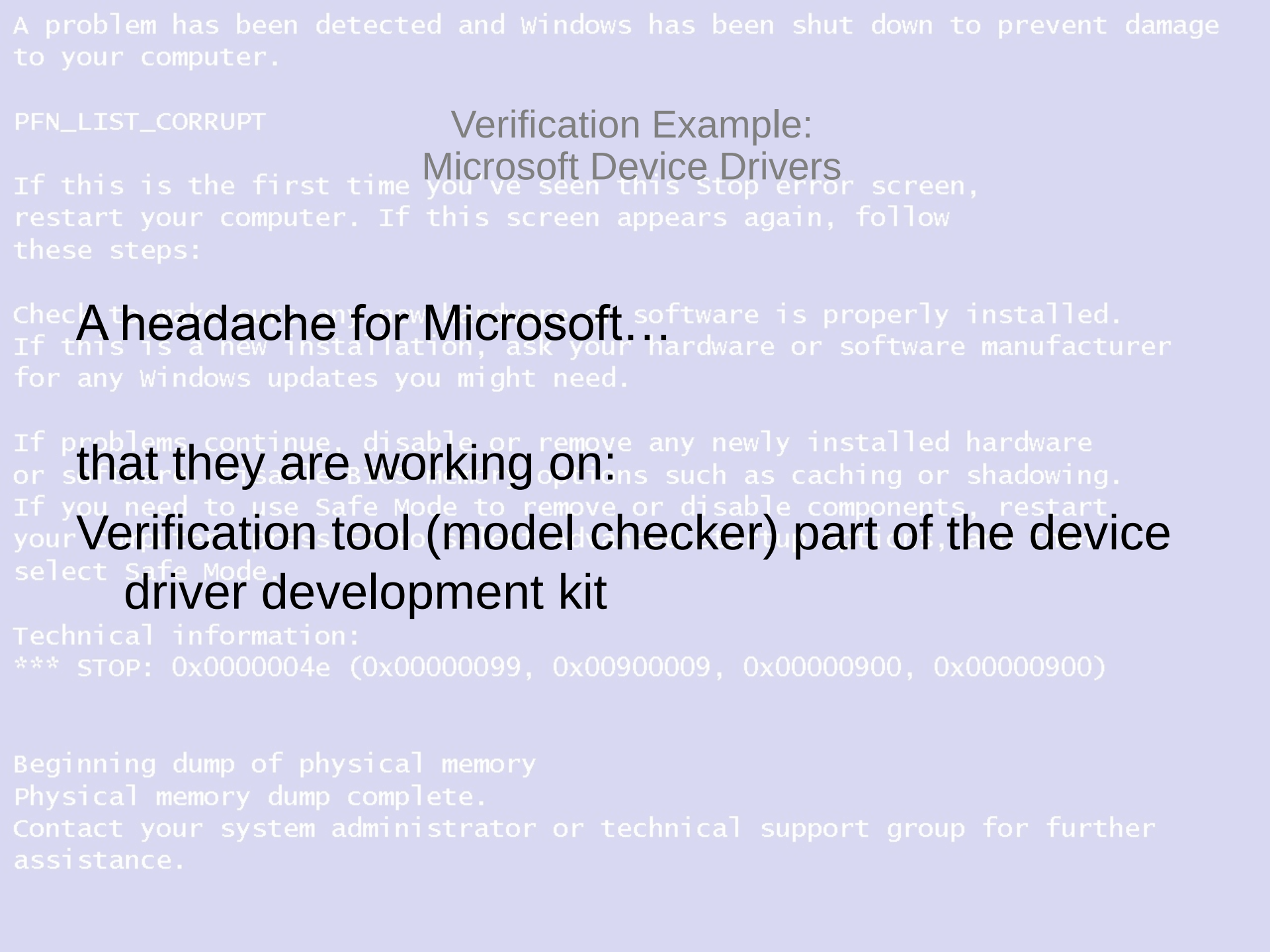
If this is the first time you've seen this Stop error screen, restart your computer. If this screen appears again, follow these steps:

Check to make sure any new hardware or software is properly installed. If this is a new installation, ask your hardware or software manufacturer for any Windows updates you might need.

If problems continue, disable or remove any newly installed hardware or software. Disable BIOS memory options such as caching or shadowing. If you need to use Safe Mode to remove or disable components, restart your computer, press F8 to select Advanced Startup Options, and then select Safe Mode.

Technical information:
*** STOP: 0x0000004e (0x00000099, 0x00900009, 0x00000900, 0x00000900)

Beginning dump of physical memory
Physical memory dump complete.
Contact your system administrator or technical support group for further assistance.



Verification Example: Microsoft Device Drivers

A headache for Microsoft…

that they are working on:

Verification tool (model checker) part of the device driver development kit

State of the Art

Verification is standard

- In VLSI design
- In MS Windows development
- At Facebook
- At Amazon
- ...

Plan

Dynamic Algorithms. *Get more from testing*

- Deadlocks: Eraser & Locktree
- Memory use: Valgrind & Purify
- Symbolic Execution

Static Algorithms Prove absence of bugs

- Symbolic Execution
- Java Path Finder
- Static Analysis
- Hoare Logic
- Abstraction and refinement: Microsoft's SLAM