Masterarbeit „Programmverstehen von Access Control-Mechanismen in komplexer Open Source Software auf Basis von großen Sprachmodellen“

Software Security ist ein immer wichtiger werdendes Teilgebiet der Informationssicherheit. Die Relevanz ist u.a. durch die zunehmende Digitalisierung, insbesondere durch die vermehrte Einführung von mobilen und IoT-Anwendungen, begründet. Sicherheitslücken in Software können zu entsprechenden Risiken führen, was letztendlich das Vertrauen untergräbt.

Es gibt unterschiedliche Ansätze, die Security von Software-Systemen zu erhöhen, wie z.B. Fuzzing, SAST (Static Security Application Testing) oder die architekturelle Risikoanalyse (Threat Modeling). Ein oft vernachlässigter Bereich ist das (manuelle) **Programmverstehen von Quelltext bzgl. der Security**, da es sich um einen sehr aufwendigen Vorgang handelt, der zudem viel Expertise verlangt. Auf der anderen Seite wird immer mehr Open Source Software (OSS) verwendet, auch in proprietären Systemen. Ein besseres Verständnis von oft komplexer OSS zu erhalten ist daher ein wichtiger Schritt zur Verbesserung der Sicherheit oder zur Stärkung des Vertrauens solcher implementierten Sicherheitsmechanismen, insbesondere da der Quelltext offen verfügbar ist und sonst diese Quelle nicht entsprechend für Security Audits genutzt wird.

Ziel der Masterarbeit ist es zu untersuchen, inwieweit **große Sprachmodelle** (LLMs) hier unterstützen können. LLMs bieten den Vorteil, dass sie z.B. unabhängig von der Programmiersprache agieren und ggf. sogar Konfigurationsdateien mit einbeziehen können. Des Weiteren haben LLMs unter Beweis gestellt, dass sie auch komplexeren Programmcode erklären können. Auf der anderen Seite liefern sie in einigen Fällen fehlerhafte Ergebnisse („Halluzinieren“) und können daher zu Missverständnissen führen.

Im Rahmen der Arbeit wird eine verallgemeinerbare Methodik entwickelt, wie LLMs zum Verstehen der Implementierung von **Access Control-Mechanismen** in unterschiedlichen OSS-Projekten eingesetzt werden können. Geplant ist hier eine Untersuchung von AndroidOS (Systemdienste, in Java und in C/C++ implementiert; Android Binder; SELinux-Konfigurationen) und von Kubernetes.

Insbesondere werden Chancen und auch Grenzen eines solchen LLM-gestützten Programmverstehens ermittelt. Wenn bei dem LLM-basierten Ansatz Probleme auftreten sollten, wird untersucht, inwieweit Agentic AI ergänzend zur Verbesserung der Ergebnisse eingesetzt werden kann.

Es ist zudem eine Publikation auf einer wissenschaftlichen Security-Konferenz geplant, wenn die Ergebnisse interessant genug sind. Das Projekt wird in Zusammenarbeit mit der School of Computing and Augmented Intelligence an der Arizona State University (Prof. Dr. Gail-Joon Ahn) durchgeführt.

Voraussetzungen:

1. Informationssicherheit (z.B. ISEC oder andere Vorlesungen aus diesem Bereich)
2. Kenntnisse in der Softwaretechnik und von LLMs sind wünschenswert, aber nicht zwingend erforderlich
3. Eigenmotivation, an einem forschungsnahen Thema zu arbeiten

Kontakt:

PD Dr. Karsten Sohr, TZI der Universität Bremen

E-Mail: sohr@tzi.de

Tel.: 218 63922