الجمهورية الشعبية الديمقراطية الجزائرية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
المدرسة العليا للإعلام الآلي 08٠ ماي 5491٠ بسيدي بلعباس
École Supérieure en Informatique
-08 Mai 1945- Sidi Bel Abbès

THESIS

To obtain the diploma of **Master's Degree**
Field: **Computer Science**
Specialty: **Intelligence Artificielle et Science de Données (IASD)**

Theme

Enhancing Software Security Using Transformer-Based Language Models

---

Presented by:
Abdechakour MECHRI

Submission Date: **Sept, 2024**
In front of the jury composed of:

Dr. DIF Nassima	President
Dr. BOUSMAHA Rabab	Examiner
Dr. KHALDI Belkacem	Supervisor
Dr. FERRAG Mohamed Amine	Co-Supervisor

Academic Year : 2023/2024

Abstract

As software becomes increasingly integral to various industries, the need for robust software security has become paramount. Ensuring the protection of software from attacks and cyber threats throughout its lifecycle—from development to deployment and maintenance—requires innovative approaches. This thesis explores the application of Transformer-Based Language Models to enhance software security. We begin by reviewing the current state of software security and machine learning, then delves into various approaches, evaluating the efficacy of these models in identifying and analyzing vulnerabilities, and comparing their performance to RNN-based models. The results demonstrate that Transformer-Based Language Models hold significant promise in vulnerability detection, offering a powerful tool to advance the field of software security. Through rigorous evaluation and comparison, we establish these models as superior in their capability to identify and address software vulnerabilities, paving the way for more secure software development practices.

Keywords— Transformer-Based Language Models, Large Language Model, Generative Pre-trained Transformers, Static Analysis, Vulnerability Detection, Cyber threats, Software lifecycle, Software Security

الملخص

مع تزايد أهمية البرمجيات في مختلف الصناعات، أصبحت الحاجة إلى أمن قوي للبرمجيات أمرًا بالغ الأهمية. يتطلب ضمان حماية البرمجيات من الهجمات والتهديدات السيبرانية طوال دورة حياتها - من التطوير إلى النشر والصيانة - نهجًا مبتكرة. تستكشف هذه الأطروحة تطبيق نماذج اللغة المعتمدة على المحولات لتعزيز أمن البرمجيات. نبدأ بمراجعة الوضع الحالي لأمن البرمجيات والتعلم الآلي، ثم نتعمق في مختلف النهج، مقيمين فعالية هذه النماذج في تحديد وتحليل نقاط الضعف، ومقارنة أدائها بنماذج الشبكات العصبية المتكررة (RNN). تظهر النتائج أن نماذج اللغة المعتمدة على المحولات تحمل وعودًا كبيرة في الكشف عن نقاط الضعف، مما يوفر أداة قوية لتطوير مجال أمن البرمجيات. من خلال التقييم الدقيق والمقارنة، نثبت تفوق هذه النماذج في قدرتها على تحديد ومعالجة نقاط الضعف في البرمجيات، مما يمهد الطريق لممارسات تطوير برمجيات أكثر أمانًا.

الكلمات المفتاحية--- التحليل الثابت، اكتشاف الثغرات، التهديدات السيبرانية، نموذج اللغة الكبيرة، أمان البرمجيات، الأمن، المحولات التوليدية المدربة مسبقًا.

Résumé

À mesure que les logiciels deviennent de plus en plus essentiels dans divers secteurs, le besoin d'une sécurité logicielle robuste est devenu primordial. Assurer la protection des logiciels contre les attaques et les menaces cybernétiques tout au long de leur cycle de vie - du développement au déploiement et à la maintenance - nécessite des approches innovantes. Cette thèse explore l'application des modèles de langage basés sur les Transformers pour améliorer la sécurité des logiciels. Nous commençons par examiner l'état actuel de la sécurité des logiciels et de l'apprentissage automatique, puis nous nous penchons sur diverses approches, évaluant l'efficacité de ces modèles dans l'identification et l'analyse des vulnérabilités, et comparant leurs performances à celles des modèles basés sur les RNN. Les résultats démontrent que les modèles de langage basés sur les Transformers sont très prometteurs pour la détection des vulnérabilités, offrant un outil puissant pour faire progresser le domaine de la sécurité des logiciels. Grâce à une évaluation et une comparaison rigoureuses, nous établissons la supériorité de ces modèles dans leur capacité à identifier et à traiter les vulnérabilités logicielles, ouvrant la voie à des pratiques de développement de logiciels plus sécurisées.

Keywords— Modèles de langage basés sur les Transformers, Grand modèle de langage, Generative Pre-trained Transformers, Analyse statique, Détection de vulnérabilités, Menaces cybernétiques, Cycle de vie des logiciels, Sécurité logicielle