Master Thesis Behavioral Authentication

"Domain-specific dataset repository"

Background

Passwords are used for securing computer systems for a long time. Although they are still many times the only protection, people (re-)use short and weak passwords for their own accounts that can easily be guessed or may be revealed by big identity leaks. Apart from possession-based mechanisms like hardware tokens or access cards, behavioral biometrics are usable to verify claimed identities of persons, too and could provide a higher security and usability. Typical examples are the verification by the way a person walks (gait), by the way a person types on a keyboard or touchscreen (keystroke/touchstroke) or by routines based on the browsed webpages, app interaction, or visited locations.

Problem

Typically, researchers conduct user experiments to evaluate their behavioral authentication system proposals. In these experiments they sample human behavior using, e.g., sensors in smartphones. This data collection requires much effort as it requires the design of scenarios with a minimal bias on the participant's behavior but also label as much interaction specifics as possible with a robust collection infrastructure. Furthermore, once an approach provides good results or only the different components of the included biometric systems (preprocessing, matching, and decision) are subject to optimization, there is no need for individual data collection and experiments.

In these cases, an easy way to integrate existing datasets for own evaluation purposes would be very helpful. Although some researchers publish their datasets and some repositories exist, a dedicated dataset repository and integration into typical pipelines for behavioral authentication systems is needed, but yet missing.

Goal

In this master thesis, you should develop and evaluate such a dataset repository:

- Extensive related work analysis which datasets and repositories exist
- Develop a unified behavioral authentication dataset domain model
- Implement a dataset repository that allows an easy integration of new datasets and an easy integration into machine learning pipelines
- Evaluation of the repository by integrating datasets from related work

Contact

Eric Klieme
eric.klieme@hpi.de
H1.18
0331 5509-559