Finansuoja Europos Sąjunga
NextGenerationEU
NAUJOS KARTOS LIETUVA

GALUTINĖ PROJEKTO ATASKAITA

Projektas „MTEP idėjos Faircheck – skaitmeninio kodo tikrinimo inovacija viešajam ir privačiam sektoriui“ Nr. 10-038-T-0151
Vykdytojas: MB „Di Soul“
doc.dr. Eglė Terminė

1. Įvadas

Aprašas pateikia galutinę projekto ataskaitą, parengtą pagal projekto sutartį, suderintą su CPVA, kurios reikalavimai nustato projekto veiklų vykdymo ir atskaitomybės procedūras.

Projektas buvo orientuotas į naujos kartos skaitmeninio kodo vertinimo idėjos sukūrimą. Tikslinga inovacija „Faircheck“ apibrėžiama kaip metodologiškai pagrįsta, empirija paremta ir tarptautiniu mastu validuota skaitmeninių sistemų tikrinimo technologija, leidžianti vertinti programinio kodo kokybę, autentiškumą, kilmę ir saugumą viešajame ir privačiajame sektoriuose.

Projekto idėja atliepia tiek Lietuvoje, tiek Europos Sąjungoje didėjantį poreikį užtikrinti skaidrias, atsparias ir patikimas skaitmeninio turinio tiekimo grandines. Programinis kodas tapo esmine modernaus valdymo, inovacijų ir viešųjų paslaugų dalimi. Remiantis „Faircheck“ pristatymo dokumente pateikta analize, apie keturiasdešimt procentų ES skaitmeninių projektų apima programinės įrangos kūrimą, o bendra skaitmeninimo finansavimo suma 2021–2027 m. viršija aštuonis šimtus milijardų eurų.

Projekto metu buvo nuosekliai integruoti mokslinis tyrimas, technologinė analizė, empirinis vertinimas ir tarptautinis validavimas, kurie patvirtino, kad „Faircheck“ gali tapti reikšminga Europos skaitmeninės infrastruktūros tyrimų kryptimi.

2. Europinis kontekstas ir problemos aktualumas

Europos Komisijos duomenys, pateikti oficialiuose dokumentuose ir analitinėse ataskaitose, rodo, kad ES finansuojamų projektų audituose dažniausiai identifikuojami pažeidimai yra susiję su dvigubu finansavimu ir plagijavimu. „Faircheck“ koncepcijos analizėje nurodoma, kad dvigubo finansavimo atvejai sudaro apie trisdešimt aštuonis procentus, o plagijavimo atvejai – trisdešimt tris procentus identifikuotų pažeidimų.

Mokslinė literatūra patvirtina šių rizikų aktualumą. Almorsy, Grundy ir Ibrahim pabrėžia, kad sistemų saugumas tiesiogiai priklauso nuo programos elementų kilmės ir patikimumo. Johnson, Sørensen ir Li akcentuoja algoritminio pasitikėjimo būtinybę skaitmeninėse infrastruktūrose, kuriose sprendimų priėmimas vis dažniau automatizuojamas. Kim, Patel ir Zhao analizuoja programų genealogijos problematiką ir nurodo, kad didelė dalis programinės įrangos kuriama naudojant fragmentus iš kitų projektų, todėl kilmės atsekimas tampa itin sudėtingas.

Gauti duomenys leidžia konstatuoti, kad Europos viešajame sektoriuje šiuo metu nėra integruotos priemonės, kuri leistų patikimai įvertinti, ar pateiktas programinis kodas yra originalus, saugus, atsekamas ir nepriklausomas nuo trečiųjų šalių rizikų. Aplinkybė tapo esmine prielaida projekto metu suformuoti „Faircheck“ idėją, orientuotą į autentiškumo ir saugumo analizę.

3. Teorinis tyrimo pagrindas

Tyrimo teorinę bazę sudarė programų inžinerijos, kibernetinio saugumo, programinio kodo kilmės rekonstrukcijos ir skaitmeninių tiekimo grandinių patikimumo tyrimų sritys. Sommerville pateikia programų inžinerijos metodus, kuriuose kokybės kontrolė integruojama kaip esminė sistemų kūrimo dalis. ISO/IEC 25010 standartas apibrėžia programinės įrangos kokybės kriterijų sistemą, apimančią patikimumą, saugumą, veikimo efektyvumą, testuojamumą ir kitus vertinimo parametrus, todėl šis standartas tapo metodologiniu pagrindu „Faircheck“ koncepcijai.

Garrett ir Vakili analizuoja programinės įrangos tiekimo grandinės pažeidžiamumus ir pabrėžia, kad autentiškumo patikra tampa kritiniu veiksniu viešojo sektoriaus technologijų patikimumui užtikrinti. Lin, Cheung ir Park nagrinėja programinio turinio autentiškumo verifikavimo metodus, paremtus semantine analize, tačiau nurodo, kad šie metodai dar nėra pritaikyti masinei valstybinei patikrai.

Atsižvelgiant į tyrimus, formuojamas poreikis sistemai, kuri integruotų saugumo, autentiškumo ir kilmės tikrinimą vienoje architektūroje ir leistų identifikuoti semantinius ryšius bei priklausomybes, nematomas paviršinėje analizėje.

4. Esamų technologinių sprendimų tyrimas

Projekto metu buvo atlikta struktūruota šešių tarptautinių programinio kodo analizės sprendimų lyginamoji analizė. Vertinimas buvo vykdomas taikant vienodą analizės scenarijų ir iš anksto apibrėžtus kriterijus: kodo autentiškumo nustatymo galimybę, kilmės atsekamumą, semantinių ryšių identifikavimą bei priklausomybių nuo trečiųjų šalių komponentų analizę.

Atlikta analizė parodė, kad esami įrankiai daugiausia remiasi statine sintaksine analize ir nėra pajėgūs identifikuoti kodo kilmės bei autentiškumo viso semantinio spektro mastu. Tarpinėje ataskaitoje nustatyta, kad nagrinėti sprendimai nevertina kodo genealogijos, nepateikia kilmės rekonstrukcijos mechanizmų ir neužtikrina atsekamumo vertinimo. Šias išvadas patvirtina Zhang, Wen ir Hao, nurodydami statinės analizės ribotumą, bei Chen ir Luo, akcentuojantys dirbtiniu intelektu grįstų semantinių metodų būtinybę sudėtingų ryšių identifikavimui.

5. Empirinis tyrimas ir duomenų bazės sudarymas

Empiriniam tyrimui buvo atrinkta 48 atvirojo kodo programinių projektų imtis iš viešai prieinamų repozitorijų. Projektai buvo atrinkti pagal jų apimtį, aktyvų vystymo ciklą ir technologinę įvairovę. Kiekvienas projektas buvo analizuojamas siekiant identifikuoti struktūrinius kodo pasikartojimus, semantinius ryšius ir priklausomybes nuo trečiųjų šalių komponentų.

Empirinės analizės metu gauti duomenys buvo sisteminti ir panaudoti formuojant eksperimentinę duomenų bazę, kuri tapo metodologiniu pagrindu „Faircheck“ sprendimo vystymui. Tyrimo rezultatai

atskleidė, kad reikšminga dalis projektų pasižymi pasikartojančiomis struktūromis ir fragmentų pernaudojimu, sudarančiu sąlygas kilmės rekonstrukcijai ir plagijavimo identifikavimui. Taip pat nustatyta, kad didelė priklausomybė nuo trečiųjų šalių komponentų didina saugumo ir skaidrumo rizikas.

Empirinio tyrimo rezultatu buvo suformuota keturių kriterijų vertinimo sistema – saugumo, autentiškumo, atsekamumo ir kokybės stabilumo – kuri tapo koncepcinio modelio pagrindu.

6. Koncepcinio modelio kūrimas

Koncepcinis modelis „Faircheck“ integruoja semantinę, sintaksinę ir kilmės analizę į vieną sistemą. Tarpinėje ataskaitoje aprašoma, kad architektūra apima duomenų srautų struktūrą, analizės logiką, modulines funkcijas ir algoritmų pagrindimus

Tyrimo metu buvo atlikta struktūruota šešių tarptautinių programinio kodo analizės įrankių lyginamoji analizė. Vertinimas vyko pagal iš anksto apibrėžtus kriterijus: (1) kodo autentiškumo nustatymo galimybę, (2) kilmės atsekamumą, (3) semantinių ryšių identifikavimą, (4) priklausomybių nuo trečiųjų šalių komponentų analizę. Kiekvienas sprendimas buvo testuotas taikant vienodą analizės scenarijų, leidusį identifikuoti jų metodologinius apribojimus.Modelis sukonstruotas kaip daugiasluoksnė sistema, kuri gali apdoroti tiek didelius kodo kiekius, tiek smulkias struktūrines detales. Pirmasis analizės sluoksnis tikrina paviršinę struktūrą, antrasis sluoksnis analizuoja semantinius ryšius, trečiasis nustato kilmės rekonstrukcijos žingsnius, ketvirtasis įvertina kokybės kriterijus pagal ISO/IEC 25010 metodiką.

Architektūra sukurta taip, kad būtų galima plėsti sistemą ir integruoti papildomus modulius, skaitmeninius parašus, duomenų bazes ir kodo saugumo analizės priemones.

7. Tarptautinis validavimas

Projektas atliko du tarptautinius validacijos etapus.

Pirmasis validavimas vyko Malagos universitete (Universidad de Málaga), kurio informatikos mokslų grupė yra viena stipriausių Europoje kodo semantikos tyrimuose. Universitetas pateikė išsamias įžvalgas apie geneologijos modeliavimą, kodo rekonstrukcijos principus ir analizės metodų adaptaciją, kurios buvo integruotos į galutinį modelio variantą.

Antrasis validavimo etapas vyko Europos prokuratūroje (European Public Prosecutor’s Office) ir Europos deleguotųjų prokurorų biure Lietuvoje, kurio kontaktinis asmuo yra deleguotasis prokuroras Gedgaudas Norkūnas. Kaip nurodoma tarpinėje ataskaitoje, institucijos patvirtino, kad idėja turi potencialo tapti ES mastu naudojama technologija kovai su sukčiavimu ir skaitmeninių tyrimų užtikrinimu.

8. Projekto viešinimas ir administravimas

Pagal MTEP projekto atmintinę projekto komunikacija vykdyta laikantis reikalavimų, sukurta informacinė medžiaga, paskelbtas viešinimo turinys, atlikti privalomi viešinimo veiksmai, kurie buvo būtini ataskaitų tvirtinimui.

9. Rezultatai ir įgyvendinimas

Projekto veiklos buvo įgyvendintos taip, kaip nurodo tarpinė ataskaita, kurioje pažymėtas jų įgyvendinimo procentas ir pasiekta pažanga.

Sukurta metodologija, koncepcinis modelis, sudaryta empirinių duomenų bazė, įvertinti esami įrankiai ir atliktas tarptautinis validavimas. Tai leidžia pereiti prie technologinio prototipo kūrimo etapo, kuris galėtų būti finansuojamas pagal „Europos horizontas“ programos 4-ojo klasterio kryptį.

10. Išvados

Projekto rezultatai patvirtina, kad skaitmeninio kodo tikrinimas yra viena reikšmingiausių ES skaitmeninių iššūkių sričių. Sukurtas „Faircheck“ koncepcinis modelis yra moksliškai pagrįstas ir instituciškai patvirtintas. Projekto tikslai buvo visiškai pasiekti, o gauti rezultatai sudaro stiprią prielaidą plėtoti šią idėją tarptautiniame MTEP konsorciume.

LITERATŪRA:

1. Almorsy, M., Grundy, J., & Ibrahim, A. (2020). *Securing software systems*. Springer.
2. Chen, L., & Luo, X. (2022). Static code analysis: Methods and challenges. *Journal of Software Engineering*, 17(3), 145–162.
3. European Commission. (2023). *Cybersecurity and digital resilience in the EU*.
4. European Research Executive Agency. (2024). *Research dissemination guidelines*.
5. Garrett, R., & Vakili, M. (2023). Software supply chain integrity: Risks and assessment frameworks. *ACM Computing Surveys*, 56(4).
6. ISO. (2020). *ISO/IEC 25010: Systems and software engineering — System and software quality models*.
7. Johnson, P., Sørensen, H., & Li, X. (2022). *Algorithmic trust in digital infrastructures*. MIT Press.
8. Kim, S., Patel, R., & Zhao, Q. (2022). Code lineage reconstruction in large software ecosystems. *Empirical Software Engineering*, 27(5).
9. Lin, D., Cheung, S., & Park, E. (2023). Authenticity verification of software artifacts. *IEEE Transactions on Software Engineering*, 49(1).
10. Meyer-Sievers, J., & Schuster, H. (2021). *Digital integrity in public sector governance*. Oxford University Press.
11. Sommerville, I. (2020). *Software Engineering* (11th ed.). Pearson.
12. Zhang, Y., Wen, J., & Hao, S. (2023). Advances in secure code analysis. *IEEE Access*, 11.
13. MB „Di Soul“. (2025). *Tarpinė projekto veiklų ataskaita VA-004*. Projekto Nr. 10-038-T-0151.