

De Onderzoeksgroep  
Software Languages Lab

nodigt U graag uit op de openbare verdediging van het proefschrift van

## Ward Muylaert

ter behaling van de graad van Doctor in de Wetenschappen

Titel van het proefschrift:

### Data flow and Control flow Analysis of Problematic Commits

Promotor:

**Prof. dr. Coen De Roover**

De verdediging heeft plaats op

**Maandag 22 april 2024 om 16u in  
promotiezaal D.2.01**

**Samenstelling van de jury**

Prof. dr. Viviane Jonckers (VUB, voorzitter)

Prof. dr. Lynn Houthuys (VUB, secretaris)

Prof. dr. Kris Steenhaut (VUB)

Prof. dr. Josep Silva Galiana (Universitat  
Politècnica de València, Spanje)

Prof. dr. Jean-Rémy Falleri (Université de  
Bordeaux, Frankrijk)

### Curriculum vitae

Ward Muylaert behaalde een bachelor diploma wiskunde en een master diploma computerwetenschappen aan de VUB. Doorheen zijn PhD publiceerde hij vier papers die zijn werk ondersteunen. Ward was over de jaren heen assistent voor drie verschillende 1<sup>e</sup> bachelorvakken en begeleidde een bachelor- en twee masterstudenten met hun thesis. In zijn vrije tijd heeft Ward bijgedragen aan verschillende open-source projecten. Hij is lid van de raad van bestuur van een vzw met een speciale focus op kansarme kinderen en was in het verleden lid van de raad van bestuur van een andere vzw gericht op data- en kennisverspreiding.

### Abstract van het doctoraatsonderzoek

Bij het schrijven en onderhouden van programma's gebruiken softwareontwikkelaars versie controle software om verschillende versies van de broncode op te slaan. Versie controle software gebruikt commits als bouwstenen. Commits hebben een dubbele rol: ze representeren een versie van het programma op een zeker tijds punt; ze representeren ook de veranderingen ten opzichte van de vorige versie van het programma.

In deze dissertatie onderzoeken we twee types commits die moeilijker verstaanbaar zijn voor ontwikkelaars. Eerst kijken we naar samengestelde commits. Samengestelde commits groeperen verschillende ongerelateerde veranderingen. De veranderingen behandelen verschillende taken, zoals het herstellen van een bug of het toevoegen van een eigenschap. Behalve moeilijker verstaanbaar te zijn, zijn samengestelde commits ook moeilijker voor ontwikkelaars om ongedaan te maken of te integreren en voor onderzoekers om te analyseren. Ten tweede behandelen we samenvoegcommits. Versies van het programma divergeren wanneer verschillende ontwikkelaars aan verschillende eigenschappen werken of bugs herstellen. Een samenvoegcommit voegt de gedivergeerde versies weer samen. Indien de broncode overlapt of het gedrag mekaar beïnvloedt, dan moet de ontwikkelaar een oplossing voorzien na eerst de broncode van de verschillende versies te begrijpen.

Wij stellen een algoritmische onderbouwing voor hulpmiddelondersteuning voor beide types van problematische commits voor. Ons eerste algoritme ontwaart samengestelde commits aan de hand van een dataverloop gebaseerde aanpak. Ons algoritme gebruikt een programma afhankelijkheidsgraaf en fijnmazige veranderingen aan de abstracte syntaxboom. Het algoritme groepeerde de fijnmazige veranderingen op basis van de stukken uit de programma afhankelijkheidsgraaf waar ze toe behoren. Ons tweede algoritme analyseert samenvoegcommits aan de hand van een controleverloop gebaseerde aanpak. Het gebruikt symbolische executie om padcondities van de verschillende versies van het programma te verzamelen. We definiëren de programmasemantiek in functie van deze padcondities. De padcondities worden vergeleken met regels die de aanwezigheid van een semantisch samenvoegsconflict aanduiden.

We evalueren beide algoritmes. Door het analyseren, verfijnen en gebruiken van een erkende dataset van samengestelde commits besluiten we dat ons ontwaralgoritme kan vaststellen of een commit samengevoegd is. De groepen fijnmazige veranderingen zijn over het algemeen kleiner dan de taken in de commit, maar blijven erdoor begrensd. We evalueren ons semantisch samenvoegsconflictalgoritme op twee manieren. Eerst evalueren we de correctheid aan de hand van mutatietesten. Ten tweede evalueren we het empirisch door het toe te passen op echte samenvoegcommits. We beschrijven ook uitdagingen bij het empirisch evalueren van semantische samenvoegsconflicten. Onze evaluatie laat zien dat onze aanpak in specifieke gevallen een veelbelovende uitbreiding is op bestaande mechanismen.