WETENSCHAPPEN &
BIO-INGENIEURSWETENSCHAPPEN

De Onderzoeksgroep
Software Languages Lab

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

Ruben Opdebeeck

ter behaling van de graad van Doctor in de wetenschappen

Titel van het proefschrift:

Static Analysis for Quality Assurance of Ansible Infrastructure as Code

Promotor:
Prof. dr. Coen De Roover

De verdediging heeft plaats op

Vrijdag 25 oktober 2024 om 17:00u

Campus Etterbeek van de Vrije Universiteit Brussel, Pleinlaan 2, Elsene in auditorium I.2.01.

Samenstelling van de jury

Prof. dr. Viviane Jonckers (VUB, voorzitter)
Prof. dr. Antonio Paolillo (VUB, secretaris)
Prof. dr. Kris Steenhaut (VUB)
Prof. dr. Beat Signer (VUB)
Prof. dr. Dimitris Mitropoulos (National and Kapodistrian University of Athens, GR)
Prof. dr. João F. Ferreira (Universidade de Lisboa, PT)

Curriculum vitae

Ruben Opdebeeck behaalde in 2019 de graad van MSc in Computerwetenschappen aan de VUB. Daarna startte hij een doctoraat aan het Software Languages Lab (SOFT), gefinancierd door een SB-beurs van het Fonds voor Wetenschappelijk Onderzoek – Vlaanderen.
Ruben's onderzoek resulteerde in acht publicaties in internationale peer-reviewed tijdschriften en conferenties (7 als eerste auteur), alsook een hoofdstuk in een boek als mede-auteur. Daarnaast presenteerde hij ook zijn onderzoek in twee industriële tentoonstellingen en het wetenschappelijke "Dagstuhl" seminarie. In 2023 spendeerde hij 3 maanden bij de MCIS-onderzoeksgroep van Queen's University, Canada, waar hij onderzoek uitvoerde onder begeleiding van Prof. Dr. Bram Adams. Naast zijn onderzoek heeft hij ook tien bachelorstudenten en acht masterstudenten begeleid tijdens hun thesis.

Abstract van het doctoraatsonderzoek

De huidige software-intensieve industrie gebruikt steeds complexere digitale computerinfrastructuren, mogelijk bestaande uit honderden clouddiensten zoals virtuele computerinstanties en beheerde databases. Het handmatig beheren van dergelijke infrastructuren is arbeidsintensief en foutgevoelig. Daarom stelt Infrastructure as Code (IaC) beoefenaars in staat om de voorziening, configuratie en orkestratie van infrastructuren te automatiseren met uitvoerbare code. De correctheid en veiligheid van dergelijke code is cruciaal, omdat defecten rampzalige onderbrekingen kunnen veroorzaken, terwijl beveiligingszwaktes de infrastructuur kwetsbaar maken voor cyberaanvallen. Bestaande kwaliteitsborgingsanalyses voor IaC zijn echter gebrekkig en presteren ofwel ondermaats door niet te redeneren over het gedrag van de code, of zijn zeer middelenintensief om toe te passen. Bovendien richten ze zich alleen op de infrastructuurcode en negeren ze de ondersteunende softwaretoevoerketens, die tegenwoordig een belangrijke vector voor cyberaanvallen vormen.

Deze thesis tracht deze problemen aan te kaarten met vier bijdragen gericht op Ansible, een van de meest gebruikte IaC tools. De eerste is de Program Dependence Graph (PDG) voor Ansible, een graafvoorstelling van infrastructuurcode die diens gedrag omvat. We introduceren een analyse om een PDG voor Ansible-scripts statisch te extraheren, d.w.z. zonder de code uit te voeren, en overbruggen hierdoor de kloof tussen eenvoudige maar ondermaatse aanpakken en nauwkeurige maar middelenintensieve aanpakken.

De tweede bijdrage is een toepassing van de PDG-voorstelling om semantische "codegeuren" te detecteren die kunnen duiden op defecten in infrastructuurcode. Door patronen in de grafen te detecteren, kan de analyse zes foutgevoelige codeerpraktijken gerelateerd aan Ansible-variabelen blootleggen. We passen deze analyse toe in een grootschalig empirisch onderzoek om de prevalentie en levensduur van deze geuren in open-source Ansible-code te onderzoeken.

GASEL, de derde bijdrage, is een toepassing van de PDG-voorstelling om beveiligingszwaktes in IaC te detecteren. Het gebruikt graafqueries op de PDG om zeven soorten "beveiligingsgeuren" te identificeren die kunnen leiden tot beveiligingszwaktes of aanvalsvectoren. De gedragsinformatie omvat in de grafen stelt GASEL in staat om beter te presteren dan bestaande analyses. We passen GASEL toe in een grootschalig empirisch onderzoek naar beveiligingsgeuren in open-source Ansible-code.

De vierde bijdrage is een empirische studie van de softwaretoeleveringsketen van Ansible-software. We stellen een geautomatiseerde analyse voor die afhankelijkheden van derden van Ansible-plugins identificeert en passen deze toe om de soorten afhankelijkheden te bestuderen die in de praktijk voorkomen.

De empirische studies die in deze thesis worden gepresenteerd, vragen om verbeteringen in de kwaliteitsborgingspraktijken van Infrastructure as Code, terwijl geavanceerde tooling hiervoor wordt mogelijk gemaakt door de voorgestelde statische analyses.