

## CLONE DETECTION IN REPOSITORIES OF BUSINESS PROCESS MODELS

Reina Uba

### Motivation

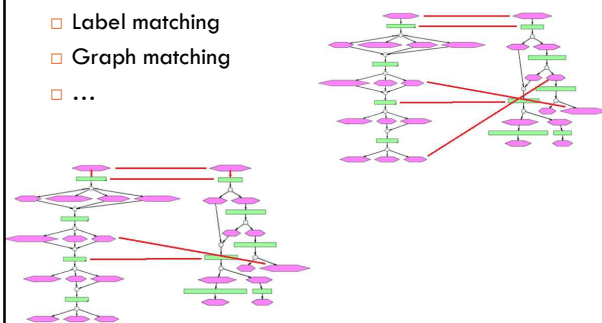
- Companies in various industries maintain collections of hundreds or even thousands of business process models, with sizes ranging from dozens to hundreds of elements per model

### Process management

- Similarity search
- Process merging
- Clone detection

### Similarity search and merging

- Label matching
- Graph matching
- ...



### Clone detection

- managing the overlap between models:
  - ▣ new process models are created by copying and merging fragments from other models

### Problems with clones

- clones make individual process models larger than they need to be
- clones are modified independently, sometimes by different stakeholders, leading to unwanted inconsistencies across models
- process model clones hide potential efficiency gains

### Clone detection

- Pairwise comparison
- The index structure

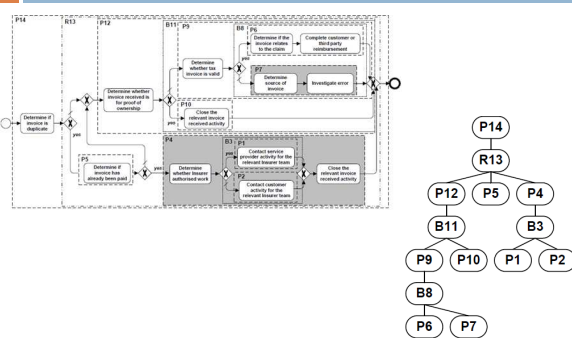
### The index structure

- All retrieved clones are single-entry, single-exit (SESE) fragments
- All retrieved clones are exact clones so that every occurrence can be replaced by an invocation to a single subprocess
- Only maximal clones are retrieved
- Retrieved clones must have at least two nodes (no "trivial" clones)

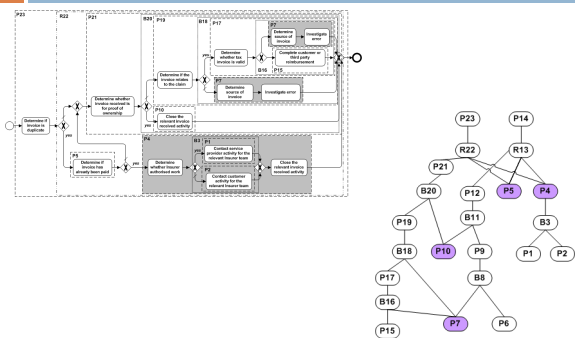
### SESE fragments, RPST

- The RPST is a parsing technique that takes as input a process model and computes a tree representing a hierarchy of SESE fragments
  - ▣ Trivials
  - ▣ Polygons
  - ▣ Bonds
  - ▣ Rigids
- We construct RPS Dag that consists all RPSTs of the process models

### RPST



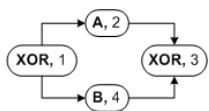
### RPST



### How to know if region already exists?

- Graph isomorphism
- Canonical hash code

## Canonical hash code



$$\begin{pmatrix} A & B & \text{XOR1} & \text{XOR3} \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} A & B & \text{XOR3} & \text{XOR1} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}$$

A.B.XOR.XOR.X.0.0.0.1.1.0.0.0.1.2.1.1.0.0.3.0.0.0.0.X

A.B.XOR.XOR.X.0.0.1.0.1.0.0.1.0.2.0.0.0.0.3.1.1.0.0.X

## Results

modelset	# models	min/max/average # nodes	Total # nodes	insertion time (ms)	# clones	refactoring gain
SAP	469	5 / 119 / 21.8	13254	14.05	279	1558 (11.75%)
Insurance	121	5 / 84 / 31.9	9815	21.32	118	356 (3.63%)
BIT A	269	5 / 47 / 17.3	4577	9.28	90	284 (6.2%)
BIT B3	247	5 / 39 / 13	3191	9.25	28	332 (10.4%)

Thank you!