CAREER: SaTC: Privacy-Compliant Software Development for Software Supply Chain

Overview:

Much software tailors user experiences based on personal data, but improper handling of the data violates privacy laws and leads to serious legal consequences. Hence, privacy compliance is now critical in software development. Software developers achieve privacy compliance with various measures, such as always providing clear privacy notices and controls for personal data handling to users. Applying these measures is feasible for software containing code entirely built from scratch, but fundamentally difficult for real-world software where the use of third-party components is common. In particular, third-party components comprise over 80% of the software code base, yet their data handling practices are unclear to developers and can vary depending on how they are used in different software, which presents challenges in addressing those practices using the above measures. Today's software developers tackle the challenges with a variety of ad-hoc, non-mechanical approaches, such as building privacy labels in a way decoupled from the actual code. The outcomes are one-size-fits-all privacy measures that do not account for the varied uses of components, only covering those components that software relies on superficially.

True software privacy compliance should adapt to varied uses of all components in the entire software supply chain, a goal that is not possible to attain with existing software development practices. This proposal presents a comprehensive research and education plan aimed at transforming the way software developers create privacy-compliant software. The key idea is to mechanically connect privacy compliance measures to software (including component) code, so that inherent in the software development process, these measures can be reliably built, adapted and audited based on varying use of components. To ensure it has a solid foundation, the project will develop a formal and unified model to capture the fine-grained implications of software code on privacy compliance, and define basic operations on the model to represent the customized use of different components. To replace the ad-hoc and non-mechanical approaches used today, it will build an automation framework that builds the model up from software code, and automates the implementations of various privacy compliance measures for software supply chain. When the implementations are not feasible due to privacy compliance bugs (implementation flaws or even intentional malicious practices such as deception) in current software components, the project will design and implement new methods for detecting and mitigating the bugs.

Intellectual Merit:

The main contributions of this project are: (1) a formal model that captures the privacy compliance implications of diverse use of components in software, (2) an automation framework for software privacy compliance based on the model, and (3) new methods for detecting and mitigating privacy compliance bugs in software supply chain. All of the tasks will be implemented and validated on real-world software applications that feature the diverse and complex use of third-party components.

Broader Impacts:

This project, by providing foundational methods and developer tooling for privacy-compliant software development, is directly aligned with the National Privacy Research Strategy (NPRS) and the recent National Strategy to Advance Privacy-Preserving Data Sharing and Analytics. The PI will field software artifacts resulting from this research with collaborators at Meta and Microsoft, and all research results will be made freely available to advance interdisciplinary research in multiple communities such as privacy, security, software engineering, and formal methods. The PI will train two PhD students in software supply chain security and privacy automation and integrate this research into his courses on computer security & privacy, as well as K-12 student events like the Camp Connect Summer programs. UCF is a Hispanic-Serving Institute, and the PI has engaged multiple and will continue to engage students from underrepresented groups in UCF undergraduate programs and other programs such as CAHSI.

Keywords: Privacy, Applied; Software; Formal Methods and Language-based Security