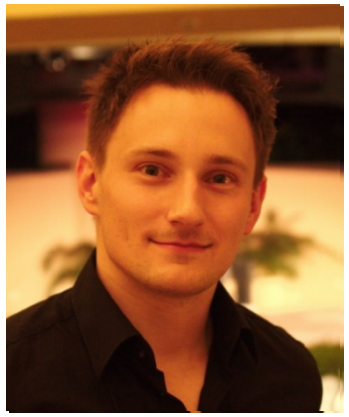
Steven She

UWaterloo

Rafael Lotufo

Auvik Networks Inc.



Thorsten Berger

Chalmers | University of Gothenburg

Andrzej Wasowski

ITU Copenhagen

Krzysztof Czarnecki

UWaterloo

the variability model of the Linux kernel

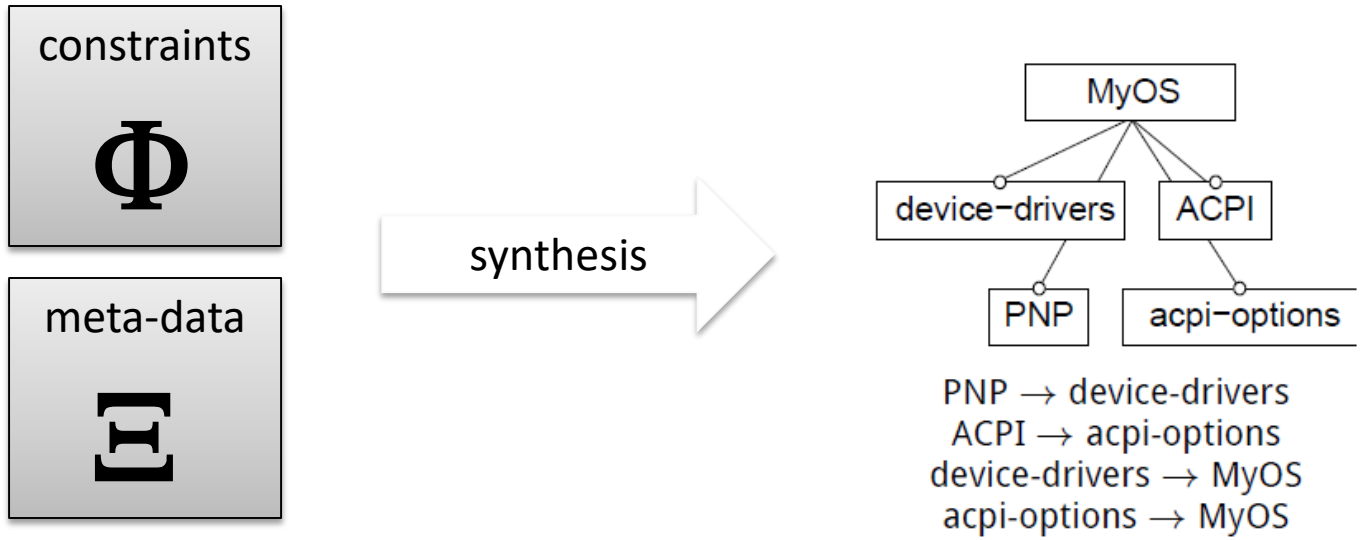
presented at **VaMoS 2010**

received **VaMoS 2020** most influential paper award

VaMoS 2020: 14th Working Conference on Variability Modeling of Software-Intensive Systems

start of my PhD studies

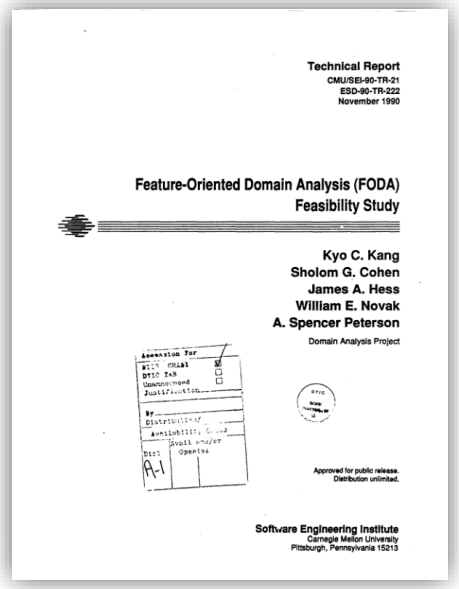
goal: feature model synthesis



Czarnecki, Wasowski. *Feature diagrams and logics: There and back again.* SPLC. *2007*

Benavides, Trinidad, Ruiz-Cortés: *Automated reasoning on feature models*. CAiSE 2005.

Batory. *Feature models, grammars, and propositional formulas*. SPLC. 2005



Technical Report
CMU/SEI-90-TR-21
ESD-90-TR-222
November 1990

Feature-Oriented Domain Analysis (FODA)
Feasibility Study

Kyo C. Kang
Sholom G. Cohen
James A. Hess
William E. Novak
A. Spencer Peterson

Domain Analysis Project

Approved for public release.
Distribution unlimited.

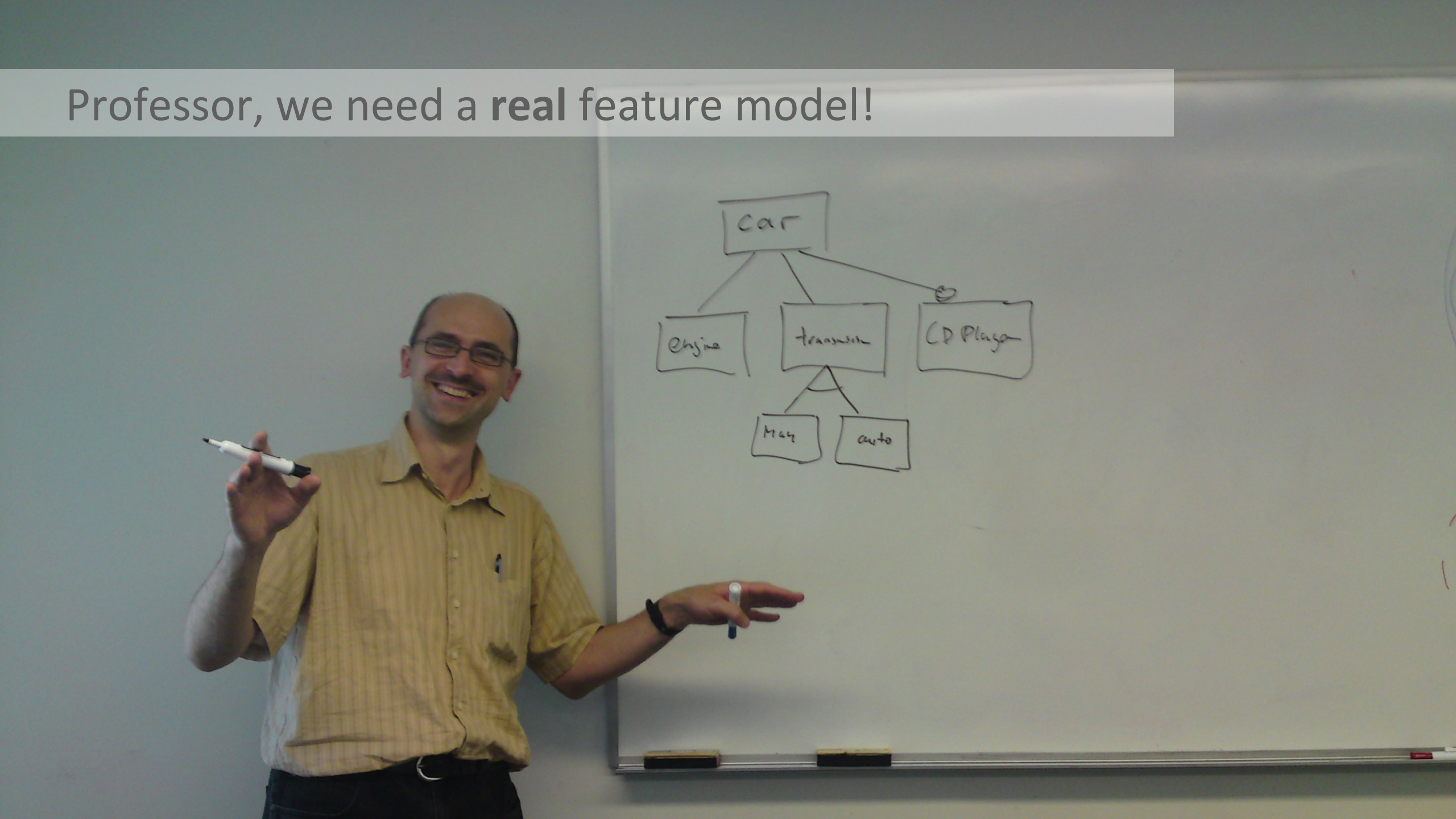
Software Engineering Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

1616 citations (January 11, 2010)

4856 citations (now!)

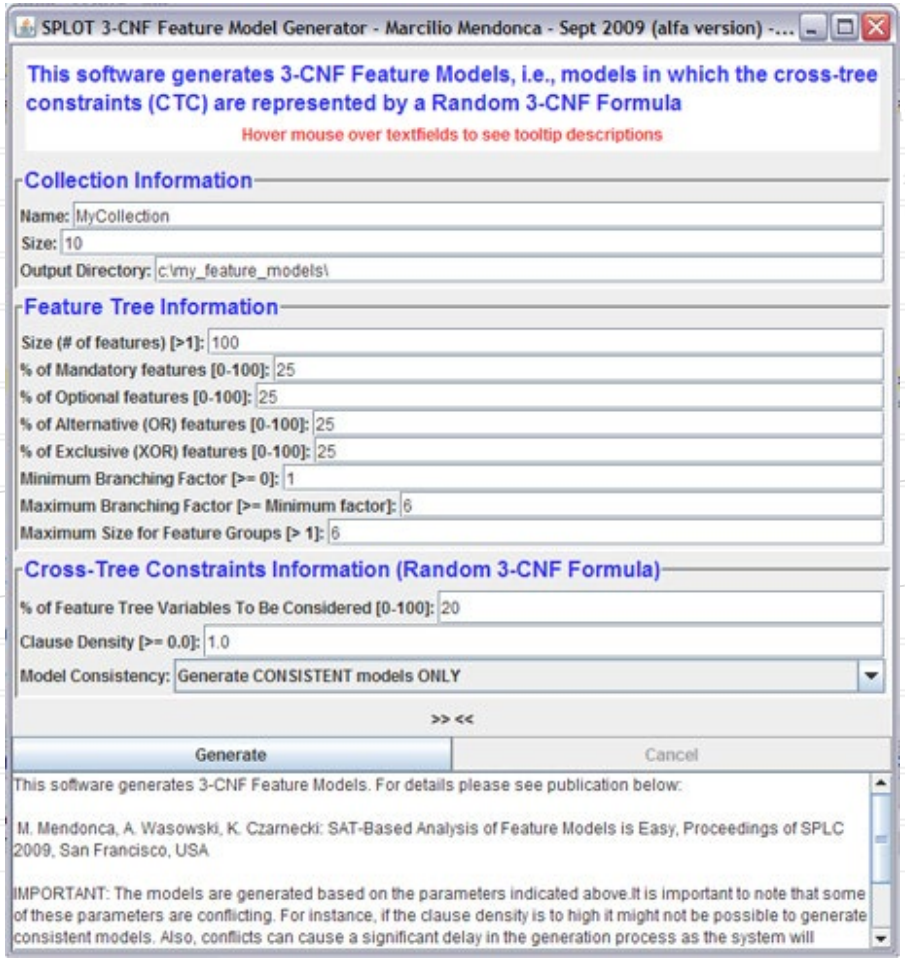
thousands of publications build upon feature modeling:
model configuration, analysis, evolution, verification, reverse-engineering

Professor, we need a **real** feature model!



problem: toy models

propositional models

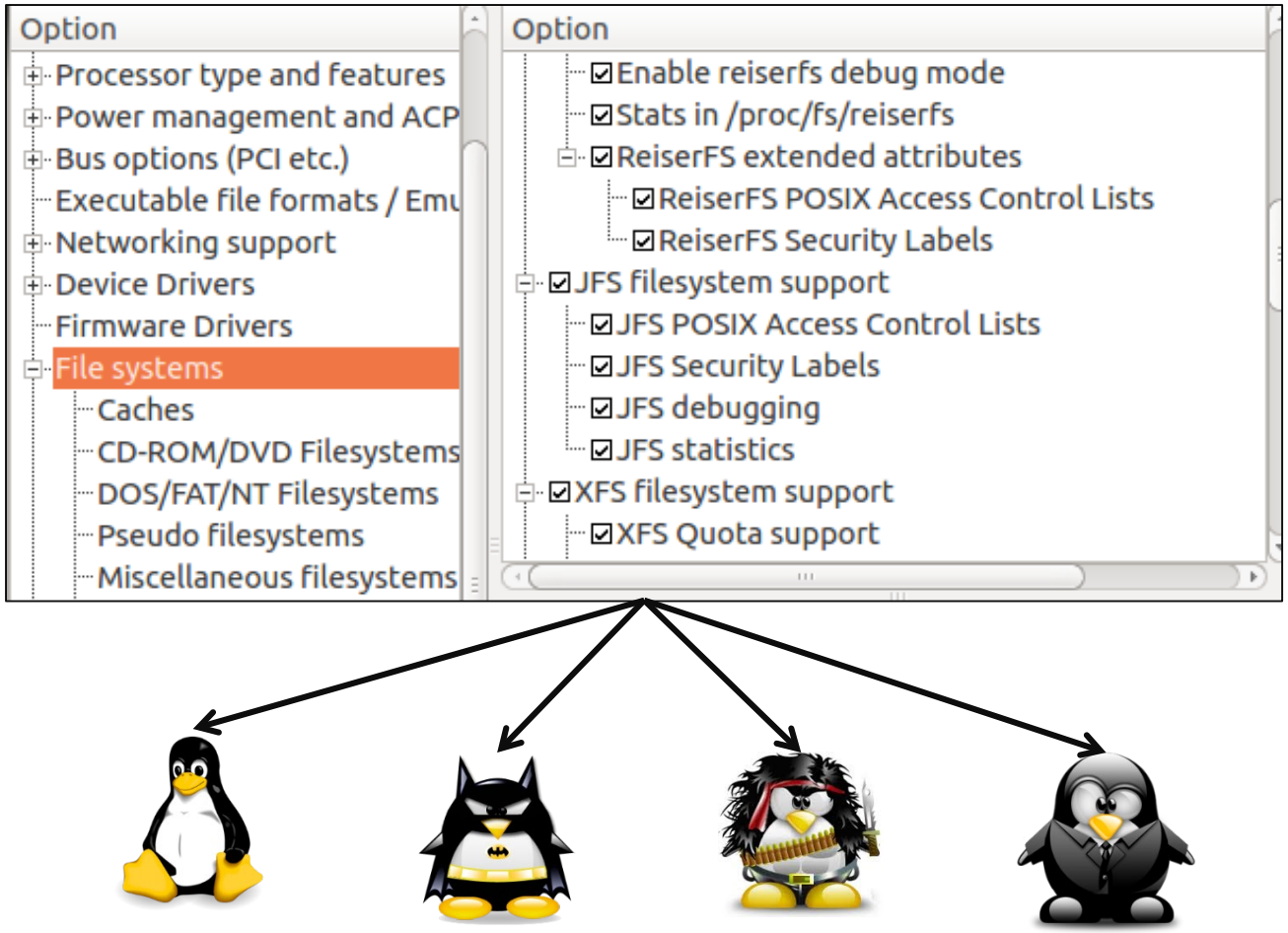


SPLOT model repository

www.splot-research.org

highly configurable systems software

Linux kernel
now ~15,000 options
(features)



not the first looking at Linux and benchmarking

Sincero, Schirmeier, Schröder-Preikschat, Spinczyk. **Is the Linux Kernel a Software Product Line?**. *OSSPL*. 2007

Sincero, Schröder-Preikschat. **The Linux Kernel Configurator as a Feature Modeling Tool**. *ASPL*. 2008

Tartler, Sincero, Schröder-Preikschat, Lohmann. **Dead or Alive: Finding zombie features in the Linux kernel**. *FOSD*. 2009.

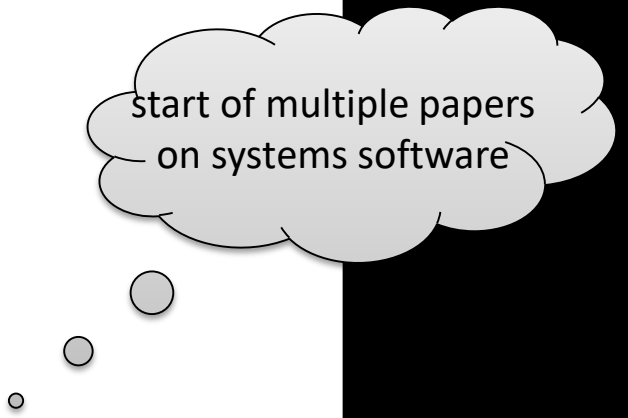
Segura, Cortés. **Benchmarking on the Automated Analyses of Feature Models: A Preliminary Roadmap**. *VaMoS*. 2009

The Variability Model of the Linux Kernel

Steven She, Rafael Lotufo, Thorsten Berger,
Krzysztof Czarnecki, Andrzej Wąsowski

University of Waterloo
University of Leipzig
IT University of Copenhagen

January 26, 2010



Kconfig declaration

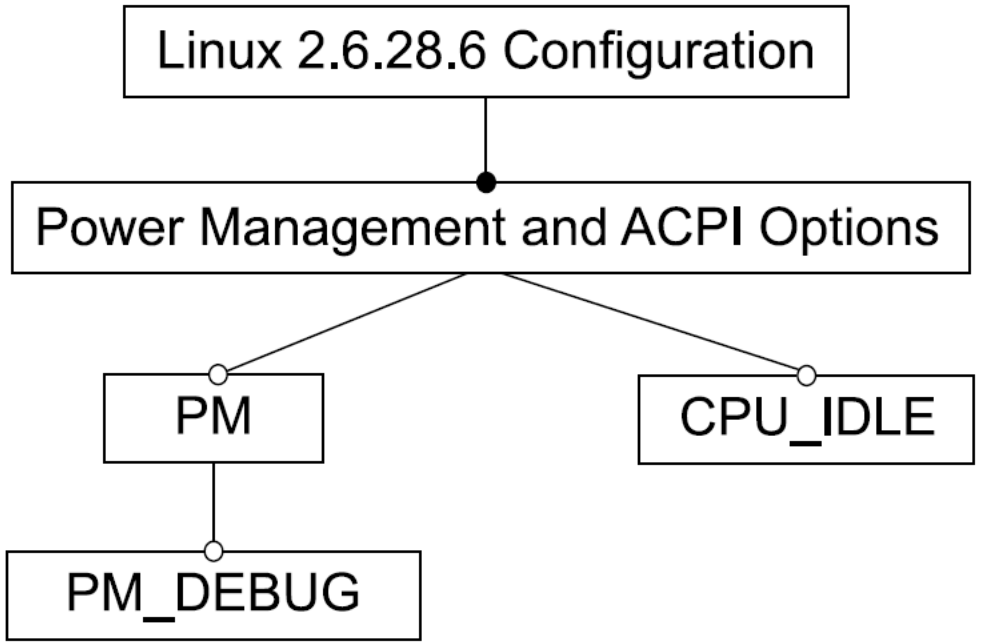
```
menu "Power management and ACPI options"
  depends on !X86_VOYAGER

  config PM
    bool "Power Management support"
    depends on !IA64_HP_SIM
    ---help---
        "Power Management" means that …

  config PM_DEBUG
    bool "Power Management Debug Support"
    depends on PM

  config CPU_IDLE
    bool "CPU idle PM support"
    default ACPI
endmenu
```

Kconfig feature model



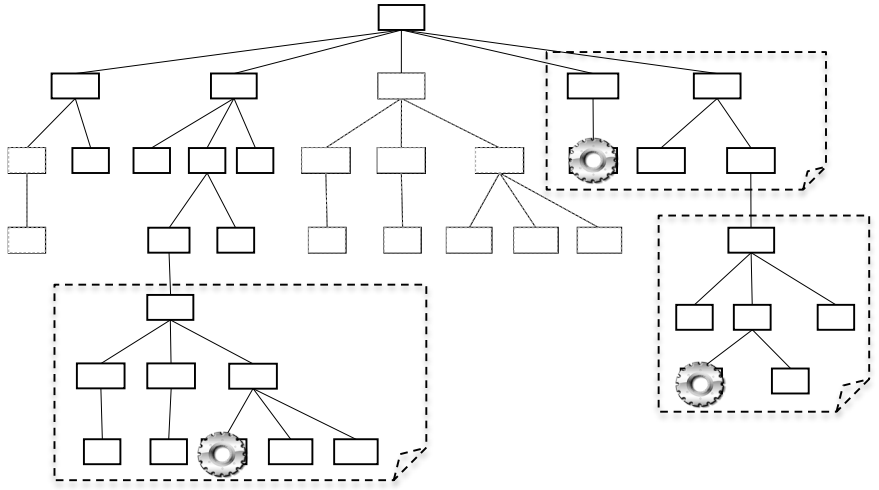
$$\text{PM\_MENU} \rightarrow \neg\text{X86\_VOYAGER}$$
$$\text{PM} \rightarrow \neg\text{IA64\_HP\_SIM}$$
$$\text{ACPI} \leftrightarrow \text{CPU\_IDLE}$$

Berger, She, Lotufo, Wasowski, Czarnecki, *A Study of Variability Models and Languages in the Systems Software Domain*, *IEEE Transactions on Software Engineering*. 2013.

Kconfig goes well beyond feature modeling!

scalability concepts

- visibility
- modularization
- derived defaults / derived features
- hierarchy manipulation



expressive constraints

- three-state logics (follows Kleene‘s rules) for binding modes
- comparison, arithmetic, and String operators

domain-specific vocabulary

Linux KConfig → Feature Model

Analyzed four aspects of the Linux 2.6.28.6 Kconfig model in terms of feature modeling concepts:

- characterized features,
- model hierarchy,
- constraints,
- and natural language properties.

Comparing with published models

Compared Linux statistics with 32 published models[1].

- 19 models - software product lines
- 8 models - other product lines (e.g. hardware, business)
- 5 models - domain models (e.g. eCommerce systems)

Only 5 models describe real, existing software systems

[1] http://www.splot-research.org

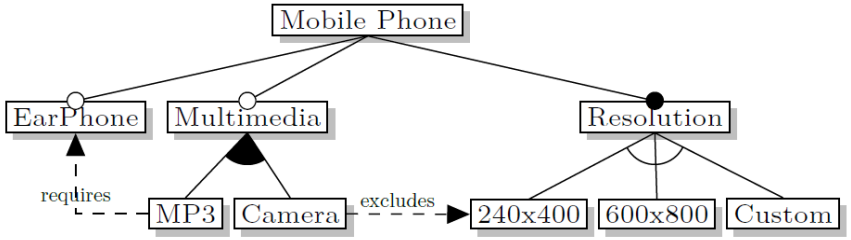
Linux feature statistics

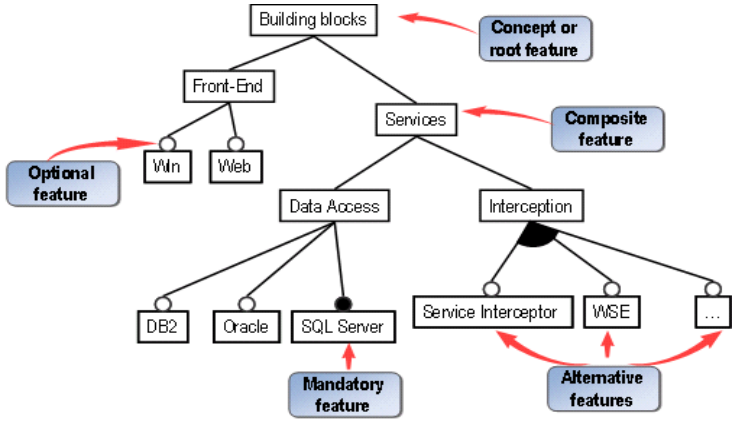
Kconfig Concept	Features	Mand.	Grouped	XOR + OR
Config	5323	0	146	0
Non / User-Sel.	547 + 4744			
Boolean	2005	0	136	0
Tristate	3130	0	10	0
Int	132	132	0	0
Hex	29	29	0	0
String	27	27	0	0
Menu	71	38	0	0
Choice	32	31	0	30 + 2
Total	**5426**	257	146	30 + 2

Published models vs. Linux

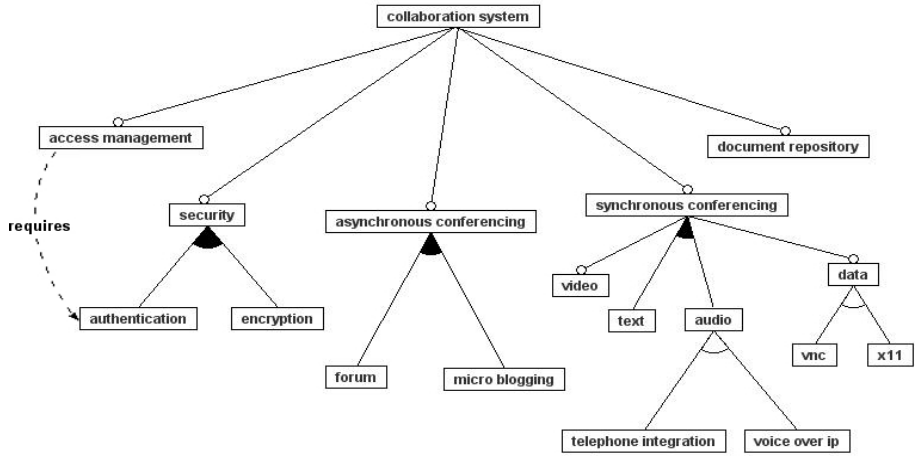
Concept	Published Models (%)			Linux (%)
	median	min	max	
mandatory features	25	0	66	4.74
grouped features	44	0	75	2.69
groups	16	0	35	0.59
XOR	9	0	30	0.55
OR	6	0	16	0.04

published models





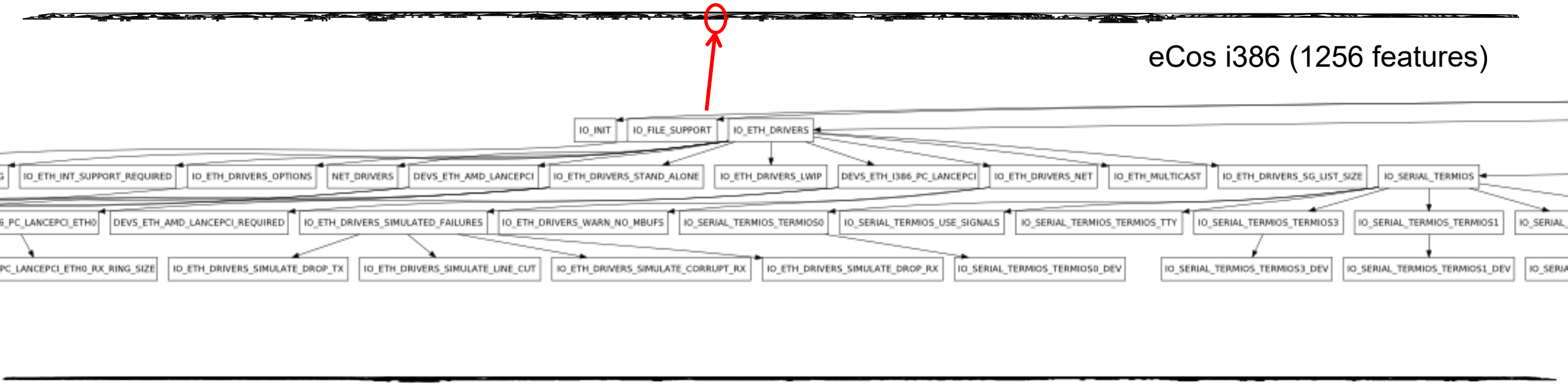
msdn.microsoft.com/en-us/library/aa925157.aspx



code.google.com/p/dslvariantmanagement/wiki/DemoShowCase

Berger, She, Lotufo, Wasowski, Czarnecki, *A Study of Variability Models and Languages in the Systems Software Domain*, *IEEE Transactions on Software Engineering.* 2013.

systems software models



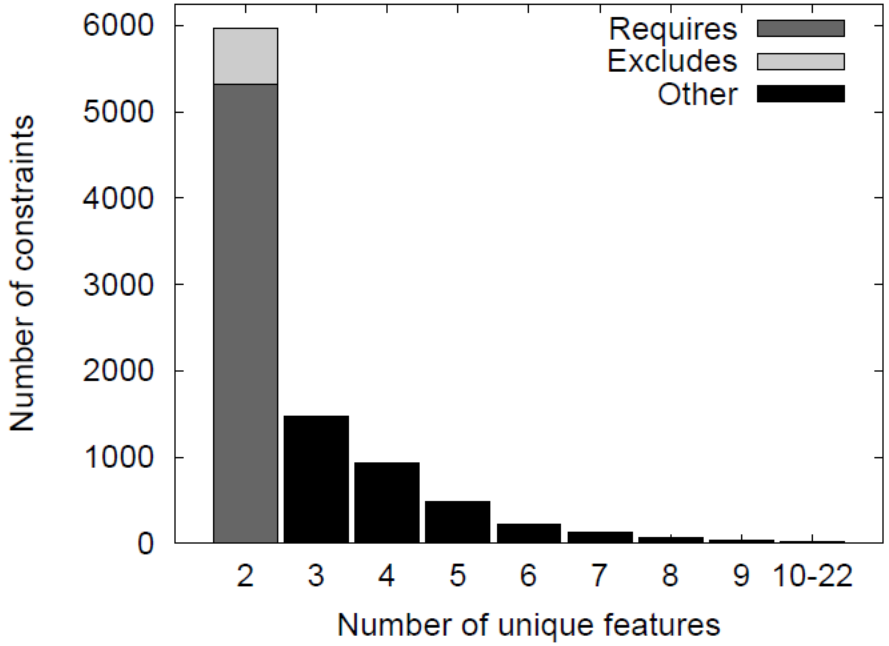
Code-granularity of features

Two heuristics for automatic feature selection in the Linux configurator: *allyes*, *allno*.

Metric	**allyes**	**allno**	Δ	θ
Features	3,448	61	3387	1
Files	10,326	973	9,353	2.76
SLOC	4,266,171	210,302	4,055,869	1,197.48

$$\Delta_i = \textbf{allyes}_i - \textbf{allno}_i; \theta_i = \Delta_i / \Delta_1$$

Constraint statistics



- 9291 constraints
- 82% features referenced
- 89% requires constraints
- some v. large constraints

Significant number of cross-tree constraints!

systems software models

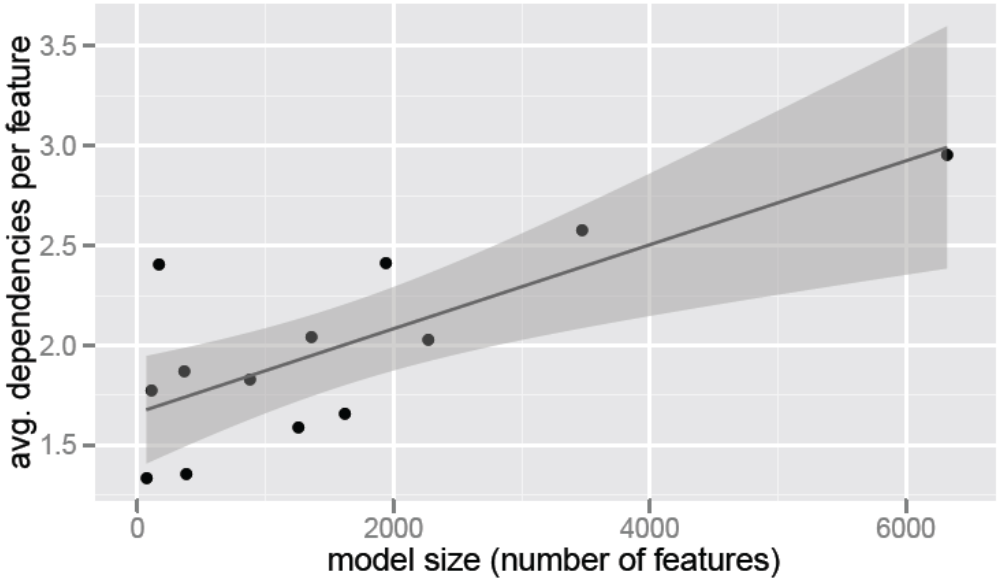
Berger, She, Lotufo, Wasowski, Czarnecki, *A Study of Variability Models and Languages in the Systems Software Domain*. IEEE Transactions on Software Engineering. 2013.

Lotufo, She, Berger, Czarnecki, Wasowski, *Evolution of the Linux Kernel Variability Model*. SPLC. 2010.

dependency graph is still sparse

linear dependency between size of model and dependencies (confirmed in temporal studies)

good news!



indicates that feature-based architectures scale well

Identifiers, prompts and descriptions

```
config PM
  bool "Power Management support"
  depends on !IA64_HP_SIM
  ---help---
```

"Power Management" means that parts of your computer are shut off or put into a power conserving "sleep" mode if they are not being used. There are two competing standards for doing this: APM and ACPI. If you want to use either one, say Y here and then also to the requisite support below.

Power Management is most important for battery powered laptop computers; if you have a laptop, check out the Linux Laptop home page on the WWW at <http://www.linux-on-laptops.com/> or Tuxmobil - Linux on Mobile Computers at <http://www.tuxmobil.org/> and the Battery Powered Linux mini-HOWTO, available from <http://www.tldp.org/docs.htmlhowto>.

Natural language properties

artifact	no. of characters			no. of words		
	median	min	max	median	min	max
identifiers	13	2	58	2	1	9
prompts	27	2	82	4	1	13
description	-	-	-	29	2	392

Size of textual artifacts

text source	most frequent domain terms					
identifier	*usb*	*snd*	*md*	*serial*	*fb*	*debug*
prompt	*usb*	*ethernet*	*pci*	*intel*	*scsi*	*pcmcia*
description	*usb*	*linux*	*scsi*	*ethernet*	*pci*	*howto*

Top domain terms

Conclusions

bitbucket.org/tberger/
variability-models

- Low number of mandatory features and groups.
- Each feature crosses roughly 2.8 source files and 1200 SLOC.
- Average leaf depth of 4, many single childs and long tail.
- Significant number and size of cross-tree constraints.

[1] Available for download at `http://fm.gsdlab.org`

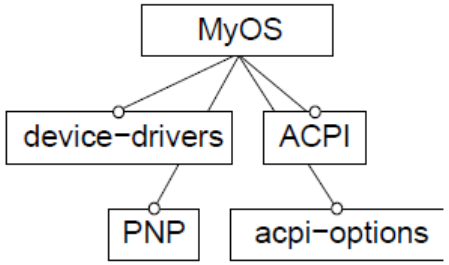
impact

together with follow-up works on systems software: >650 citations

used to evaluate new techniques for:

- quality assurance
  - Chen, Nair, Krishna, Menzies. *"Sampling" as a Baseline Optimizer for Search-Based Software Engineering*. IEEE Transactions on Software Engineering. 2019.
  - A. von Rhein. *Analysis Strategies for Configurable Systems.* Ph.D. dissertation, University of Passau, 2016.
  - Johansen, Haugen, Fleurey. *An Algorithm for Generating t-wise Covering Arrays from Large Feature Models*. SPLC. 2012.
  - ...
- evolution of feature models
  - Knüppel, Thüm, Mennicke, Meinicke, Schaefer. *Is there a Mismatch Between Real-World Feature Models and Product-Line Research?*. FSE. 2017.
  - Arcaini, Gargantini, Vavassori. *Automated Repairing of Variability Models*. SPLC, 2017.
  - ...
- automated software configuration
  - Xiong, Zhang, Hubaux, She, Wang, Czarnecki. *Range Fixes: Interactive Error Resolution for Software Configuration*. IEEE Transactions on Software Engineering. 2014.
  - Krieter, Thüm, Schulze, Schröter, Saake. *Propagating Configuration Decisions with Modal Implication Graphs*. ICSE. 2018.
  - ...

Kconfig semantics are highly intricate



PNP → device-drivers
ACPI → acpi-options
device-drivers → MyOS
acpi-options → MyOS

constraints

Φ

LVAT

Undertaker

Kconfigreader

kconfig-sat

kclause

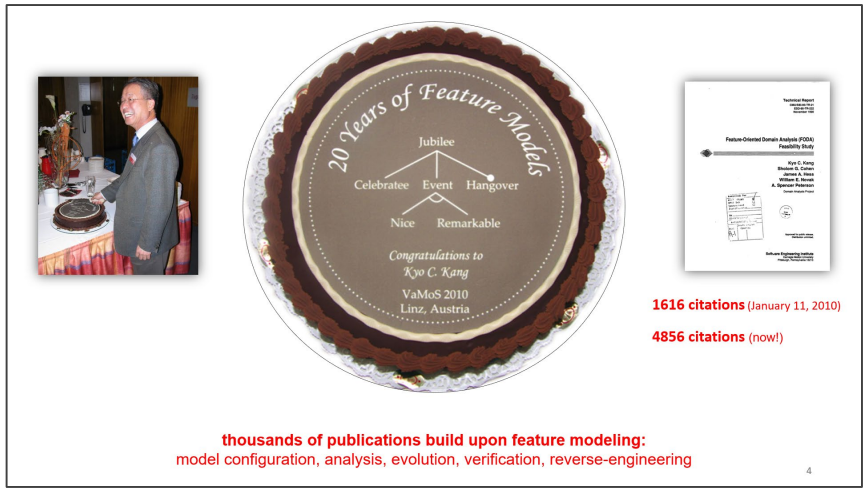
El-Sharkawy, Krafczyk, Schmid,. Analysing the Kconfig semantics and its analysis tools. GPCE. 2015

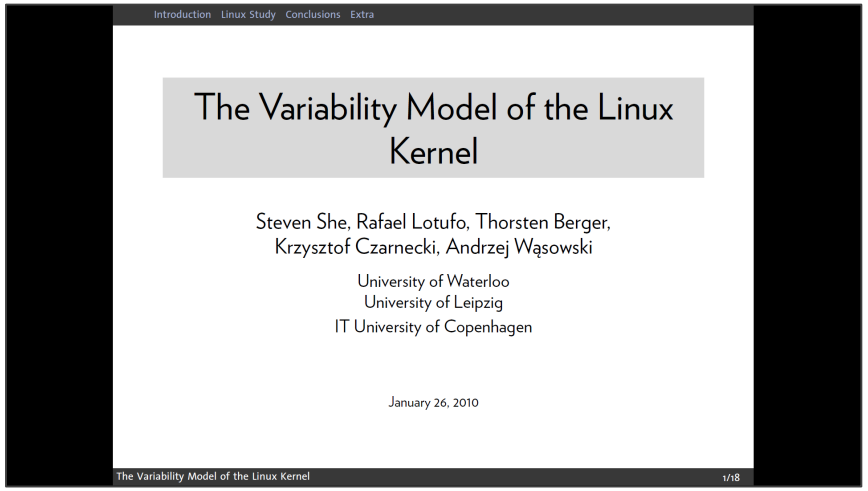
conclusion

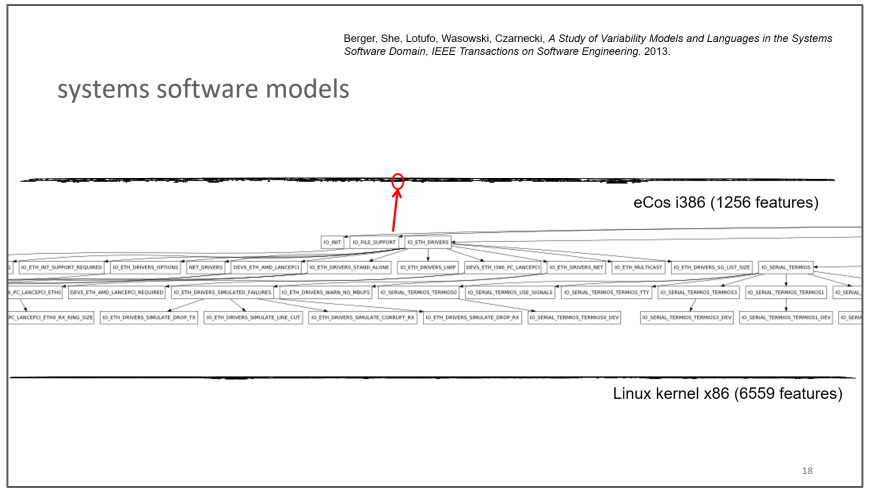
Kconfig was and is still a mess (but a very useful mess)

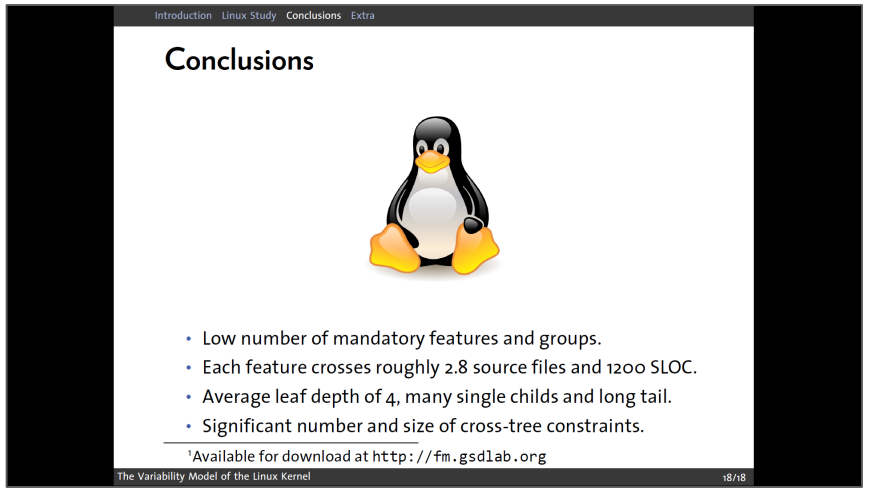
impact and relevance

- we did not really foresee it!
- moving from the state shown by SPLOT to the community working with real and complex systems
- methodological impact on feature modeling researchers
- moved our careers strongly towards work with real cases
- helped the community to publish in mainstream SE venues (among other factors)









the variability model of the Linux kernel

Steven She, Rafael Lotufo, Thorsten Berger, Andrzej Wasowski, Krzysztof Czarnecki