Software Diversity for Security

Dr. Julio Hernandez-Castro

School of Computing
University of Kent, UK

Software Diversity for Security

The Irish potato famine

The sad story about *Gros Michel* and *Cavendish* bananas

Why diversity?

- Low diversity
  - *Monoculture considered harmful* - Geer
  - Millions of exact clones
  - Exposed to the exact vulnerabilities
  - Botnets, criminal profit and risk are huge

Why diversity?

High diversity

- Millions of similar products
- Not exposed to same exact vulnerabilities
- Botnets, criminal profit and risk are lower

OK, I'm convinced!

Let's impose high diversity in software, then!

- Not trivial
- Developers are reluctant
- More time consuming
- Costlier to maintain

Ohhh...

What has been done

In the origins, it was safety, not security

N-Version programming

Interesting, but costly, not automatic

And working worse than initially though

What has been done

- Diversity by combination
  - Combining off-the-shelf software
    - Get different versions of your OS, Web server, browser, patches, etc. and hope for the best
    - Many advantages
    - Some limitations

What has been done

Diversity by combination

Combining off-the-shelf software

It's being proved to work

Also, it's proved to have serious limitations

But it is clearly one step forward

What has been done

Diversity by inner change

Address Space Layout Randomization

> *“**randomly** arranging the positions of key data areas, usually including the base of the executable and position of libraries, heap, and stack, in a process's address space”* - Wikipedia

Now mainstream: OpenBSD, Linux, Windows, OS X (post 2007), Android, iOS, etc.

Tested and Trusted technique, with limitations but OK

What has been done

- Diversity by inner change
  - Obfuscation
    - Not really properly tested
    - Not sure about results

What has been done

Diversity by inner change

- Of course, any of these proposals has to be automatic
- We saw the problems of N-versioning
- How to do it?

What should be done

If Diversity is important, we need a way of measuring it

- Inspired by Ecology literature, we are doing it using a way that is new in Computer Security
- It's called Shannon-Weiner index
- Its units are equivalent species
- May link in the future diversity & extiction probability

What should be done

Diversity by inner change

My proposal:

Random replacement by functionally equivalent code (RaReFun!)

- Randomly replace equivalent instruction sets (for-while-repeat)
- Randomly change data estructures (matrix-array-list-set)
- Randomly change variable types (float-double-long int)
- But guaranteeing functionality

Inspired by the watermarking tool *Hydan*

Hydan

Hydan used as a watermarking tool

Original code			*Encoding 00*		
83 e8 30	sub	%eax, $0x30	83 c0 d0	add	%eax, $-0x30
83 f8 36	cmp	%eax, $0x36	83 f8 36	cmp	%eax, $0x36
77 e5	ja	$-27	77 e5	ja	$-27
83 c0 08	add	%eax, $0x8	83 c0 08	add	%eax, $0x8
89 04 24	mov	%eax, [%esp]	89 04 24	mov	%eax, [%esp]
Encoding 01			*Encoding 11*		
83 c0 d0	add	%eax, $-0x30	83 e8 30	sub	%eax, $0x30
83 f8 36	cmp	%eax, $0x36	83 f8 36	cmp	%eax, $0x36
77 e5	ja	$-27	77 e5	ja	$-27
83 e8 f8	sub	%eax, $-0x8	83 e8 f8	sub	%eax, $-0x8
89 04 24	mov	%eax, [%esp]	89 04 24	mov	%eax, [%esp]

Table 1. Encoding the values `00`, `01`, and `11` using equivalent instructions (highlighted).

What's better?

- What's better, introducing diversity by combination or by inner change?
  - This can be put alternatively
    - Is is better to change at the block level, or inside the blocks themselves?
    - Blocks = OS, applications, servers, etc.
  - No evidence so far for prefering one to another
    - But in another biological analogy influenza uses both types of changes to be so successful - Rossman

The End

Thanks for your time

Any questions?

Postscript

Talk to your fellow biologist

They know much more about security than you though, just in a different context

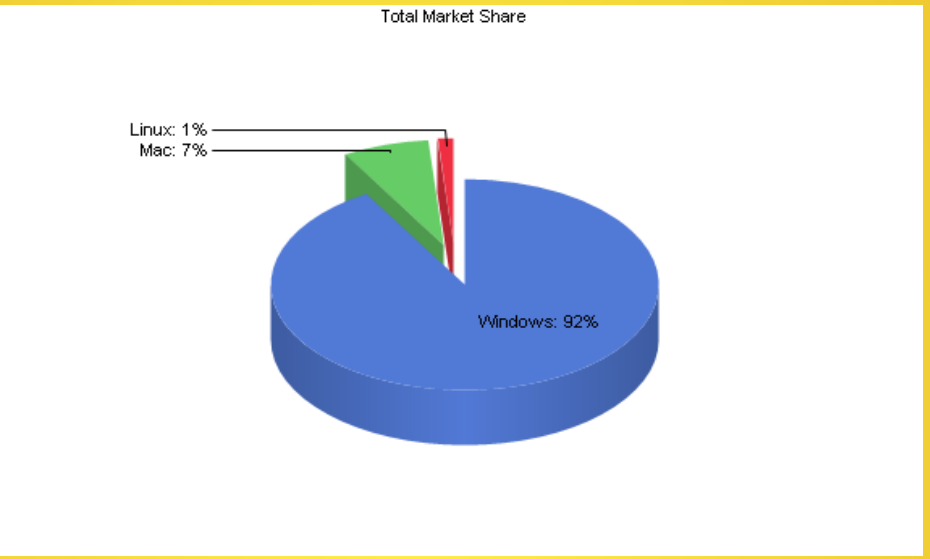
Right questions get right answers

Postscript

- Some clear trends in current malware:
  - Most doesn't work
    - Prototypes, proof of concepts, or simply wrong
  - Vast code reuse
    - Entry level increasignly complex
  - Large majority from a previous successful family
    - Process similar to crossover or replication
  - Will we have specialists in the future?
    - In Biology we have Influenza, HIV, Malaria specialists
    - Likely that this will happen in Computer Security
  - The search for an ubermalware
    - Paralelisms with creation of bioterror weapon

Poor diversity: Examples

1.386 equivalent species



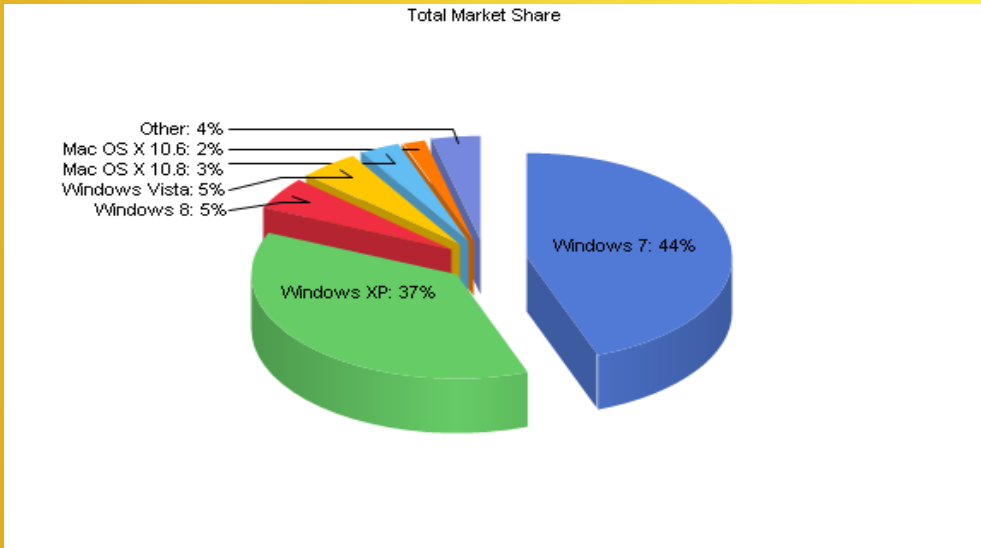
Desktop Operating System Market Share **June, 2013** **Operating System**	**Total Market Share**
Windows	91.51%
Mac	7.20%
Linux	1.28%

Poor diversity: Examples

3.971

equivalent

species



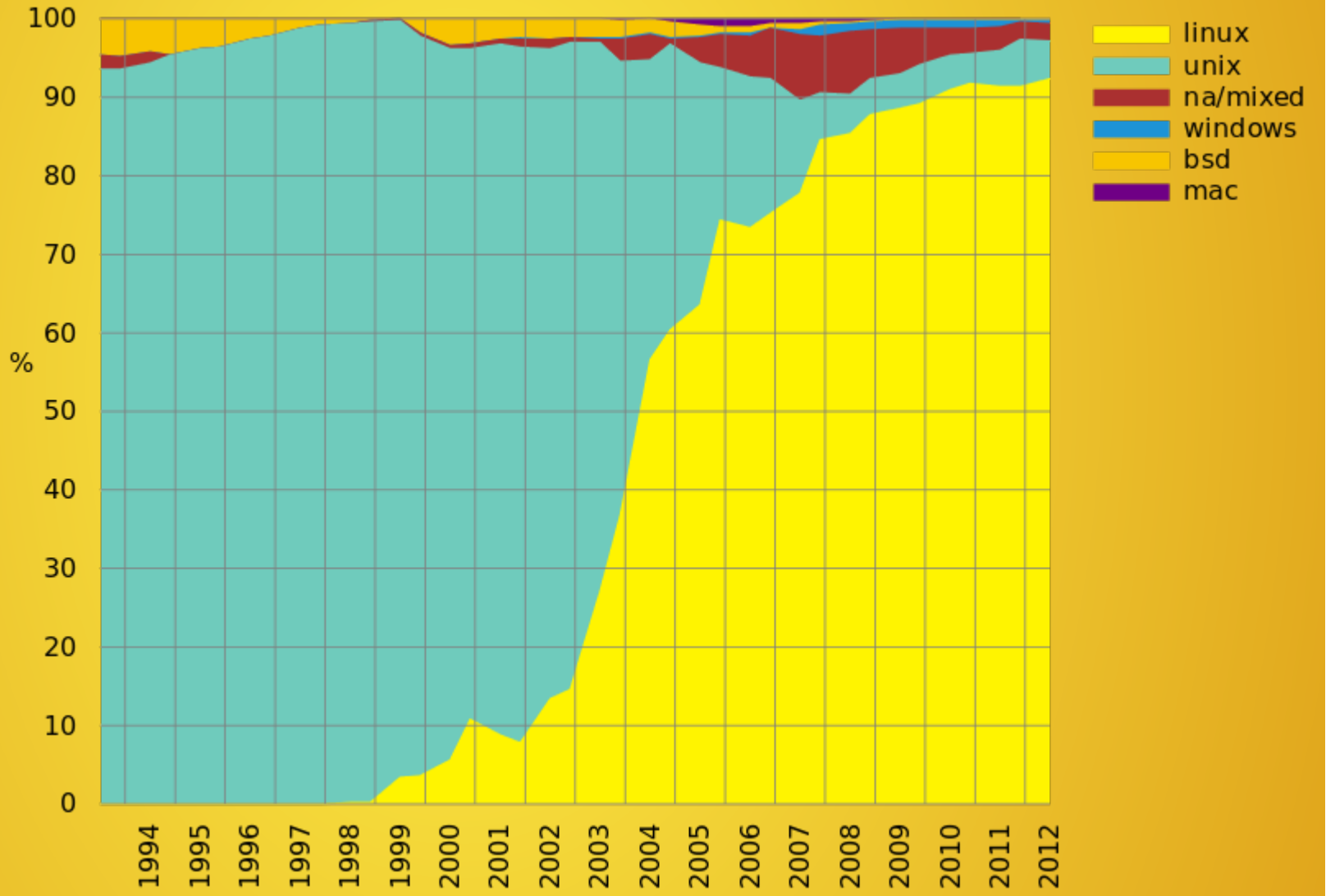
Operating System	Total Market Share
Windows 7	44.37%
Windows XP	37.17%
Windows 8	5.10%
Windows Vista	4.62%
Mac OS X 10.8	3.14%
Mac OS X 10.6	1.76%
Mac OS X 10.7	1.73%
Linux	1.28%
Mac OS X 10.5	0.43%
Windows NT	0.19%
Mac OS X 10.4	0.10%
Windows 2000	0.04%
Mac OS X 10.9	0.02%
Mac OS X (no version reported)	0.02%
Win64	0.01%
Windows 98	0.00%
Mac OS X Mach-O	0.00%
Windows ME	0.00%

Poor diversity: Examples

TOP 500

Supercomputers

1.269 equivalent species!



Poor diversity: Open questions

Is it true that the less diverse ecosystems are worst affected by malware?

- Malware developers are intelligent, after all
- Motivations & rewards clearly depend on it

This question needs further research

- My guess is that it is a combination of diversity and size
- But if true, a direct implication of this is that we'll soon see malware directly targeted to supercomputers in the Top500

Low diversity could be a security weakness in itself

- even if there are no outstanding problems after an audit
- Nmap plugin to diagnose the diversity of a Company (IP range)