Reverse Engineering Product Lines



KV Product Line Engineering (343.354)

Dr. Roberto Lopez-Herrejon

Dr. Rick Rabiser







Lesson Overview

- Motivation for Reverse Engineering Software Product Lines
 - Basic ideas and examples
- Four main challenges
 - Reverse Engineering feature models
 - Traceability
 - Safe composition and feature oriented refactoring
 - Maintenance and Evolution



And Software Product Lines?

- Software Product Line
 - A set of similar software systems distinguished by the set of features they implement.
- Feature
 - Increment in program functionality

Fact: software products are marketed by features





MagicDraw® 17.0.3 FR featu	ires
----------------------------	------

			Edition		
UML support	Personal	Standard	Professional	Architect	Enterprise
Support for UML 2.4.1 metamodel and notation.	+	+	+	+	+
Support for UML 2 metamodel and notation.	+	+	+	+	+
Import of UML 1.4 metamodel.	+	+	+	+	+
Class diagram - includes Package and Objects diagrams.	+	+	+	+	+
Composite structure diagram.	+	+	+	+	+
Use Case diagram.	+	+	+	+	+

How are they built?

Business Process Modeling Notation support 1.x***	Personal	Standard	Professional	Architect	Enterprise
Support for Business Process Modeling Notation 1.x (BPMN).	+	+	+	+	
Business Process Modeling Notation (BPMN) export to BPEL 1.1 (BEA flavor).			+	+	
Business Process Modeling Notation support 2.0	Personal	Standard	Professional	Architect	Enterprise
Support for Business Process Modeling Notation 2.0 (BPMN).		+*	+*	+**	+**
UML extensions (profiles and diagrams)	Personal	Standard	Professional	Architect	Enterprise
Generic numbering mechanism in DSL models.		+	+	+	+
WSDL profile and diagram.				+	+
XML schema profile and diagram.				+	+
CORBA IDL profile and diagram.				+	+
Database structure profile and diagram: Generic DDL and Oracle DDL diagram				+	+
Web Application Extensions (WAE) profile and diagram.		+	+	+	+
Content diagram.		+	+	+	+



A tale of success...

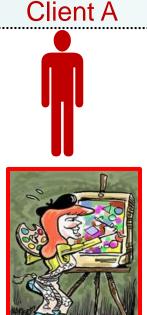
Requirements **Specification**



Design



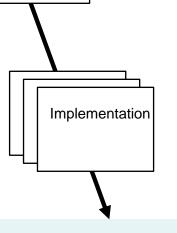
Implementation











Design



2016-06-15

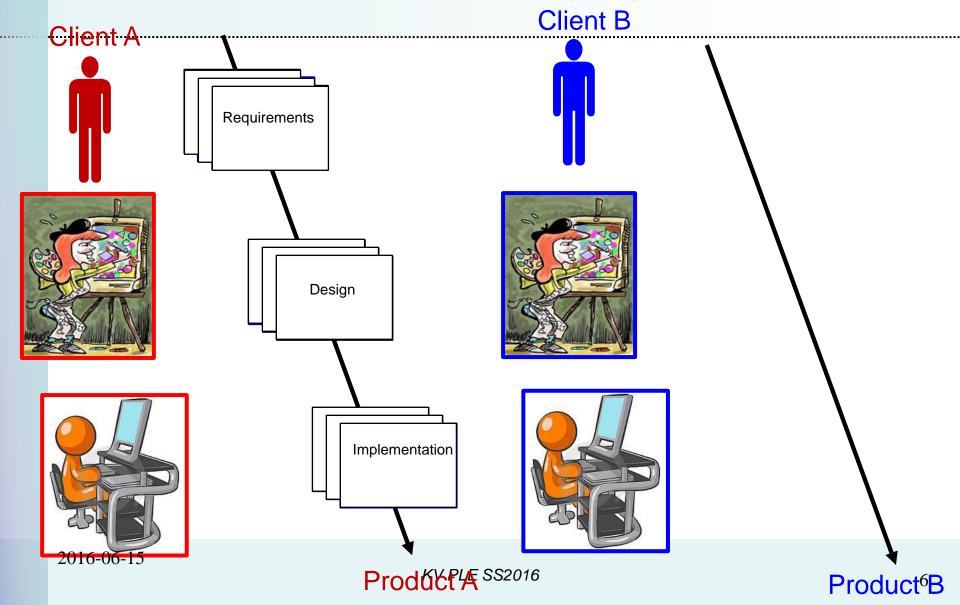
KV PLE SS2016

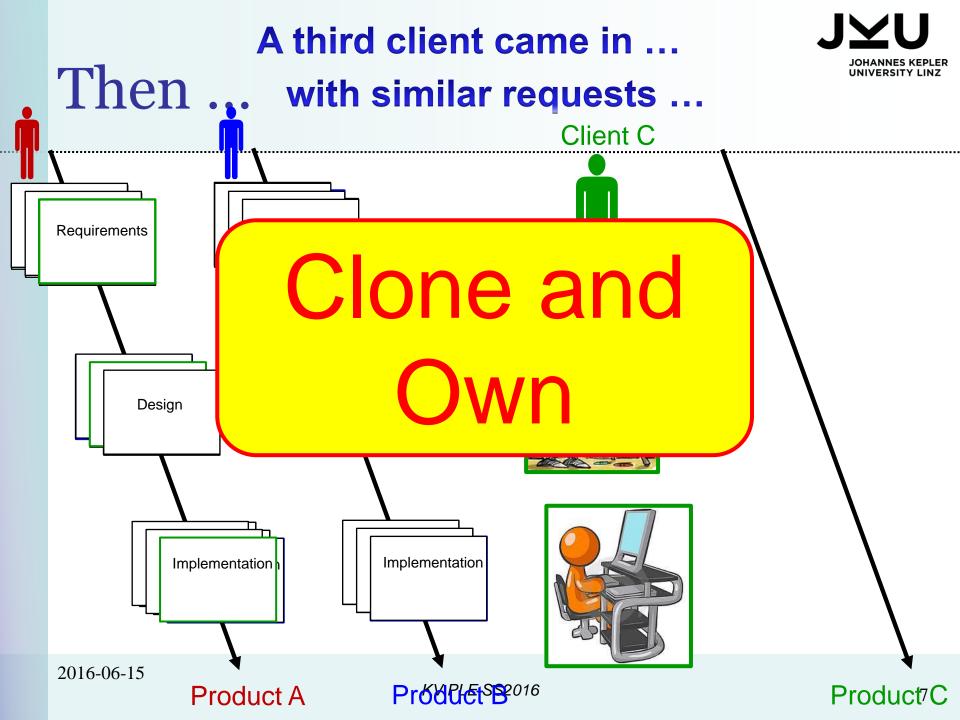
Product A

A second client came in ...



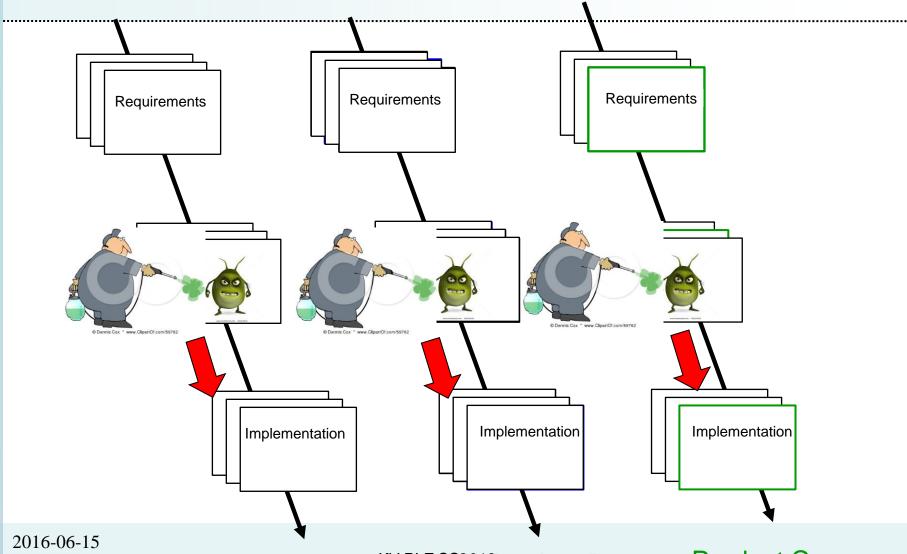
Then ... with similar requests ...





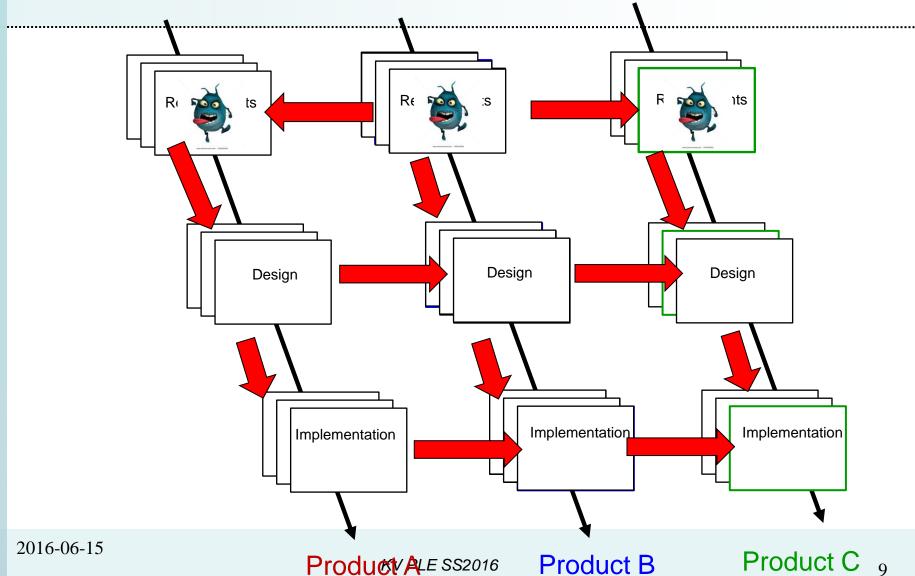


Then problems arise ...



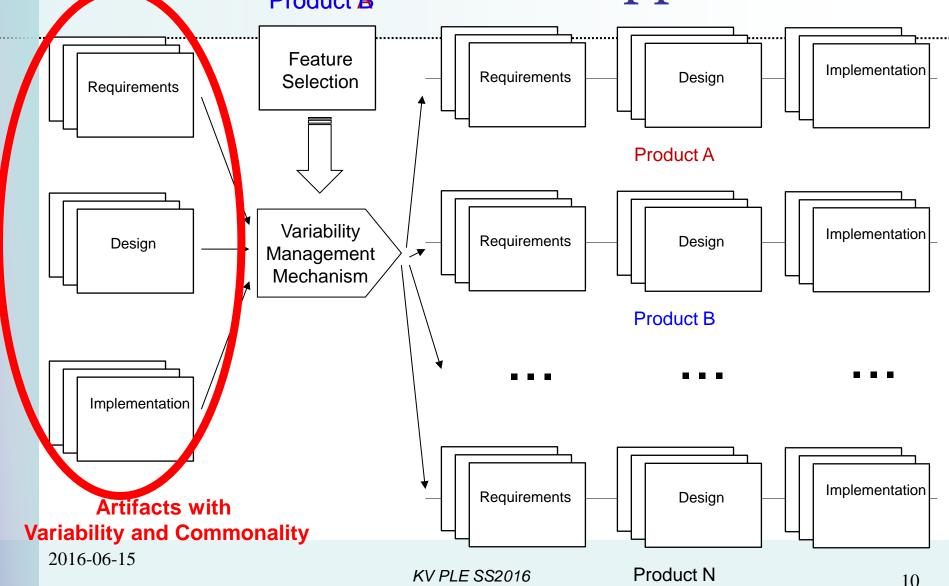


It gets even worse ...

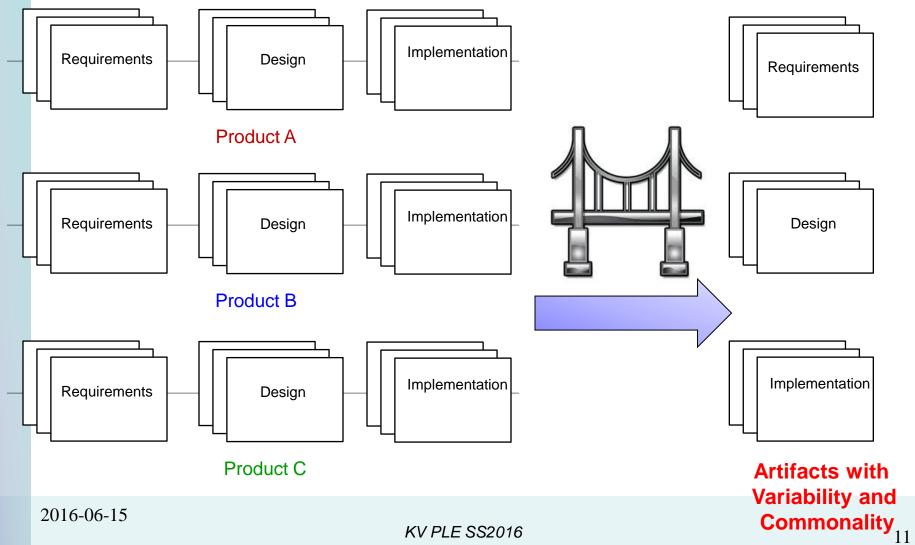




Software Product Line Approach



Reverse Engineering Variability



2016-06-15

KV PLE SS2016



VARiability In safety-critical Embedded Systems Project

VARIES Consortium





23 Partners

7 Countries

Belgium, United Kingdom, Norway, Finland, Denmark, Germany, Spain

Large Industries

Barco, Spicer, Autronica, Metso, Berner & Mattner, TÜV Süd

SMEs

Softkinetic Sensors, Macq, Mobisoft, HiQ, Pure-Systems, Hi-Iberia, Atego

Research Institutes

Sirris, FMTC, VTT, Imdea, Technalia, Fraunhofer, Vlerick, Sintef, IT University START

May 2012

DURATION

36 months

TOTAL INVESTMENT

13.2 M€





SEARCH-BASED SOFTWARE ENGINEERING (SBSE) – REMINDER



Search-Based Software Engineering (SBSE)

Search-Based Software Engineering focuses on the application of search-based optimization techniques to problems in software engineering [Harman10]

- Typical techniques are:
 - Basic searches, e.g. hill-climbing, simulated annealing
 - Techniques based on evolutionary computation



Hill Climbing Illustration

- Looks at a neighborhood of SampleSize states and selects the one with best fitness
- Main problem
 - Can get stuck in a local maxima

Algorithm 1 Steepest Ascent Hill Climbing

```
 X ← random initial state

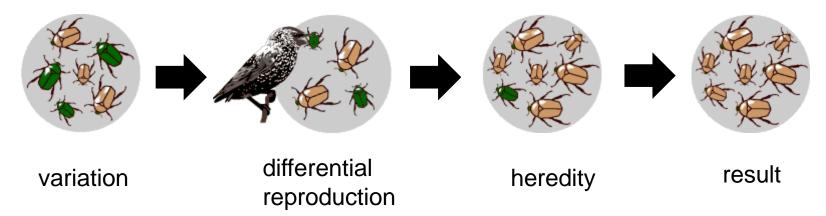
 2: I ← 0
 3: Best ← X
 4: while (I<MaxIter) ∧ (evaluate(Best) ≠ BestFitness) do</p>
        S \leftarrow 0
        while S < S ample S ize do
            X' \leftarrow move(Best)
            if evaluate(X') better than evaluate(X) then
                X \leftarrow X'
            end if
           S \leftarrow S + 1
    end while
        if evaluate(X) better than evaluate(Best) then
            Best \leftarrow X
        end if
        I \leftarrow I + 1
17: end while
18: return Best
```



Basic Ideas [Eibeno3]

Darwinian evolution:

- Given an environment able to host a limited number of individuals
- Basic instinct of individuals is to reproduce
- Natural Selection favours those that can compete for the available resources more effectively
 - a.k.a. survival of the fittest



© University of California Museum of Palaeontology's Understanding Evolution (http://evolution.berkeley.edu) 2016-06-15

KV PLE SS2016 16

Individuals, Mutation and Populations



Phenotypic traits

■ Behavioural and physical features of an individual that affect the response to the environment → fitness

Mutations

- Random variations on the phenotypic traits
- Can be accumulated to new combinations of traits

Population

- Consist on a number of individuals
- After time pases, because of reproduction and mutation the population changes



Evolutionary Computation

Evolutionary Computation

 Includes several stochastic search methods which computationally simulate the natural evolutionary process

Example techniques

- Genetic algorithms
 - Individuals are typically represented as binary strings, commonly used in numerical optimization problems
- Genetic programming
 - Individuals encode programs typically represented as treestructures whose fitness function evaluate how well the programs execute a computational task

Evolutionary Computation Illustration



- Randomly creates an initial population
- Evaluates the initial population
- At each generation
 - 1. select the individuals with best fitness
 - mutate their characteristics
 - 3. re-evaluate them

Algorithm 1 Basic Evolutionary Algorithm

- 1: $t \leftarrow 0$
- 2: initialize P(t)
- 3: evaluate P(t)
- 4: while not termination condition do
- 5: $t \leftarrow t+1$
- 6: select P(t) from P(t-1)
- 7: $mutate\ P(t)$
- 8: evaluate P(t)
- 9: end while



CHALLENGES IN RAISING A SOFTWARE FAMILY



Four Core Challenges

- 1. Know your family members
 - Reverse engineering feature models
- 2. Know the family members whereabouts
 - Traceability Feature-Artifact
- 3. Identify boundaries and enforce them
 - Safe composition and feature refactoring
- 4. Cope with growing pains
 - Evolution and Maintenance

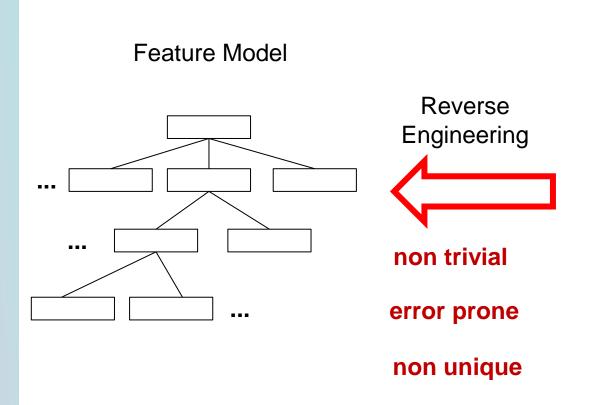


CHALLENGE 1. KNOW YOUR FAMILY MEMBERS



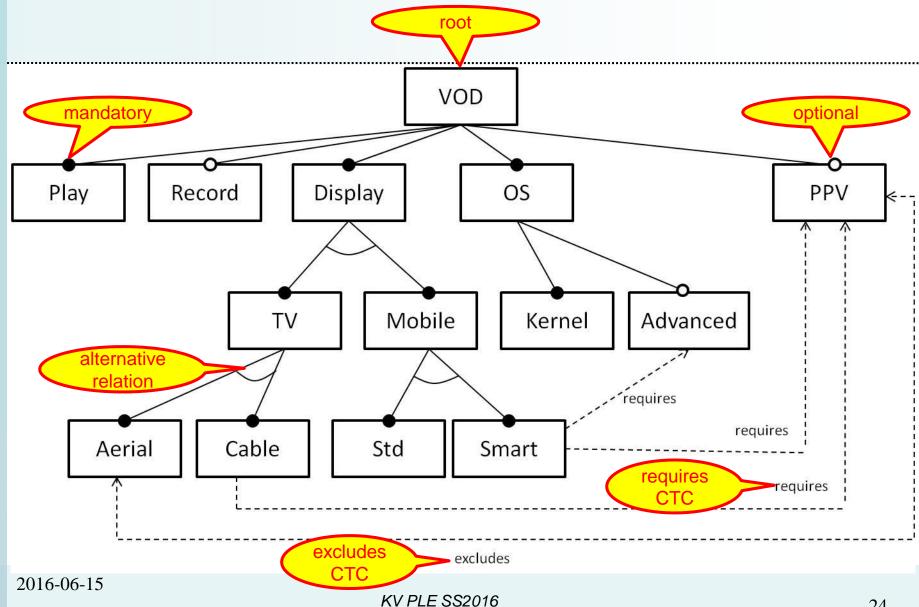
Big Picture

Goal: Model all the products of a SPL and their features



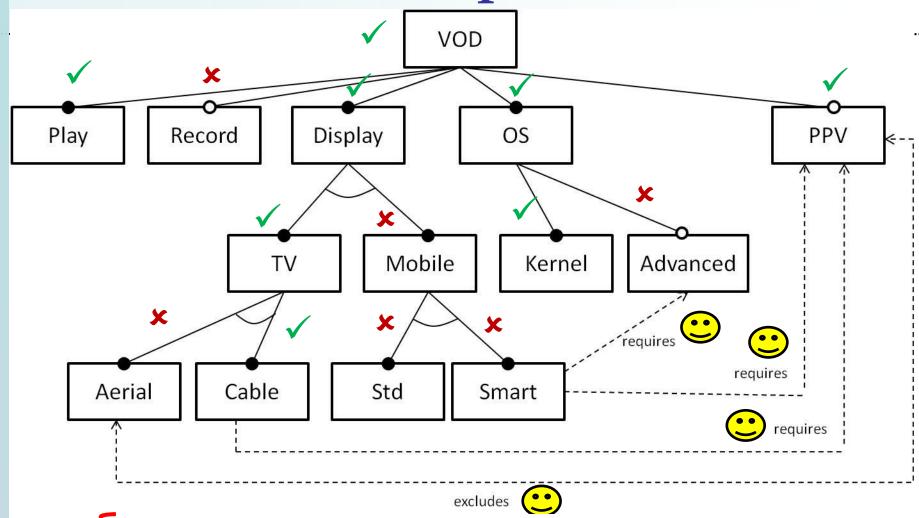
Feature Sets											
Α	В	С		N							
✓		✓									
	\checkmark			\checkmark							
✓	✓	✓									
		✓		\checkmark							
	✓	✓		✓							
\checkmark	\checkmark	✓									
	/	/		/							

Feature Model – Notation Examples KEPLER





Feature Set Example



<mark>feature</mark> 2<mark>9</mark>46-06-15 Selected {VOD, Play, Display, OS, TV, Cable, Kernel, PPV}

Not Selected {Record, Mobile, Aerial, Mobile, Std, Smart, Advanced}



Running Example -- Feature Set Table

	FSet		Play	Rec	Disp	OS	TV	Mob	Sm	Std	Ker	Adv	Aer	Cab	PPV
	FSI	√	V								./			√	√
	FS2	\checkmark	√		\checkmark	\checkmark	\checkmark				✓			\checkmark	V
	FS3	\checkmark	√	✓	✓	√	V				√		√		
	FS4	✓	✓		✓	✓	✓				✓		✓		
	FS5	√	√	✓	√	√		✓		✓	✓				√
	FS6	√	✓	✓	✓	√		✓		√	√				
	FS7	√	√		✓	✓		✓		✓	✓				
	FSS	\checkmark	\checkmark		✓	\checkmark		✓		✓	✓				✓
	FS9	√	√	✓	√	√	√				✓	✓		√	√
	FS1O	✓	√		✓	√	√				√	✓		√	✓
	FS11	✓	✓	✓	✓	✓	✓				✓	✓	✓		
	FS12	√	✓		✓	√	✓				✓	✓	✓		
	FS13	✓	✓	√	✓	✓		✓		√	✓	✓			✓
	FS14	√	✓	✓	✓	√		✓		√	√	√			
	FS15	√	√		✓	√		✓		√	√	✓			√
	FS16	✓	√		✓	✓		✓		√	✓	✓			
5	FS17	✓	✓	✓	✓	✓		✓	✓		✓	✓			✓
	FS18	✓	✓		✓	✓		✓	✓		✓	✓			✓

2016-06-15



Our contributions

- An ad hoc algorithm [WCRE11, FASE2013]
 - Provides one feature model solution with arbitray feature hierarchy
- Search-based approach based on genetic algorithm [SBSE12]
 - Can provide more than one solution alternatives

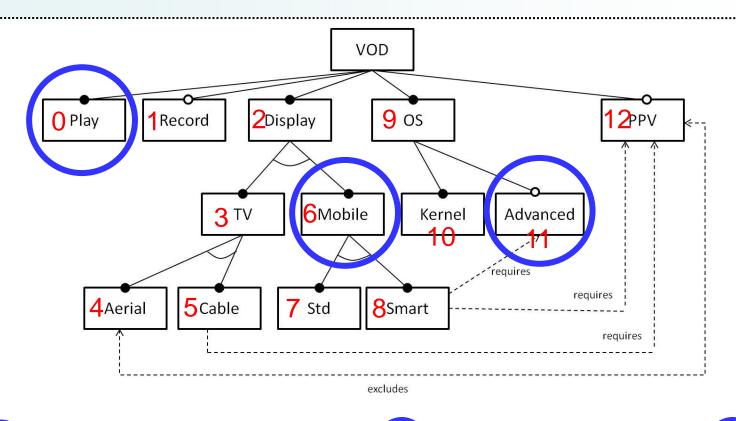


Feature Model Encoding (1)

- First part encodes the structure of the feature model
- Each chromosome is a tuple <PR, CN>
 - PR relation with parent
 - M mandatory
 - Op optional
 - Alt alternative relation
 - Or or relation
 - CN denotes the number of children
- A Depth-First Traversal determines the tuple order
 - Starting from the root of FM tree
 - The root is not encoded



Structural Encoding Example



L														_
	0	1	2	3	4	5	6	7	8	9	10	11	12	
V	M,0	Op,0	M,2	Alt,2	Alt,0	Alt,0	Alt,2	Alt,0	Alt,0	M,2	M,0	Op,0	Op,0	
N	M,0	Op,0	M,2	Alt,2	Alt,0	Alt,0	Alt,2	Alt,0	Alt,0	M,2	Μ,	0	0 Op,0	0 Op,0 Op,0

2016-06-15



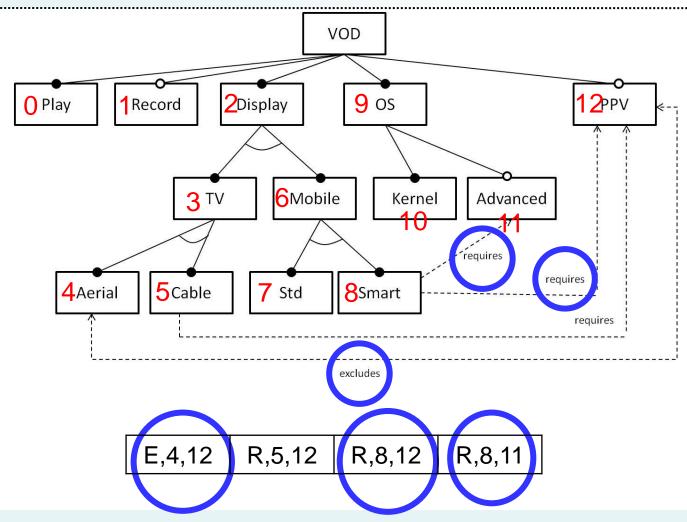
Feature Model Encoding (2)

Second part encodes the Cross-Tree Constraints

- Each chromosome is a tuple <TC, O, D>
 - TC type of constraint
 - R requires
 - E excludes
 - O origin feature denoted with DFT value
 - D destination feature denoted with DFT value

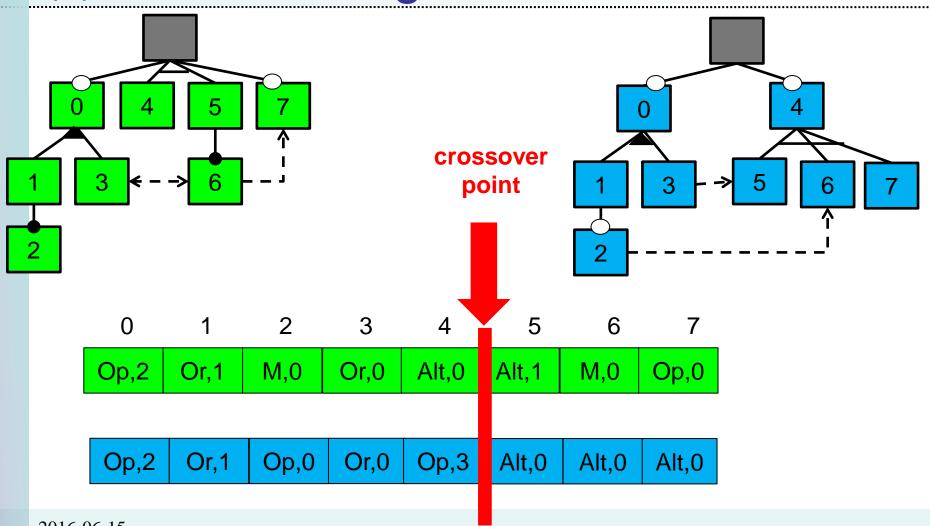


CTC Encoding Example



JAU JOHANNES KEPLER UNIVERSITY LINZ

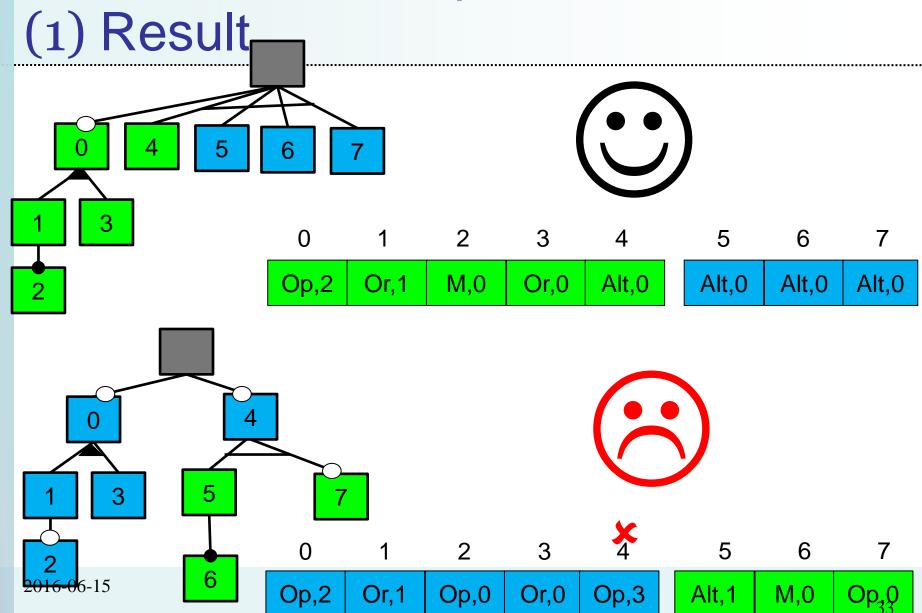
Crossover — One point (1) Feature Diagram



2016-06-15

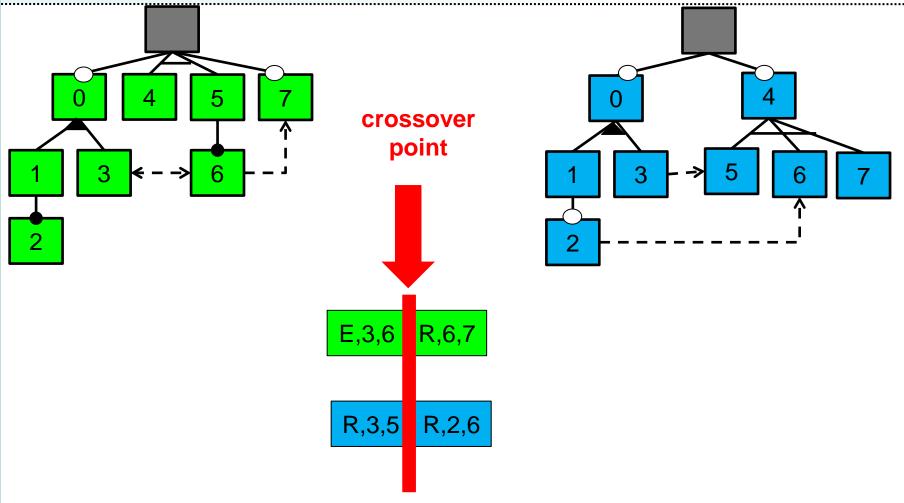
Crossover — One point





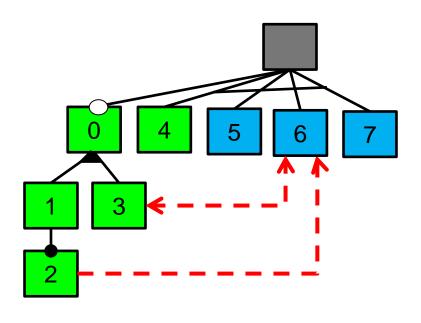


Crossover — One point (2) Cross-Tree Constraints





Crossover — One point (2) Result



E,3,6 R,2,6



Mutation Operators

- Four operators applied with a configurable probability
 - Operator 1. Changes randomly a relation between two features from one kind to any other kind. For example, from mandatory (M) to optional (Op) or from Op to Alternative (Alt).
 - Operator 2. Changes the number of children CN, to a number selected from 0 to a maximum branching factor parameter set up.
 - Operator 3. Changes the type of cross-tree constraint, from excludes to requires and vice versa.
 - Operator 4. Changes either the origin or destination feature (with equal probability) of a cross-tree constraint.
- Validity checks
 - Identification and repair of infeasible feature models
 - E.g. A CTC does not have the same origin and destination values.



Evaluation Overview

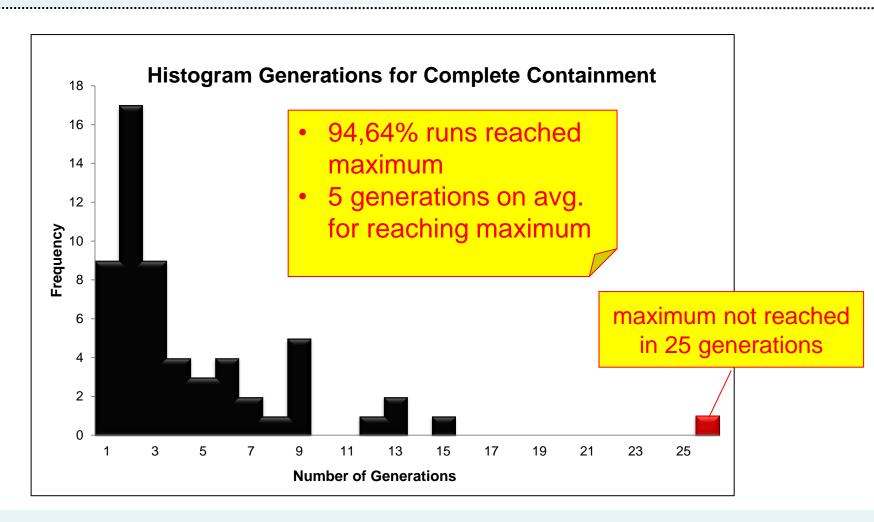
FFRelaxed(sfs, fm) = containedFSets(sfs, fm)

- Case studies
 - 59 feature models from SPLOT repository
 - No. products 1...896
 - No. features 9 ... 27
- Executions
 - 10 runs for each feature model
 - 16 cores at 2.40 GHz, 25GB RAM, Cent OS, Java 1.6

	Parameter	Value	
	Selection strategy	Roulette-wheel	
	Crossover strategy	One-point	
	Crossover probability	0.7	
	Mutation probability	0.01	
	Initial population size	100	
	Infeasible individuals	Replace	
	Maximum generations	25	

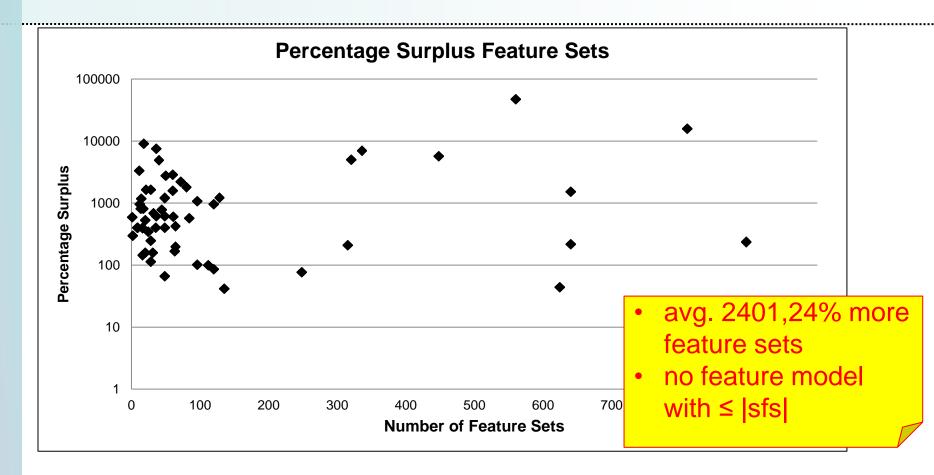


FFRelaxed Results (1)





FFRelaxed Results (2)



Surplus(sfs,fm) = $\underline{\text{products(fm) - |sfs|}} \times 100$ |sfs|



Open questions

- Non-binary feature property combinations
 - Not only yes/no but other real values, e.g. nonfunctional properties
- Effective use of domain knowledge to structure the feature hierarchy
 - For example based on ontologies
- More expressive feature model representations and operators
 - Genetic programming, variability-aware operators



CHALLENGE 2. KNOW THE FAMILY MEMBERS WHEREABOUTS



Big Picture

• Goal:

 Compute traces between features and the realization artifacts

Our contribution

 Basic algorithm to incrementally trace features and feature interactions to artifact fragments [SPLC13]





```
class Line {
   Point startPoint, endPoint;
   Line(Point start) {...}
   void paint(Graphics g) { ... }
   void setEnd(Point end) {...}
}
class Canvas {
   List<Line> lines = new LinkedList<Line>();
   void paintComponent(Line I) { ... }
}
```

	Variant V1
c1	Point Line.startPoint
c2	Point Line.endPoint
сЗ	Line.Line(Point)
c4	void Line.paint(Graphics)
c5	void Line.setEnd(Point)
с6	List Canvas.lines
c7	void Canvas.paintComponent(Line)



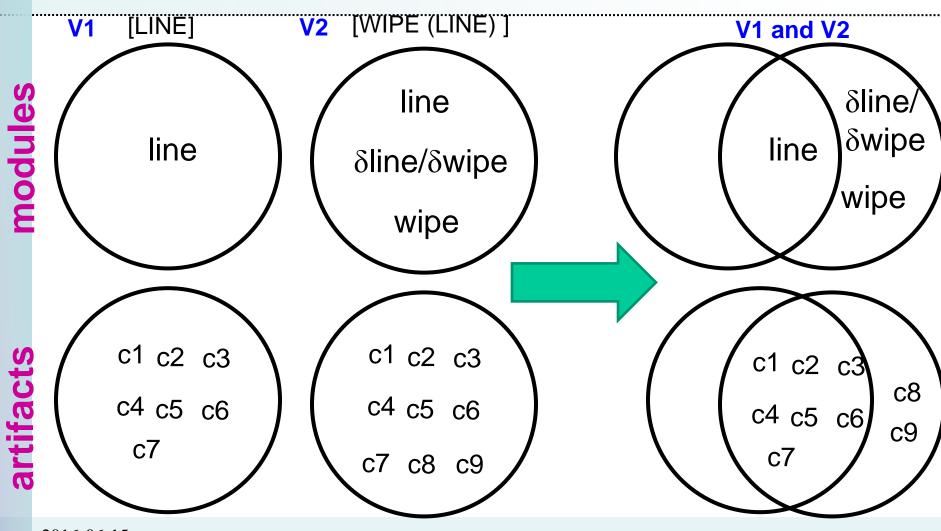


```
class Line {
 Point startPoint, endPoint;
 Line(Point start) {...}
 void paint(Graphics g) {...}
 void setEnd(Point end) {... }
class Canvas {
 List<Line> lines = new LinkedList<Line>();
 void paintComponent(Line I)
 void wipe() {
   lines.clear(); -
   repaint();
```

	Variant V2
c1	Point Line.startPoint
c2	Point Line.endPoint
c3	Line.Line(Point)
c4	void Line.paint(Graphics)
c5	void Line.setEnd(Point)
с6	List Canvas.lines
c7	void Canvas.paintComponent(Line)
с8	void Canvas.wipe()
с9	lines.clear()



First Traceability Refinement



2016-06-15

KV PLE SS2016



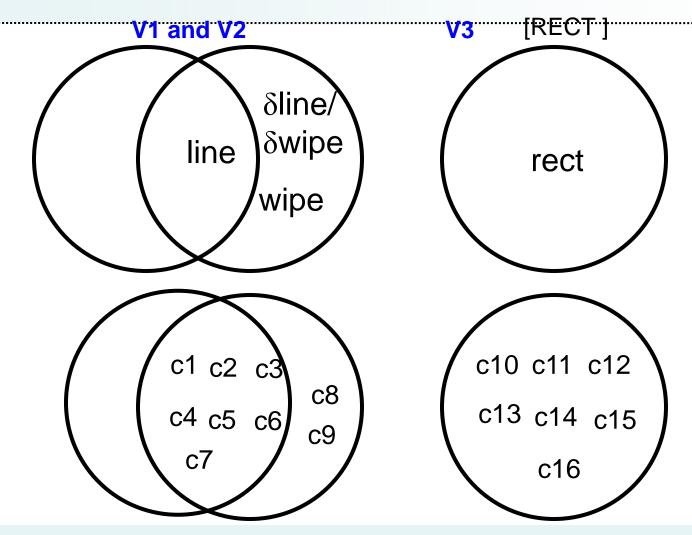


```
class Rectangle {
    Point upperPoint, lowerPoint;
    Rectangle(Point x, Point y) { ... }
    void paint(Graphics g) { ... }
    void setEnd(Point x) { ... }
}
class Canvas {
    List<Rectangle> rectangles = new
    LinkedList<Rectangle>();
    void paintComponent(Rectangle rect) { ... }
}
```

	Variant V3	
c10	Point Rectangle.upperPoint	
c11	Point Rectangle.lowerPoint	
c12	Rectangle.Rectangle(Point, Point)	
c13	void Rectangle.paint(Graphics)	
c14	void Rectangle.setEnd(Point)	
c15	List Canvas.rectangles	
c16	void Canvas.paintComponent(Rectangle)	

JOHANNES KEPLER UNIVERSITY LINZ

Second Traceability Refinement





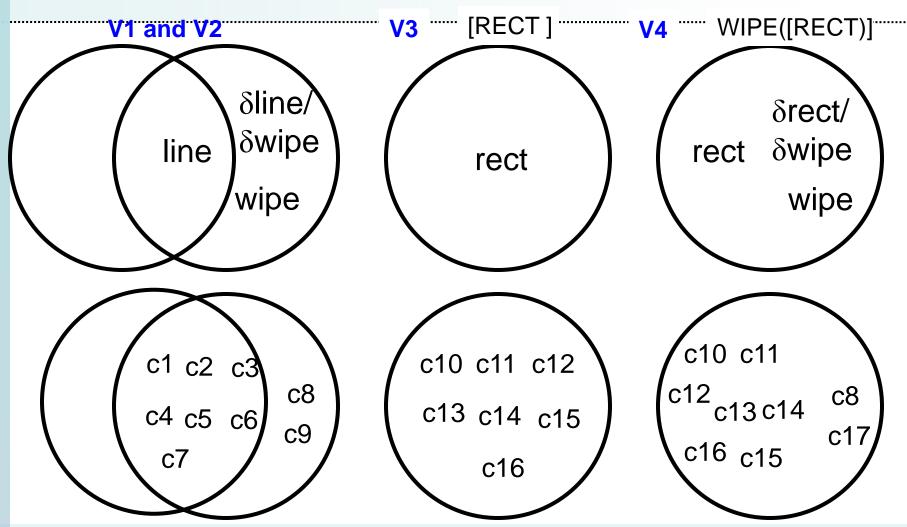


```
class Rectangle {
 Point upperPoint, lowerPoint;
 Rectangle(Point x, Point y) { ... }
 void paint(Graphics g) { ... }
 void setEnd(Point x) { ... }
class Canvas {
List<Rectangle> rectangles = new
LinkedList<Rectangle>();
void paintComponent(Rectangle rect) { ... }
void wipe() {
   rectangles.clear();
  repaint();
```

	Variant V4	
c10	Point Rectangle.upperPoint	
c11	Point Rectangle.lowerPoint	
c12	Rectangle.Rectangle(Point, Point)	
c13	void Rectangle.paint(Graphics)	
c14	void Rectangle.setEnd(Point)	
c15	List Canvas.rectangles	
c16	void Canvas.paintComponent(Rectangle)	
c8	void Canvas.wipe()	
c17	rectangles.clear();	



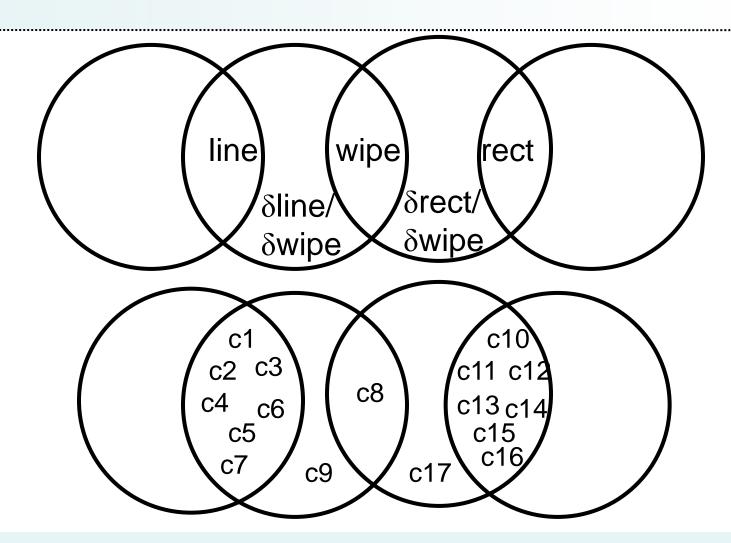
Third Traceability Refinement (1)



2016-06-15



Third Traceability Refinement (2)





Evaluation Examples

	VOD	ArgoUML	MM
Mandatory Features	6	3	6
Optional Features	5	8	8
Number of Variants	32	256	7
Lines of Code	5.3K+	340K+	5K+
Classes	42	1915	50
Fields	392	4452	223
Methods	249	16676	422
Unique Code Pieces	641	21128	645
Correctness %	100	99.4	99.6



Limitations

- Current evaluation
 - Based on synthesized variants from annotated programs
 - More "realistic" and larger examples are being collected
- Coarse grain level field, methods, basic interactions
 - Type 1 clones
- Current implementation only for Java-based systems
 - Working on extensions for EMF-based representation
- Current algorithm does no exploit
 - Runtime information execution traces
 - Development history
 - More advanced diffing and clone detection technologies



CHALLENGE 3. IDENTIFY BOUNDARIES AND ENFORCE THEM



Big Picture

- We want to answer the following questions:
- 1. How can we find if there is an error in a product?
 - For example, that we do not have references to nonexisting elements?
- 2. What can we do if we find an error in a product?
 - Can we "fix" it? How?



Our contributions

- Safe composition for multi-view models [ECMFA10, ICSR11]
 - Detection of inconsistency in models with variability
- Catalogue of Feature Oriented Refactoring patterns [SPLC11]
 - Patterns for moving code fragments across feature boundaries



Safe Composition in MVM (1)

- Multi-View Modeling (MVM)
 - Common modeling practice
 - Use of multiple yet related views is advocated
 - Example: UML multiple views
- Consistency checking
 - Description and verification of semantic relationships among views
- Challenge
 - How to detect inconsistencies in MVM with variability?
- Our approach
 - Safe composition programming languages
 - guarantee that all programs of a product line are type safe
 - All models that can be composed in a product line are type safe



57

Safe Composition in MVM (2)

- Let
 - PL_f domain constraints from the product line IMP_f an implementation constraint in MV models
- Using a SAT solver we can check if one propositional formula is satisfiable or not
- Our interest is verifying that all the product line members satisfy an implementation constraint

$$\neg$$
 ($PL_f \Rightarrow IMP_f$)

Unsatisfiable = there is <u>no</u> product that violates the constraint

Satisfiable = there is at least one product that violates the constraint

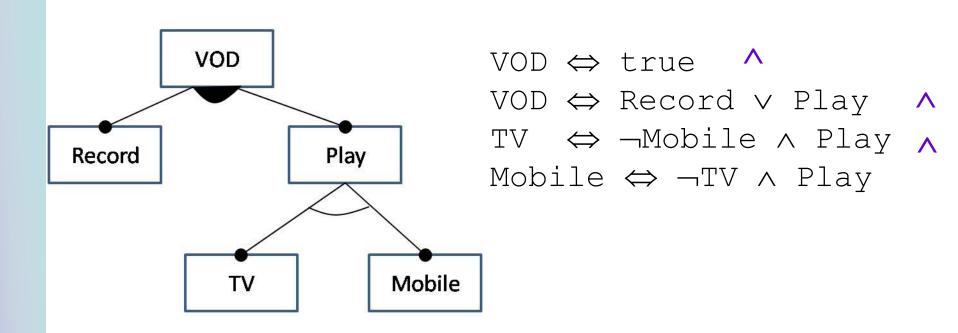


Computing PL_f

Name	Diagram Notation	Propositional Logic
Mandatory	P	$P \Leftrightarrow C$
Optional		$C \Rightarrow P$
Alternative	F1 F2 F3	$(F1 \Leftrightarrow (\neg F2 \land \neg F3 \land P)) \land (F2 \Leftrightarrow (\neg F1 \land \neg F3 \land P)) \land (F3 \Leftrightarrow (\neg F1 \land \neg F2 \land P))$
Or	F1 F2 F3	<i>P</i> ⇔ <i>F</i> 1 ∨ <i>F</i> 2 ∨ <i>F</i> 3
Requires Cross feature arrow		$A \Rightarrow B$
Excludes Cross feature arrow		$A \Rightarrow \neg B \equiv \neg (A \land B)$



Example PL_f





How to compute IMP_f?

- In MVM consistency rules
 - Establish semantic relationships among elements
 - We regard them as implementation constraints
- We identified classes of rules depending on their consistency validation
- We create one IMP_f for each constraint instance that we need to check



Requiring Rules

Each constraint instance IMP_f is computed as follows

$$IMP_f \equiv F \Rightarrow V_{i=1..k} Freq_i$$

where

F is the feature that requires another feature(s)

 $\mathtt{Freq}_\mathtt{i}$ are the features that satisfy requirement of \mathtt{F}

▶ Considering PL_f

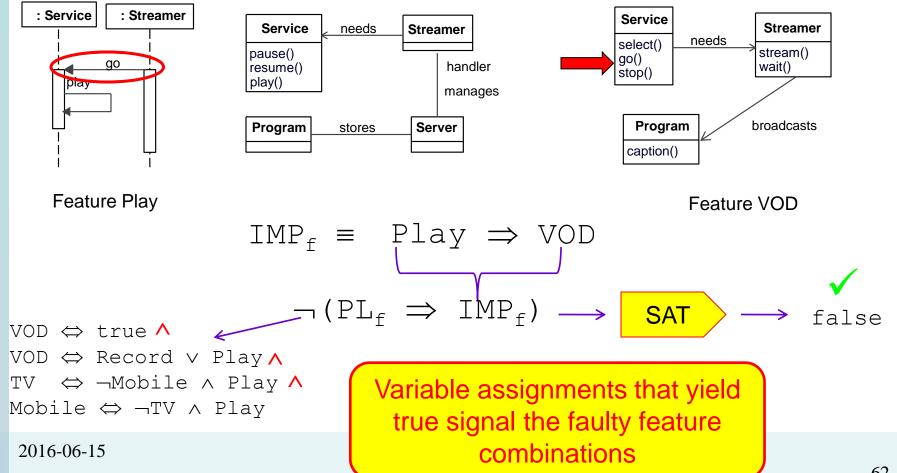
$$\neg (PL_f \Rightarrow IMP_f) \equiv PL_f \land F \land \neg Freq_i$$

i=1..k KV PLF SS2016



Requiring Rule Example

R5. Message action must be defined as an operation in receiver's class



62

Feature Oriented Refactoring Patterns

JOHANNES KEPLER UNIVERSITY LINZ

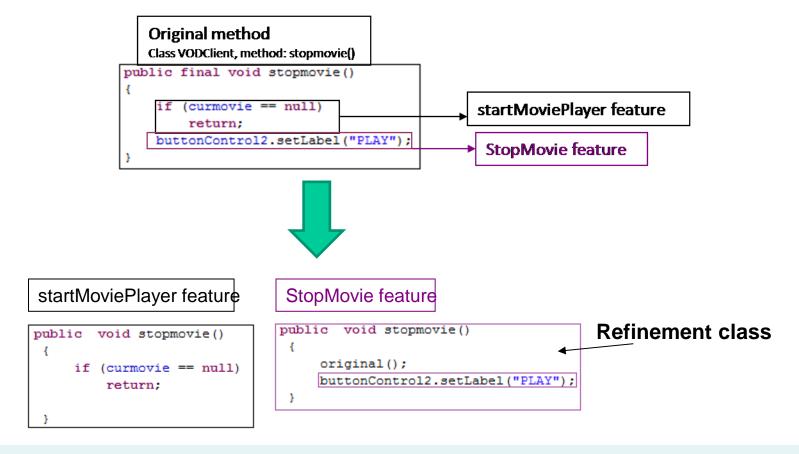
- We created two SPLs from two existing single programs by taking them appart into features
 - Starting with list of requirements, each describing a feature
 - Target features modularized with FOSD

Num.	Level	Name of refactoring
1	Method	Addition at the end of the method
2	Method	Addition anywhere with a hook method
3	Method	Addition at the beginning of the method
4	Method	Overwrite method
5	Method	Move entire method
6	Attribute	Move field
7	Attribute	Remove access declaration
8	Class	Move entire class



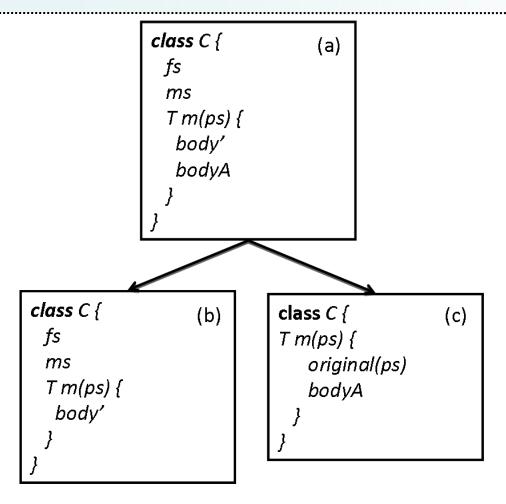
Pattern Example

Refactoring at end of the method





Pattern Example Overview



Constraints

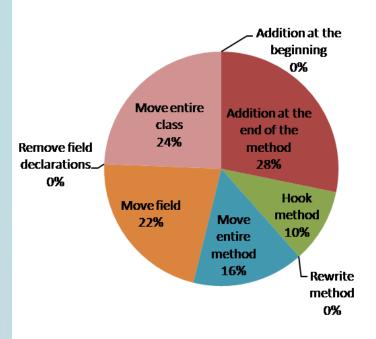
-bodyA cannot usevariables defined by body'-bodyA uses variable in psif not modified by body'



Summary of Case Studies

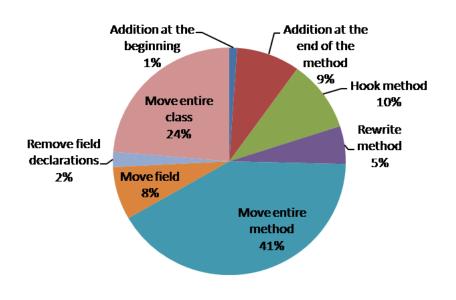
VOD Player

3.6 KLOC, 15 features,42 classes



Gantt Chart

41 KLOC, 16 features,43 packages





Making the connection with SBSE ...

Safe Composition



Feature Oriented Refactoring



Open Question:

Can we leverage work from refactoring + SBSE?



CHALLENGE 4. COPE WITH GROWING PAINS



Evolution scenarios

- Features are added, deleted, or their relationships changed in a feature model
- Realizing artifacts change feature renovation
- Challenge:
 - How to cope with the co-evolution of variability and its realization?
- Maintenance activities
 - Software Testing



Relevant References (1)

- E. N. Haslinger, R. E. Lopez-Herrejon, and A. Egyed, "On extracting feature models from sets of valid feature combinations", FASE 2013.
- R. E. Lopez-Herrejon, J. F. Chicano, J. Ferrer, A. Egyed, and E. Alba, "Multi-objective optimal test suite computation for software product line pairwise testing", ICSM 2013
- L. Linsbauer, R. E. Lopez-Herrejon, and A. Egyed, "Recovering traceability between features and code in product variants", SPLC 2013.
- R. E. Lopez-Herrejon and A. Egyed, "Towards interactive visualization support for pair-wise testing software product lines", VISSOFT 2013
- R. E. Lopez-Herrejon, J. A. Galindo, D. Benavides, S. Segura, and A. Egyed, "Reverse engineering feature models with evolutionary algorithms: An exploratory study", SSBSE 2012.
- E. N. Haslinger, R. E. Lopez-Herrejon, and A. Egyed, "Reverse engineering feature models from programs' feature sets", WCRE 2011.
- R. E. Lopez-Herrejon, L. Montalvillo-Mendizabal, and A. Egyed, "From requirements to features: An exploratory study of feature-oriented refactoring", SPLC 2011.
- M. Alferez, R. E. Lopez-Herrejon, A. Moreira, V. Amaral, and A. Egyed, "Supporting consistency checking between features and software product line use scenarios", in ICSR 2011.
- R. E. Lopez-Herrejon, A. Egyed, "Detecting inconsistencies in multi-view models with variability", ECMFA 2010.



Relevant References (2)

- A. Eiben and J. Smith, Introduction to Evolutionary Computing. Springer Verlag, 2003
- M. Harman, P. McMinn, J. T. de Souza, and S. Yoo, Search Based Software Engineering: Techniques, Taxonomy, Tutorial. LASER Summer School 2010.
- M. Harman, S. Afshin Mansouri, Y. Zhang. Search-Based Software Engineering: Trends, Techniques and Applications. ACM Computing Surveys 2012.
- Roberto E. Lopez-Herrejon, Javier Ferrer, Francisco Chicano, Alexander Egyed, Enrique Alba. Comparative Analysis of Classical Multi-Objective Evolutionary Algorithms and Seeding Strategies for Pairwise Testing of Software Product Lines. Accepted at IEEE Congress on Evolutionary Computation (IEEE CEC), Beijing, China, July 2014.
- Roberto E. Lopez-Herrejon, Javier Ferrer, Francisco Chicano, Evelyn Nicole Haslinger, Alexander Egyed, Enrique Alba. A Parallel Evolutionary Algorithm for Prioritized Pairwse Testing of Software Product Lines. GECCO 2014.