AGENTÚRA NA PODPORU VÝSKUMU A VÝVOJA

Agentúra na podporu výskumu a vývoja
Mýtna 23, P.O.BOX 839 04, 839 04 Bratislava 32

Formulár ZK

Záverečná karta projektu

Názov projektu

Ontologická reprezentácia pre bezpečnosť informačných systémov

Evidenčné číslo projektu **APVV-19-0220**

Zodpovedný riešiteľ **prof. Ing. Pavol Zajac, PhD.**

Príjemca **Slovenská technická univerzita v Bratislave - Fakulta elektrotechniky a informatiky**

Názov pracoviska, na ktorom bol projekt riešený

Projekt bol riešený na 4 pracoviskách:
1. Fakulta elektrotechniky a informatiky, Slovenská technická univerzita v Bratislave, Ilkovičova 3, 841 04 Bratislava
2. Fakulta matematiky, fyziky a informatiky, Univerzity Komenského, Mlynská dolina F1, 842 48 Bratislava
3. Ústav informatiky SAV, v. v. i., Dúbravská cesta 9, 845 07 Bratislava
4. Matematický ústav SAV, v. v. i., Štefánikova 49, 814 73 Bratislava

Názov a štát zahraničného pracoviska, ktoré spolupracovalo pri riešení

Na projekte neboli priami zahraniční partneri.

Udelené patenty/podané patentové prihlášky, vynálezy alebo úžitkové vzory, ktoré sú výsledkami projektu

Projekt nemá ako výstup patenty, vynálezy alebo úžitkové vzory.

Najvýznamnejšie publikácie (knihy, články, prednášky, správy a pod.) zhrňujúce výsledky projektu – uveďte aj publikácie prijaté do tlače

Balogh, S; Gallo, O; Ploszek, R; Spacek, P; Zajac, P: IoT Security Challenges: Cloud and Blockchain, Postquantum Cryptography, and Evolutionary Techniques, ELECTRONICS 10/21, 2021, DOI: 10.3390/electronics10212647
55 citácií (GS), z toho 21 citácií WoS
Svec, Peter, Stefan Balogh, and Martin Homola. "Experimental Evaluation of Description Logic Concept Learning Algorithms for Static Malware Detection." ICISSP. 2021.
9 citácií (GS), z toho 2 citácie WoS
Švec, P., Balogh, Š., Homola, M., Kľuka, J., & Bisták, T. (2024). Semantic Data Representation for Explainable Windows Malware Detection Models. arXiv preprint arXiv:2403.11669.
(zaslané do recenzného konania)
Svátek, V., Zamazal, O., Nguyen, V.B., Ivánek, J., Kľuka, J. and Vacura, M., 2023. Focused categorization power of ontologies: General framework and study on simple existential concept expressions. Semantic Web, (Preprint), pp.1-45.
Uhliarik, I., 2022. Enhancing and Evaluating the Product Fuzzy DPLL Solver. SN Computer Science, 3(5), p.388.

Uplatnenie výsledkov projektu

Hlavným praktickým výsledkom je komplexný ontologický model umožňujúci zachytiť dáta o bezpečnostných hrozbách. Na základe modelu sme vytvorili komprehenzívnu ontológiu PE Malware ontology, ktorá dodáva sémantickú prezentačnú vrstvu pre dáta pochádzajúce zo statickej analýzy malvéru, vychádzajúcu zo v súčasnosti najpopulárnejších malvérových datasetov EMBER a SOREL. Spolu s ontológiou sme publikovali aj viaceré predpripravené dátové množiny odvodené od datasetu EMBER ako aj SOREL, tieto sú dostupné v repozitári https://github.com/orbis-security. Okrem toho sme realizovali viaceré ďalšie dátové množiny a ontológie (podrobnejší popis viď ZS).

Vytvorené dátové množiny a ontológie umožňujú ako výskumné tak aj praktické využitie v oblasti strojového učenia a inteligentného odhaľovania malvér a súvisiacich bezpečnostných hrozieb.

Výsledky výskumu je možné využiť pri rozšírenej malvérovej analýze a klasifikácii, pre lepšie porozumenie bezpečnostných hrozieb, pre budúci vývoj bezpečnostných nástrojov, ako aj pre lepšiu spoluprácu bezpečnostných expertov a efektívnejšie zdieľanie informácií o bezpečnostných hrozbách a incidentoch.

Súhrn výsledkov riešenia projektu a naplnenia cieľov projektu v slovenskom jazyku (max. 20 riadkov)

Ako popisujeme v ZS, podarilo sa nám naplniť stanovené ciele projektu. Vybudovali sme ontologickú reprezentačnú vrstvu pre viaceré dátové zdroje v oblasti informačnej bezpečnosti s dôrazom na malvér. Navrhli sme architektúru detekčného systému a preskúmali sme viaceré teoretické modely detekčných mechanizmov na báze informačných tokov a tzv. attack-trees. Venovali sme sa tiež kryptografických vlastnostiam využiteľným v detekčných systémoch. Intenzívne sme skúmali rôzne aspekty inferenčných mechanizmov, ktoré môžu spolu s ontologickou reprezentáciou byť využité na inferenciu, napr. za účelom hľadania vysvetlení, alebo objavovania a aplikácie detekčných pravidiel.

Konkrétnym výsledkom bol napr. algoritmus pre fuzzy DPLL inferenciu, alebo implementácia abdukčného solvera pre ontológie, skúmanie metamodelovania ontológií, či publikované zlepšenia nástroja DL learner. Tieto techniky sme aplikovali nad ontologicky reprezentovanými dátami z datasetov EMBER a SOREL. Ukázali sme, že nástroje konceptového učenia sú vhodné na vysvetliteľnú klasifikáciu malvérových vzoriek. Ich praktickému využitiu nateraz bránia vyššie výpočtové nároky. Tým bola daná aj nižšia miera generalizácie a presnosti.

Výsledky sme porovnali s inými metódami, napr. s metódou učenia sa rozhodovacích stromov. Siahli sme tiež po hybridných neuro-symbolových nástrojoch. Mimoriadne silné výsledky sme dosiahli metódou LEN. Dosiahnuté vysvetlenia boli síce mierne nižšej kvality ako pri konceptovom učení, ale dokázali sme spracovať celý dataset EMBER. V tejto časti práce sa nám podarilo tiež vylepšiť algoritmus, ktorý generuje vysvetelnia z LEN (tzv. tailored LEN).

Súhrn výsledkov riešenia projektu a naplnenia cieľov projektu v anglickom jazyku (max. 20 riadkov)

As we describe in the ToR, we have succeeded in meeting the stated objectives of the project. We have built an ontology representation layer for multiple information security data sources with an emphasis on malware. We designed the architecture of the detection system and explored several theoretical models of detection mechanisms based on information flows and attack-trees. We also addressed cryptographic properties usable in detection systems. We have intensively investigated different aspects of inference mechanisms that can be used together with ontological representations for inference, e.g., in order to search for explanations or to discover and apply detection rules.

Concrete results were e.g. an algorithm for fuzzy DPLL inference, or the implementation of an abduction solver for ontologies, the investigation of ontology metamodelling, or published improvements to the DL learner tool. We applied these techniques over ontology-represented data from the EMBER and SOREL datasets. We have shown that concept learning tools are suitable for the explainable classification of malware samples. Their practical use is currently hampered by higher computational requirements. This has also given lower generalization and accuracy rates.

We compared the results with other methods, e.g., decision tree learning. We also reached for hybrid neuro-symbolic tools. We obtained particularly strong results with the LEN method. Although the achieved explanations were of slightly lower quality than the concept learning approach, we were able to process the entire EMBER dataset. In this part of the work, we were also able to improve the algorithm that generates explanations from LEN (called tailored LEN).