



## A DELTA-ORACLE FOR FAST MODEL MERGE CONFLICT ESTIMATION USING SKETCH-BASED CRITICAL PAIR ANALYSIS

Karl Kegel<sup>1</sup>, Andreas Domanowsky<sup>1</sup>, Kevin Feichtinger<sup>2</sup>, Romain Pascual<sup>2</sup>, Uwe Aßmann<sup>1</sup>

<sup>1</sup>Dresden University of Technology <sup>2</sup>Karlsruhe Institute of Technology

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - SFB 1608 - 501798263



$$(\Delta)$$

#### **GET THE NUMBER OF MODEL MERGE CONFLICTS FAST**

# ... MAYBE NOT ALL OF THEM ... MAYBE TOO MUCH

#### THREE WAY MODEL MERGE





A merge combines two models into one
 Three-way merges are aware of a common ancestor
 Successful if models are "conflict-free"

But: Merge conflicts are difficult to anticipate
 Resolving conflicts is costly
 Model merging itself is costly

Tools & metrics for conflict monitoring required
 Previous work on the *Drift* metric<sup>1</sup>



(1) Karl Kegel, Sebastian Götz, Ronny Marx, Uwe Aßmann:

## **EXCURSE: THE DRIFT METRIC FOR INCONSISTENCY MONITORING**





- Each pair of variants has a measurable number of conflicts/distance
- A set of variants has a conflict potential
   = difficulty of merging two arbitrary variants
- 1. Capture all pairwise distances in the distance matrix
- 2. Apply multi-dimensional scaling of the matrix into 3D
- 3. Measure the average median deviation
  - = scatteredness = *Drift*<sup>1</sup>



(1) Karl Kegel, Sebastian Götz, Ronny Marx, Uwe Aßmann:

## **EXCURSE: THE DRIFT METRIC FOR INCONSISTENCY MONITORING**





- Each pair of variants has a measurable number of conflicts/distance
- A set of variants has a conflict potential
   = difficulty of merging two arbitrary variants
- 1. Capture all pairwise distances in the distance matrix
- 2. Apply multi-dimensional scaling of the matrix into 3D
- 3. Measure the average median deviation
  - = scatteredness = *Drift*<sup>1</sup>



(1) Karl Kegel, Sebastian Götz, Ronny Marx, Uwe Aßmann:

## **EXCURSE: THE DRIFT METRIC FOR INCONSISTENCY MONITORING**





- Each pair of variants has a measurable number of conflicts / distance
- A set of variants has a conflict potential
   = difficulty of merging two arbitrary variants
- 1. Capture **all pairwise distances** in distance matrix
- 2. Apply multi-dimension
- 3. Measure the average
  - = scatteredness = A

(number of variants)<sup>2</sup> / 2
model comparisons



(1) Karl Kegel, Sebastian Götz, Ronny Marx, Uwe Aßmann:



# HYBRID CONFLICT MONITORING



## **HYBRID CONFLICT MONITORING**





#### CLAIM

To raise awareness of conflicts, their exact number is not important
 Fast feedback is important!

□ Full model merges are accurate but slow, e.g., *EMFCompare*<sup>1</sup>

□ *Operation-based Versioning*<sup>2</sup> is fast but difficult to set up

□ Fast, because its runtime is independent of model size!

#### □ But we can build upon the idea!

(1) EMFCompare, Eclipse Foundation https://eclipse.dev/emf/compare/

(2) Edvin Herac, Wesley K. G. Assun..o, Luciano Marchezan, Rainer Haas, and Alexander Egyed. 2023. A flexible operation-based infrastructure for collaborative model-driven engineering

(2) Alfa Yohannis. 2020. Change-Based Model Differencing and Conflict Detection

## **HYBRID CONFLICT MONITORING**





#### IDEA

Use a fast but inaccurate algorithm for fast feedback = ORACLE
 Use a traditional comparison for occasional validation = VALIDATOR





# FAST CONFLICT ESTIMATION: SKETCH-BASED CRITICAL PAIR ANALYSIS (SCPA)





 $(\Delta)$ 

Use ideas from operation-based versioning
 BUT simplify as much as possible!
 Critical pair analysis as base concept



#### A GENERIC GRAPH METAMODEL







- 1. Capture change operations (delta operations) Detailed difference metamodel
- 2. Simplify delta operations to delta operations sketches Simplified operation langua



- **1. Iterate delta sketch once to reduce internal redundancy**
- 2. Flatten the delta sketch into a set
- 3. Compare two sketches by comparing all pairs via the *Critical Pair Heuristic*

#### 4. Return the number of found conflicting pairs

A Delta-Oracle for Fast Model Merge Conflict Estimation using Sketch-based Critical Pair Analysis // Karl Kegel, Andreas Domaowsky, Kevin Feichtinger, Romain Pascual, Uwe Aßmann 10















Preprocessing / sketch building Preprocessing / sketch building **Critical pair detection** (Delete, N42) (Add, N10) (Add, N54) (Add, N11) **CP** Heuristic (Delete, E12) (Move, N22) (Delete, N43) (Delete, E12) (Delete X)  $\langle - \rangle$  (Update X) (Move, N22) (Update, N43) (Move X) <-> (Move X) (Update, N12) (Update X) <-> (Update X)







# **EVALUATION**

#### A Delta-Oracle for Fast Model Merge Conflict Estimation using Sketch-based Critical Pair Analysis // Karl Kegel, Andreas Domaowsky, Kevin Feichtinger, Romain Pascual, Uwe Aßmann 16

#### **EVALUATION**

#### **EVALUATION OBJECTIVE**

- □ Influence of model properties on SCPA accuracy
- □ Influence of edit behaviour on SCPA accuracy
- □ Influence of edit sequence length on SCPA accuracy

□ Comparison SCPA with EMFCompare (time & accuracy)

#### **Baseline**

Generated hierarchical graph models as base data set
 EMFCompare 3-way conflict count as ground-truth



## **EVALUATION SETUP**



- **1.** Generate a large number of graph models with different properties<sup>1</sup>
- 2. For each: generate two variants and edit them n-steps (<= 1000 edits)
  - 1. Serialize edits using delta metamodel
- 3. Calculate the number of conflicts for each variant pair for each step (> 5000 comparisons)
  - **1.** Using EMFCompare
  - 2. Using SCPA
- 4. Compare runtime and conflict count
- 5. Analyse correlations

(1) Karl Kegel & Collaborators: GraphGentool https://github.com/convidev-tud/emf-graph-gen

### **EVALUATION RESULTS**





- SCPA fails for models without unique IDs
- No significant influence on accuracy caused by model parameters
- Low influence of edit strategy (contextual problems caused by movements)
- Low influence of sequence length (slowly accumulating inaccuracies)
- ✓ Overall, good accuracy for the analyzed datase correlation coefficients > 0.85 up to 0.95
- Speed gain dependent on model properties (SCPA around 10x faster)



#### **EVALUATION RESULTS**









## CONCLUSION

CLUSION

#### CONCLUSION



**Comparing models for conflict monitoring is costly** 

- Possible speedup by trading accuracy for time
- Proposition: hybrid comparison pipeline for combining fast and slow comparison approaches

#### □ SCPA as a conflict oracle using edit operations

- **Runtime depends on edit size, not model size**
- Evaluation using generated graph models with varying properties
- Short runtime retaining good accuracy

#### **Given Setup Setup**

- **Evaluate more diverse metamodels**
- Integrate SCPA into generic tools
- Use SCPA to monitor conflicts in heterogenous models (e.g., views)



## THANK YOU FOR YOUR ATTENTION

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - SFB 1608 - 501798263