جامعة محمد الخامس بالرباط
Université Mohammed V de Rabat

École Nationale Supérieure d'Informatique et d'Analyse des Systèmes
Centre d'Études Doctorales en Sciences des Technologies de l'Information et de l'Ingénieur

AVIS DE SOUTENANCE DE THÈSE DE DOCTORAT

Monsieur Oualid ZAAZAA

Soutiendra publiquement sa thèse de Doctorat en Informatique

Spécialité : Informatique

Le Samedi 18 Janvier 2024 à 12h00 au Grand Amphi à l'ENSIAS

Intitulé de la thèse

A Machine Learning Based Framework for Enhancing Security In Blockchain Smart Contracts through Automated Vulnerability Detection

Président :

Pr. Faissal EL BOUANANI, PES, ENSIAS, Université Mohammed V de Rabat

Directeur de thèse :

Pr. Hanan EL BAKKALI, PES, ENSIAS, Université Mohammed V de Rabat

Rapporteurs :

Pr. Mostafa AZIZI, PES, EST, Université Mohammed I, Oujda

Pr. Yassine MALEH, MCH, ENSA-Khouribga, Université Sultan Moulay Slimane, Beni Mellal

Pr. Afaf OUADDAH, MCH, Institut National des Postes et Télécommunications, Rabat

Examinateur :

Pr. Khalid ZINE-DINE, PES, FSR, Université Mohammed V de Rabat

Résumé: Dans un domaine en pleine évolution comme celui de la Blockchain et des contrats intelligents, la détection et la correction des vulnérabilités sont essentielles pour garantir l'intégrité et la fiabilité des systèmes basés sur cette technologie. Cette thèse explore plusieurs aspects critiques de la sécurité dans les contrats intelligents, en se concentrant sur l'identification des vulnérabilités peu explorées et l'automatisation de leur détection. L'un des principaux objectifs était de développer un Framework basé sur l'apprentissage automatique pour détecter les vulnérabilités, en relevant le défi des vulnérabilités répandues mais encore insuffisamment étudiées dans la littérature. De plus, l'étude visait à réduire la dépendance aux connaissances d'experts et à accélérer la création de règles de détection pour les vulnérabilités nouvellement découvertes.

Notre enquête de la littérature a révélé plusieurs résultats clés. Tout d'abord, nous avons identifié trois vulnérabilités dans les contrats intelligents qui n'avaient pas été étudiées en profondeur auparavant dans la littérature. Ces vulnérabilités représentent des lacunes importantes dans la compréhension actuelle et soulignent la nécessité d'une enquête plus approfondie sur leurs implications et leurs stratégies de mitigation.

Deuxièmement, nous avons observé que les conventions de dénomination des vulnérabilités varient considérablement selon les articles de recherche. Cette incohérence peut conduire à une redondance et à une inefficacité dans le processus de recherche, car différentes études peuvent faire référence à la même vulnérabilité en utilisant des terminologies différentes. Pour résoudre ce problème, nous avons développé un système de codification visant à normaliser la communication des vulnérabilités entre les chercheurs, réduisant ainsi la redondance et améliorant l'efficacité des futurs efforts de recherche.

En plus du système de codification, nous avons construit plusieurs Datasets pouvant être utilisées par les chercheurs pour tester et valider leurs modèles et Framework par rapport aux vulnérabilités nouvellement identifiées.

Une contribution significative de cette recherche est le développement d'un Framework modulaire capable de détecter les vulnérabilités dès les premières étapes du processus de développement des contrats intelligents. Ce Framework s'appuie sur des techniques d'apprentissage automatique pour automatiser le processus de génération des règles, qui nécessitait traditionnellement des connaissances approfondies et un effort manuel. En intégrant les Grands Modèles de Langage (LLM), nous avons considérablement accéléré le processus d'intégration de nouvelles règles de détection et d'adaptation aux vulnérabilités émergentes.

En identifiant et en traitant les vulnérabilités dès le début du processus de développement, notre Framework améliore la sécurité globale des contrats intelligents, améliorant ainsi la fiabilité des systèmes basés sur la Blockchain.

Mots-clés: Détection de vulnérabilité, LLM, Machine Learning, Sécurité Blockchain, Sécurité logicielle, Smart Contract.

Abstract: In the rapidly evolving field of smart contract security, the discovery and mitigation of vulnerabilities are crucial for ensuring the integrity and reliability of Blockchain-based systems. This thesis explores several critical aspects of smart contract vulnerabilities, focusing on the identification of under-researched issues and the automation of their detection. A primary aim was to develop a machine learning-based framework for detecting vulnerabilities, addressing the challenge of prevalent but insufficiently studied vulnerabilities in current research. Additionally, the study sought to reduce dependence on expert knowledge and expedite the creation of detection rules for newly discovered vulnerabilities.

Our investigation of the literature revealed several key findings. First, we identified three vulnerabilities in smart contracts that had not been previously researched in the literature. These vulnerabilities represent significant gaps in current understanding and highlight the need for further investigation into their implications and mitigation strategies. Second, we observed that the naming conventions for vulnerabilities vary widely across different research papers. This inconsistency can lead to redundancy and inefficiency in the research process, as different studies may refer to the same vulnerability using different terminologies. To address this issue, we developed a codification system aimed at standardizing the communication of vulnerabilities among researchers, thereby reducing redundancy and improving the efficiency of future research efforts. In addition to the codification system, we constructed multiple datasets that can be utilized by researchers to test and validate their models and frameworks against the newly identified vulnerabilities.

A significant contribution of this research is the development of a modular framework that can detect vulnerabilities in the early stages of smart contract development. This framework leverages machine learning techniques and Large Language Models (LLMs) to automate the rule generation process, which traditionally required extensive expert knowledge and manual effort, while adapting to emerging vulnerabilities.

By identifying and addressing vulnerabilities early in the development process, our framework enhances the overall security of smart contracts, thereby improving the reliability of Blockchain-based systems.

Keywords: Blockchain Security, Dynamic analysis, Large Language Models, Machine Learning, Smart Contract, Software Security, Static analysis, Vulnerability Detection.