Masterarbeit „Verständliches Threat Modeling von Software mit großen Sprachmodellen“

Mit der zunehmenden Digitalisierung unserer beruflichen und privaten Welt wird die Frage nach der Sicherheit von Software (im Sinne der Software Security) immer relevanter. Software findet man nicht nur in PCs oder Servern, sondern in so unterschiedlichen Geräten wie Mobile Devices, Automobiltechnik, Smarthome-Geräten, Medizintechnik, intelligentem Spielzeug oder Robotern, für die alle unterschiedliche Sicherheitsanforderungen bestehen. Neue gesetzliche Regelungen wie der Cyber Resilience Act (CRA) der EU fordern u.a. von Herstellern von Produkten mit digitalen Komponenten die Umsetzung von Software-Sicherheitsmaßnahmen nach dem Stand der Technik – dies beinhaltet insbesondere den Nachweis eines effektiven Security Development Lifecycle (SDL). Ein essentieller Baustein eines SDL ist die **architekturelle Risikoanalyse** (Threat Modeling), bei der Bedrohungen und geeignete Sicherheitsmaßnahmen auf Ebene der Software-Architektur modelliert werden. Gleichzeitig lassen sich aus den architekturellen Sicherheitsmaßnahmen Testfälle für das spätere Sicherheitstesten ableiten, so dass auch andere Schritte eines SDL unterstützt werden.

Es ist jedoch bekannt, dass die architekturelle Risikoanalyse schwierig ist und somit ein entsprechendes Expertenwissen erfordert, das aber oft nicht zur Verfügung steht. Um den Zugang zur architekturellen Risikoanalyse zu vereinfachen, wird eine Plattform entwickelt, mit welcher architekturelle Bedrohungen und daraus abgeleitete Sicherheitsanforderungen *semi-automatisch* mit Hilfe von großen Sprachmodellen (Large Language Models, LLMs) erstellt und anschließend mittels **Feedbackschleife** durch Sicherheitsanalysten oder Software-Architekten geeignet angepasst werden können. Dieses Zusatzwissen kann durch Expertenkenntnisse (wenn vorhanden), Informationen über geplante Technologien, Ergebnisse statischer Programmanalysen (Extraktion von Informationen aus existierendem Quellcode), CMDB-Informationen sowie Einträge aus Software Bill of Materials (SBOMs) gewonnen werden. Außerdem bietet die Feedbackschleife den Nutzern die Möglichkeit, Halluzinationen zu erkennen.

Die Architektur selbst wird dabei in Form von Datenflussdiagrammen (DFDs) visualisiert; DFDs werden in der STRIDE-Methodik (Threat Modeling) verwendet, eine verbreitete Methode für das Threat Modeling von Microsoft. In der Masterarbeit wird untersucht, welche der zu korrelierenden Informationen dem LLM in Form von Prompts und welche mittels Retrieval Augmented Generation/Fine Tuning übergeben werden. Beim Prompt-Engineering besteht eine Herausforderung darin, dass die Prompts für die Erzeugung von Software-Architektur, Bedrohungen und möglichen Sicherheitsmaßnahmen (Sicherheitsanforderungen auf Architekturebene) aus Benutzbarkeitsgründen möglichst automatisch erstellt werden. Es wird ein **nutzerzentrierter Entwicklungsansatz** angestrebt; Benutzbarkeitsaspekte sind daher in diesem Projekt ein wesentlicher Aspekt.

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 in 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