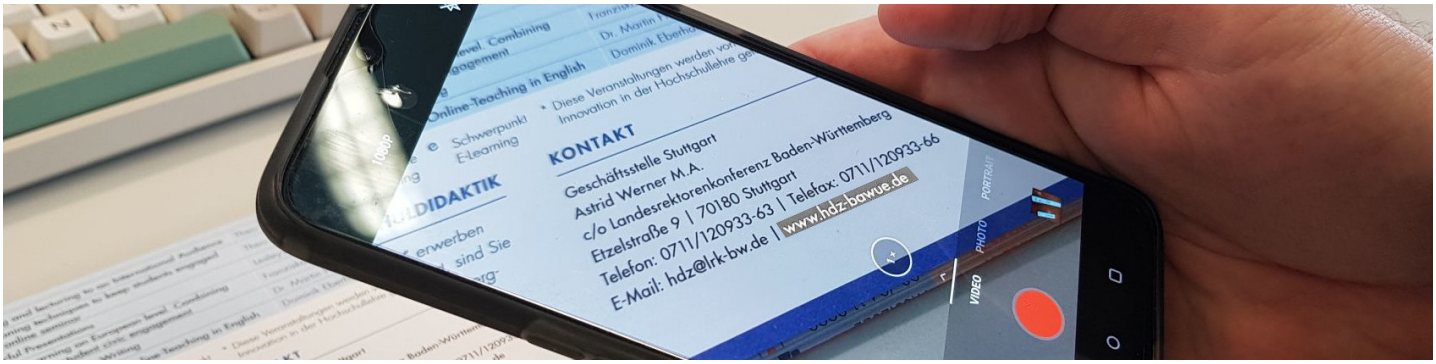
Karlsruhe Institute of Technology

Research Group Intelligent System Security (ISEC)
Institute of Information Security and Dependability
Karlsruhe Institute for Technology

Research Group Security · Usability · Society (SECUSO)
Institute of Applied Informatics and Formal Description Methods
Karlsruhe Institute for Technology

LiveText
Phishing Attacks based on LiveText

Bachelor's Thesis Proposal



Background

Optical Character Recognition (OCR) are tools that convert images of typed, handwritten or printed text in other formats. OCRs are quite common, nowadays, especially for blind and visually impaired users (e.g., text-to-speech tools). 'LiveText' is an OCR implemented in Apple iPhones and iPads mounting Apple A12 Bionic chips and successors, i.e., 2018+ iPhones and 2019+ iPads. On top of recognizing characters, LiveText highlights hyperlinks and enables the users to directly open websites by pointing the camera to printouts of the domain. As all OCRs, LiveText makes mistakes and misinterprets characters. This motivates a new real world attack vector, e.g., for phishing attacks, to take users to malicious versions of legitimate websites. However, this problem is difficult to investigate, due to the closed source nature of the LiveText system. Hence, in the first step, a surrogate model needs to be extracted. Next, the student runs adversarial attacks on it and verifies the transferability of the attacks on the original LiveText system. Finally, she develops a social engineering attack that takes advantage of the flaws and evaluates it in a user study.
The goal of this thesis is two-folds:

- Review literature on model extraction attacks and conduct a model extraction on Live Text to construct a surrogate model. Implement transferable attacks against this model. Verify the transferability of the attack on the original Live Text system quantitatively.
- Review literature on phishing attacks. Identify, motivate and evaluate social engineering attacks, scenarios and attacker models by taking advantage of the flaws in the Live Text recognition model (e.g., homographic attacks, typo-attacks, pixel-based attacks, etc.). Verify in a user study the scenarios' applicability, qualitatively and quantitatively.

Related work to start with

1. Chen and Xu, 2020, Attacking Optical Character Recognition (OCR) Systems with Adversarial Watermarks
2. Song et al., 2018, Fooling OCR Systems with Adversarial Text Images
3. Heise.de, 24.01.2022: Mehr Fehler mit QR-Codes, auch bei OnePlus - Google bietet Pixel-Update
4. Heise.de, 20.01.2022: Googles Kamera verfaelscht Links in QR-Codes
5. Tramer et al., 2016, Stealing Machine Learning Models via Prediction APIs
6. Orekondy et al., 2019, Knockoff Nets: Stealing Functionality of Black-Box Models
7. Steinmetz et al., 2021, Performing social engineering: A qualitative study of information security deceptions.
8. Thao et al., 2020, Human Factors in Homograph Attack Recognition.
9. Bhardwaj et al., 2021, Privacy-aware detection framework to mitigate new-age phishing attacks.
10. Quinkert et al., 2019, It's Not what It Looks Like: Measuring Attacks and Defensive Registrations of Homograph Domains.

Supervisors & Contact

- Maximilian Noppel, M.Sc.,, ISEC, maximilian.noppel@kit.edu
- Mattia Mossano, M.Sc. M.A., SECUSO, mattia.mossano@kit.edu