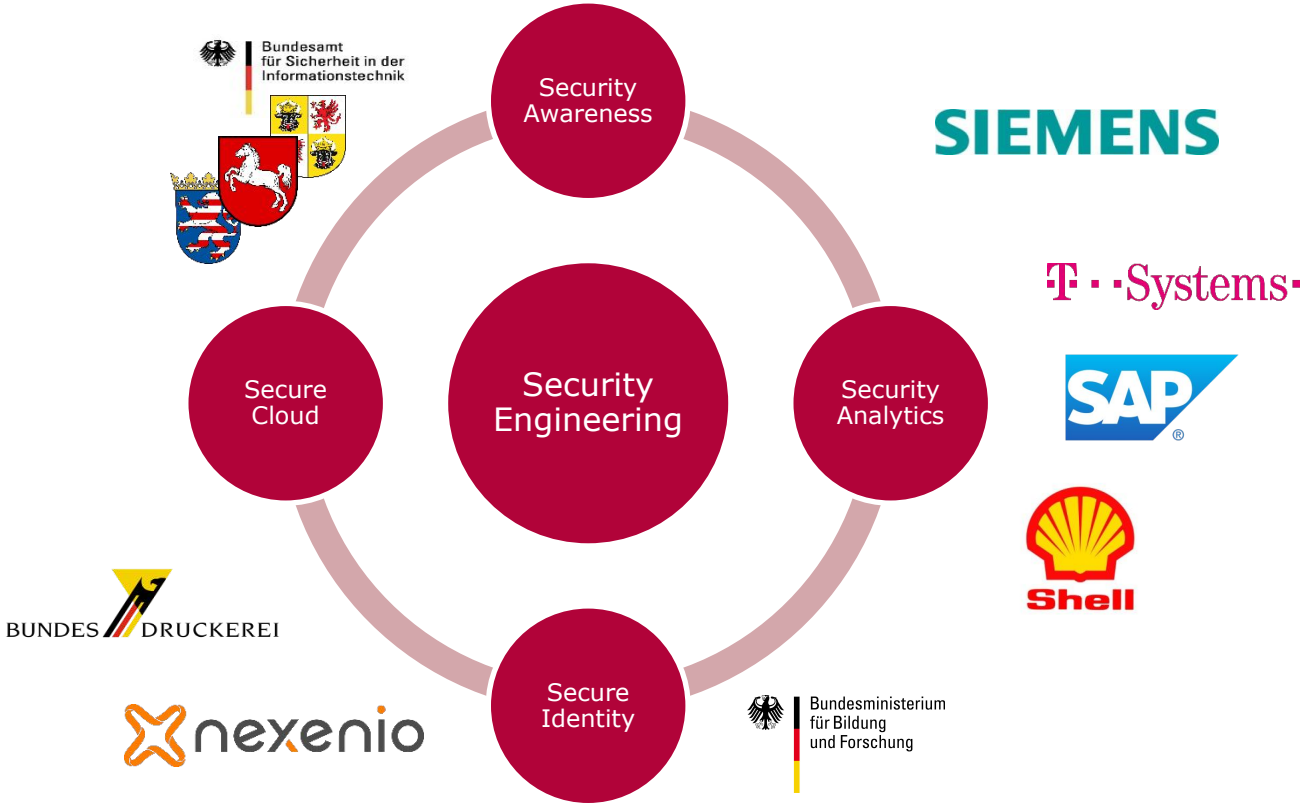
Secure Identity Lab Seminar

Summer Semester 2021

Eric Klieme, Alexander Mühle, Andreas Grüner, Daniel Köhler
Chair Internet Technologies and Systems

Summer Semester 2021

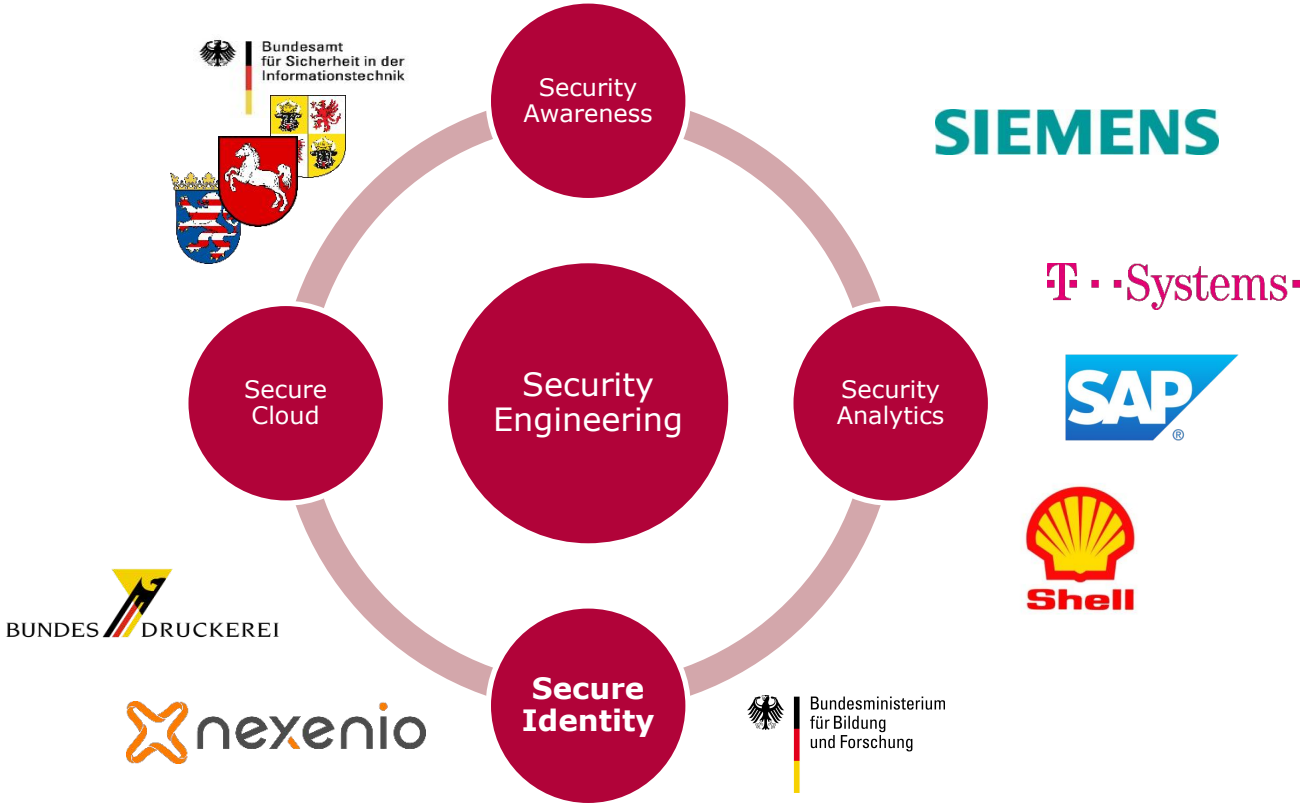
HPI Security Engineering



Secure Identity Lab Seminar

Eric Klieme
Alexander Mühle
Andreas Grüner
Daniel Köhler

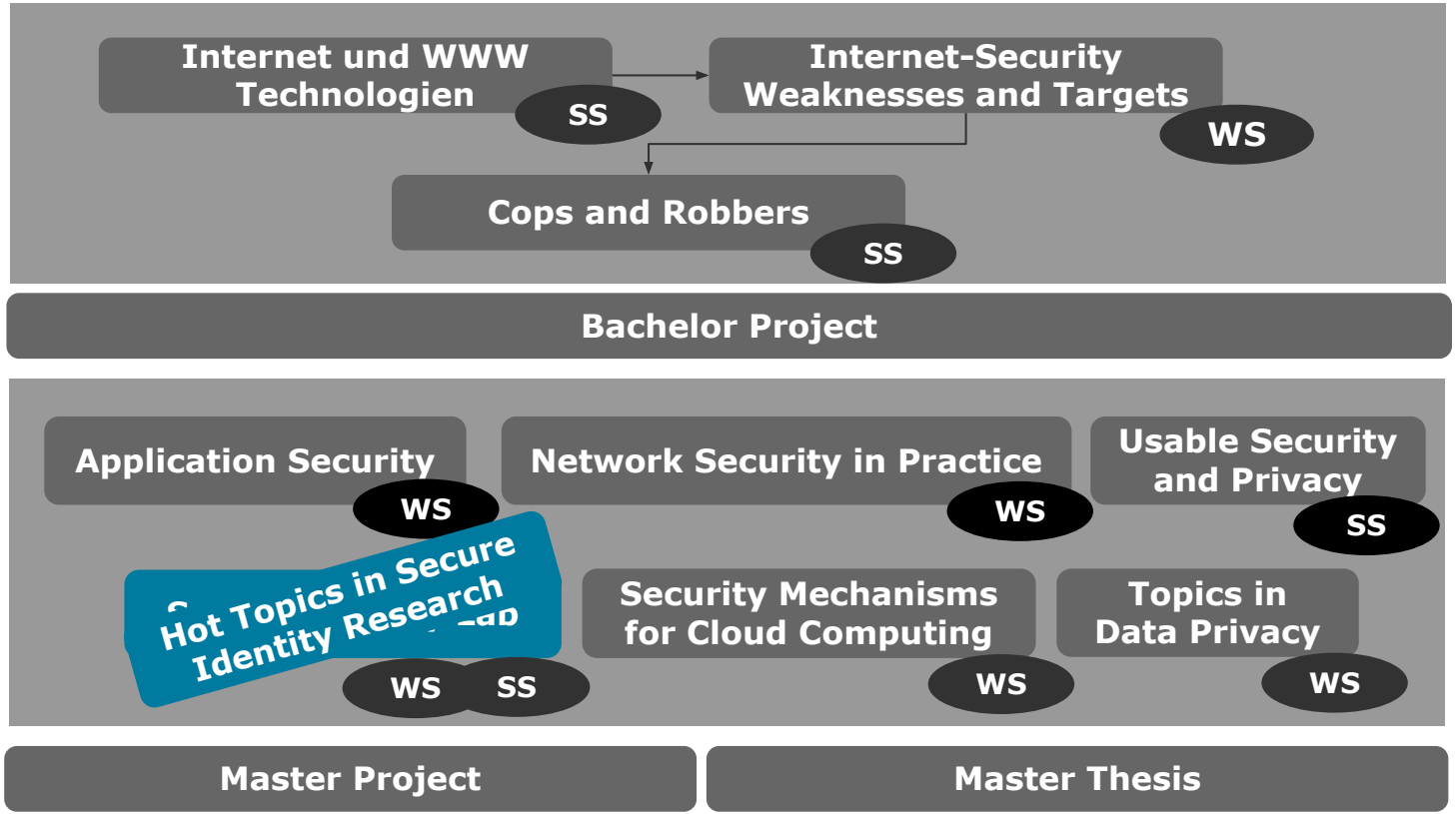
HPI Security Engineering



Secure Identity Lab Seminar

Eric Klieme
Alexander Mühle
Andreas Grüner
Daniel Köhler

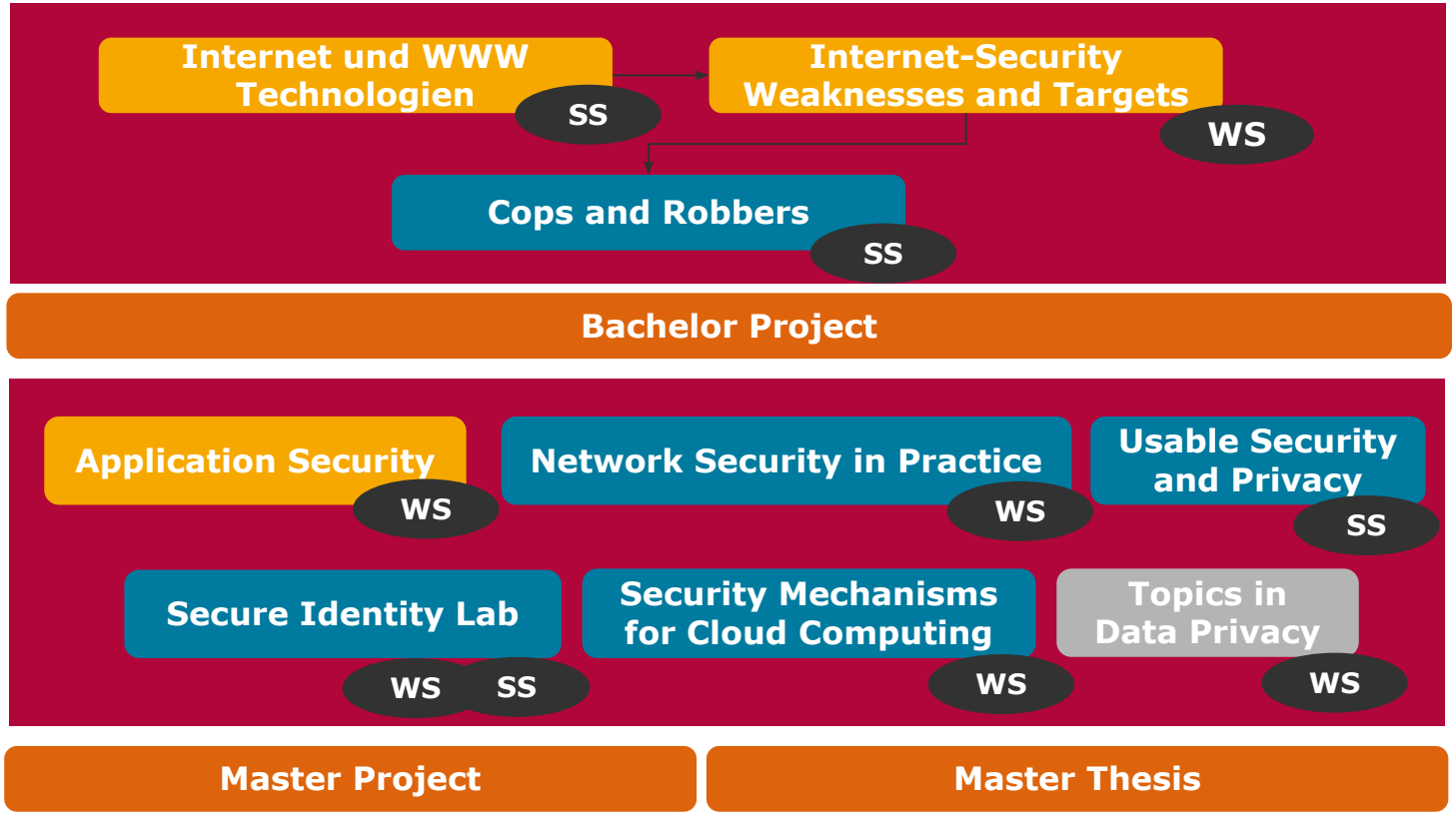
Study Plan



Secure Identity Lab Seminar

Eric Klieme
Alexander Mühle
Andreas Grüner
Daniel Köhler

Study Plan



Lecture

Seminar

Block Event

Secure Identity Lab Seminar

Eric Klieme
Alexander Mühle
Andreas Grüner
Daniel Köhler

Topics - Summary

- In this seminar we will focus on three fields of secure identity research
  - Authenticating the users through behavioural aspects (Eric)
  - Analyzing P2P networks (Alexander)
  - Self-Sovereign Identity management (Andreas, Daniel, Alexander)

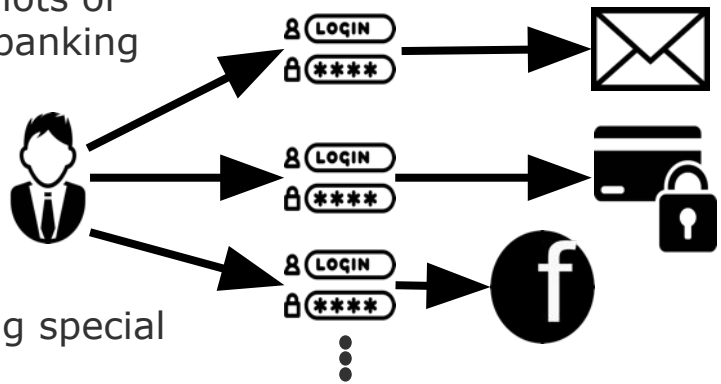
Secure Identity Lab
Behavioral Authentication

Eric Klieme
eric.klieme@hpi.de

Traditional Username/Password authentication may not be the perfect solution for today's internet service usage

Problem

- A user has 80++ passwords on average and uses lots of different services in a range from social media to banking applications



Solution (in theory)

- Different password for every service
- Each password of a certain length, maybe including special letters
- Only remembered, not written down anywhere

Solution (assumed)

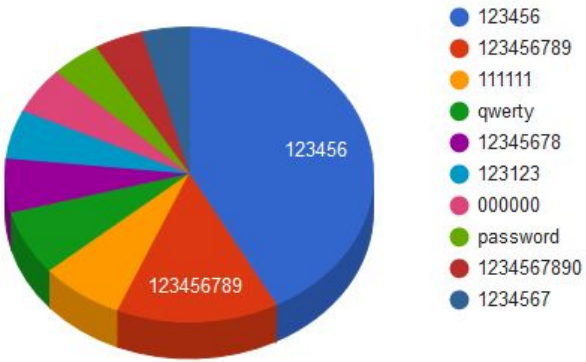
- Complex passwords hard to remember, use a much simpler
- Same passwords for different services

Topic Presentation

Eric Klieme

HPI Identity Leak Checker confirms the assumed real world situation

- Service to check if identity has leaked based on freely accessible sources of leakages
- Currently database of ~ 12 billion user accounts
- Main findings:
  - Very simple passwords used
  - A lot of services either apply no hashing at all or just weak approaches (~60%)



Distribution of top 10 leaked passwords

- https://sec.hpi.de/ilc/

Topic Presentation

Eric Klieme

Typically, we have three ways to authenticate people

1. Something you know

e.g., passwords

Advantages

- Easy implementation / Low cost
- Easy to change
- Widely used

Disadvantages

- Theft / Guessing possible
- Modern service landscape lead to many simple and reused passwords
- Once authenticated, further authentication rarely implemented

2. Something you possess

e.g., tokens, access cards

Advantages

- Needs no remembering
- „Show and access"

Disadvantages

- Loss/Theft
- Once authenticated, further authentication rarely implemented
- Sometimes additional hardware required

3. Something you are: Biometrics include physical and **behavioral** aspects but collects very private data

Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Physical Aspects

Behavioral Aspects

Advantages

- No objects needed
- No knowledge needed
- Unique for every person by default

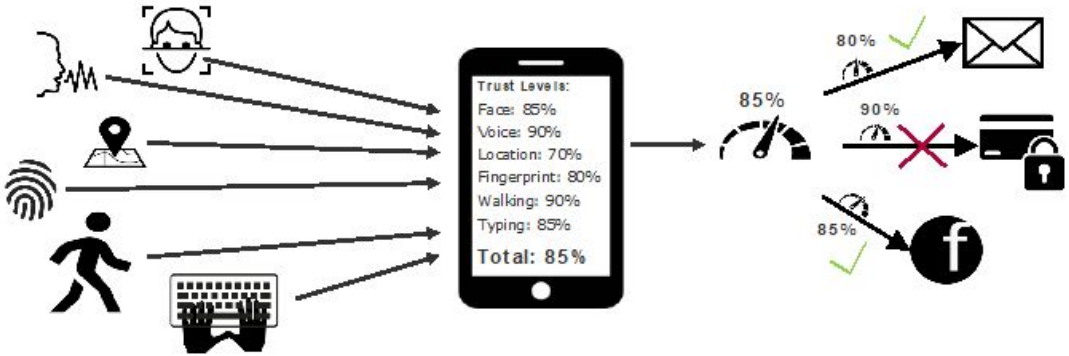
Disadvantages

- Once compromised/imitated, features are hard to change
- Once authenticated, further authentication rarely implemented
- Location or typing contains sensible information
- Only **probability** of „rightful" user, exact match difficult

Idea: Analyze user's behavior **continuously** in the background based on **behavioral biometrics**

- Sample behavior via sensors, e.g. accelerometer, microphone, APIs
  - Devices that are "worn", e.g., smartphone, smartwatches
  - Devices that are explicitly used, e.g., notebooks, smartphones
  - Devices that are physically interacted with, e.g., **doorhandles?**

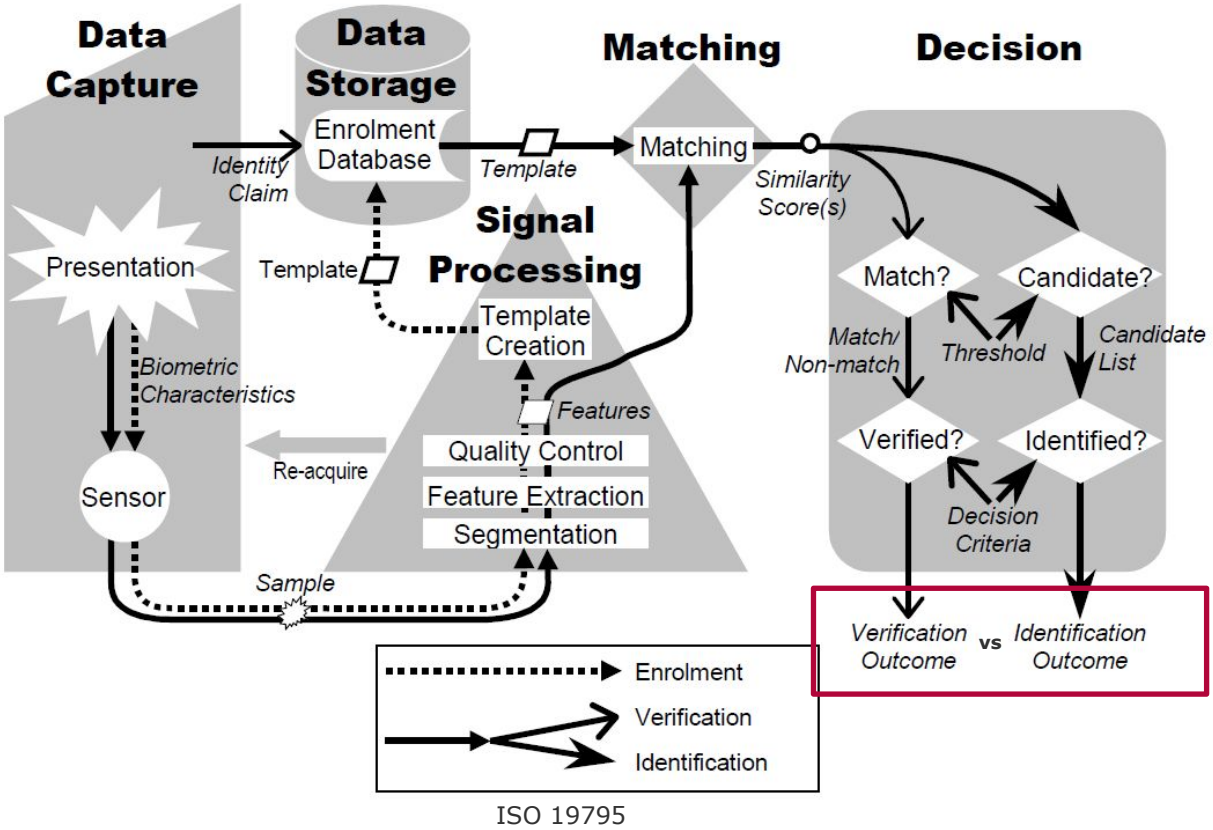
- Use biometric system to create templates of users based on behaviour and later use these templates for identification or verification
- **One Vision:** Aggregate all results for **continuous authentication**



Topic Presentation

Eric Klieme

Behavioral Authentication Systems usually include **Biometric Systems with Machine Learning**



ISO 19795

Topic Presentation
Eric Klieme

Examples of Behavioral Authentication Systems
Maturity Level: *Research Proposals*

Behavioral Characteristic	*Data Capture*	*Signal Processing*	*Matching*	*Decision*
Keystroke	Keystroke API, Keycode, Time, Duration, (Pressure)	Features: Flight time, Dwell time, further related to n-grams	SVM, DL, Random Forest, GMM,...	Thresholds, Majority Voting
Gait	Accelerometer, Gyroscope	Step detection / Windowing, FFT features: min, max, std, var, energy, power...	SVM, DL, Random Forest, GMM,...	Thresholds, Majority Voting
Location Routine	GPS API	Features: Time spent at location, #visits (per day/week/..),...	SVM, DL, Random Forest, GMM,...	Thresholds, Majority Voting

Data Storage often not specifically mentioned in research proposals (some dumps/jsons/yaf somewhere..)

Behavior-Based Authentication Research: **Status Quo & New Challenges**

Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Behavioral Authentication Research at HPI

	Behavior-based Authentication Systems	Behavior Algorithms	Experiments
Challenges & Research Questions	What is required to use behavior-based authentication systems as alternatives to password-based?	Which algorithms are suitable for the different areas of behavior to verify identities?	How can I effectively collect data in experiments and how can I verify my own approaches in the real world?
Projects	**Modeling Behavioral Authentication Systems and Evaluations:** A unified understanding and domain model of all aspects of behavioral authentication systems is required for automation and simplification of research and deployment efforts	**Robust gait-based user verification:** Smartphones are not only in your pocket when you walk, other scenarios such as reading and phone calls are also of interest. **User verification through typing sounds:** Use the smartphone to recognize the user while he's typing on the device or next to it **Smart door handles:** Use door handles that sense touch and acceleration to identify users	**Large scale data collection on smartphones:** Integrate users in labeling process in the wild and let them annotate data even further **Techniques for less supervision and more realistic behavior during experiments:** Reduce supervisor interaction with questionnaire-like experiments

Topics
Behavioral Authentication

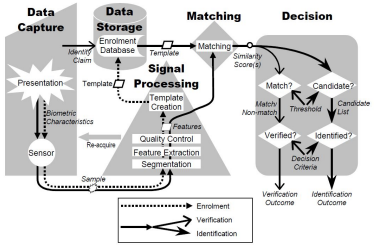
Eric Klieme
eric.klieme@hpi.de

Topic 1
Behavioral Authentication Implementation Platform

- Motivation: The core components of any behavioral system are similar
  - Data Capturing, Storage, Signal Processing, Matching, Decision
- Problem:
  - Any approach usually implements pipeline from scratch although different frameworks exist and biometric system is “formalized”
    - high effort, error prone, reproducibility difficult
- Idea:
  - Analyze existing frameworks to come up with domain specific (model-driven) implementation platform for different purpose
    - For **Evaluation** => **Algorithm Improvement** (python, R, Matlab...)
    - For **Deployment** => **Real-World Check** (android, ios, cloud container…)
    - For **Benchmarking** => **Comparison** (processing complexity, runtime, memory consumption…)

Behavioral Authentication Implementation Platform Vision

Behavioral Authentication System Proposals



(research paper, scripts..)

Frameworks

Support

Transformation

Your System

Support

Other datasets

Goal: Evaluation Deployment Benchmark

With your system: Easy, repeatable, shareable..

Topic 1
Behavioral Authentication Implementation Platform

- Contribution: „Proof-Of-Concept“
  - Search and Analyze a lot of related technologies for biometric systems / machine learning and already proposed frameworks on different platforms
  - Come up with an implementation platform design
  - Prototype platform and evaluate it based on real approaches from the related work for evaluation and real-world setups

- Nice-To-Have Skills
  - Python, Android, Java, Machine Learning (Frameworks)...
  - Strong focus on systemization

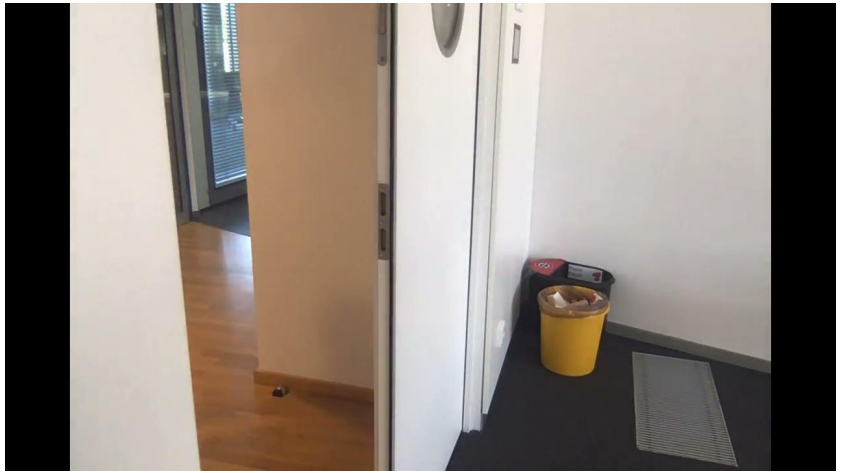
Topic 2
Door handle authentication

- **Idea:** Let users be authenticated based on the way they use a door handle
  - How they "touch" a handle (Resistive / Capacitive Touch)
  - How they "interact" with it (Acceleration)
  - How their hand looks / move like? (Computer Vision)
  - How they approach it (Bluetooth Signal, Ultrasonic)

- **Project:** Build a prototype, collect data, answer whether it is possible
  - Evaluate, choose and integrate sensors with door handle
  - Come up with a plan for a (large-scale) data collection user study, e.g. different offices, kitchens, meeting rooms etc.
    - Although Covid-19: some WiMis are in the Office, students available at Wohnheim ;)
  - Finally apply ML in a specific scenario
    - Add further focus, e.g. on deployability => real-time possible?

You will not start from scratch!
Prototype Stage 1 - “Proof-of-concept”

door handle with four touch sensors (top, bottom, front, back)



Main messages:

- Touch data seems very helpful and the overall idea realistic
- Identification is possible with > 80% taking the full interaction an no further feature engineering

You will not start from scratch!
Prototype Stage 2 - “long-term data collection”

door handle with touch sensor and accelerometer

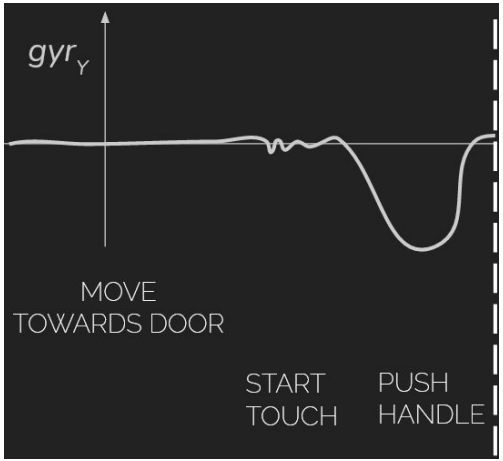
wrist bands for proximity tracking for automatic labeling of people opening the door

data collection infrastructure with dashboard

Main messages:

- Unsupervised collection requires many supervision ;)
- (HPI) Doors are not opened that often, collection takes a lot of time
- Diverse “special” ways of opening a door, proximity tracking works well
- ⇒ lessons learned, optimization required!

You will not start from scratch!
Prototype Stage 3 - “Real-world check”



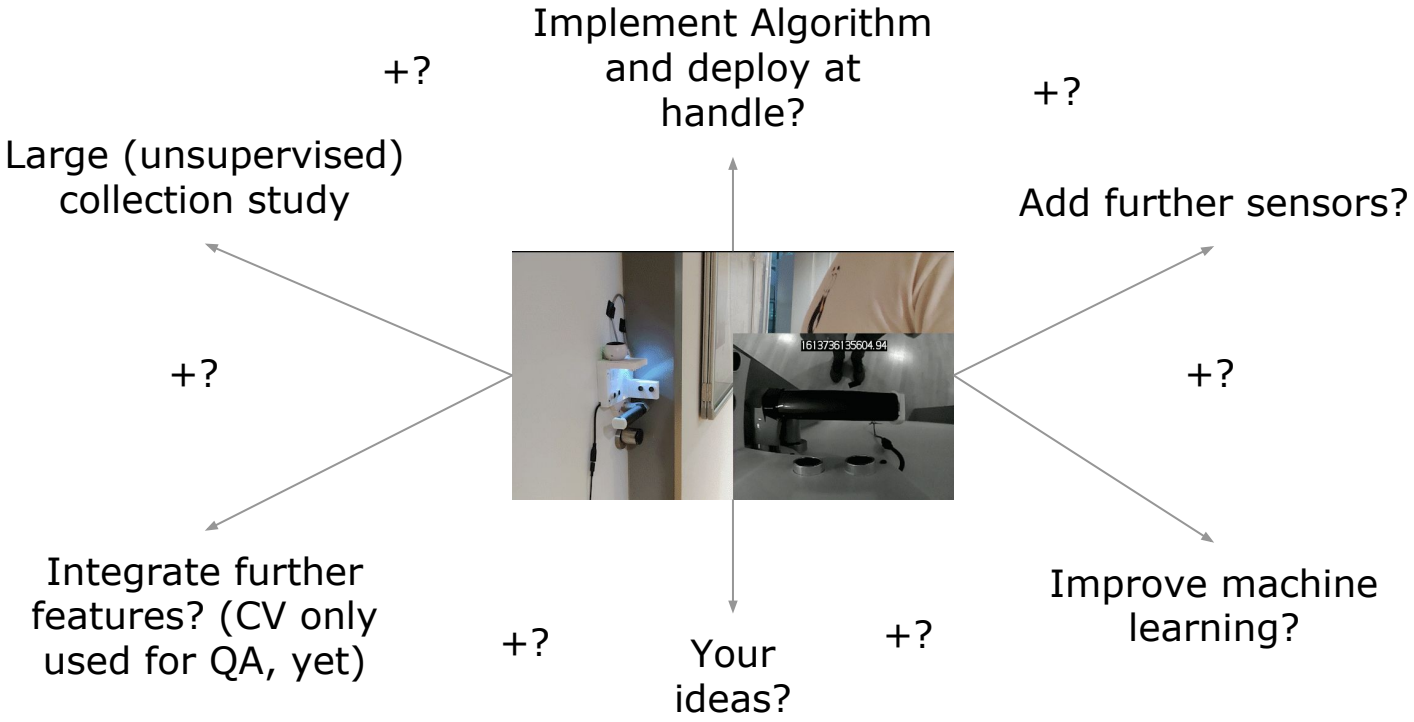
a real system needs to authenticate after door handle is pushed down

door handle with touch sensor, accelerometer, gyroscope, ultra sonic sensors, camera

Main messages:

- Only Push down phase of door handle seems sufficient for identification
- Gyro, Acc and Ultrasonic help to detect things, but touch seems still most important for identification

Topic 2: **Experiments / Algorithms**
Prototype Stage 4 - “This Seminar”



Topic 2:
Door handle authentication

- Contribution: „Proof-Of-Concept / User Study“
  - **Goal:** Verifiy User identities, Identify users from an office
  - **Data:** Touch data, accelerometer data, proximity data, CV + X?
  - **Your team’s contribution**
    - Come up with a nice processing of the data using machine learning
    - Study design and data collection for large scale collection
    - Improve data collection Infrastructure

- Nice-To-Have Skills
  - Python, Machine Learning, Raspberry Pi & Friends (3D Printing?)
  - Strong communication skills, creativity
  - Interest in conducting studies

Bring Your Own Ideas

Do you have other interesting ideas on what to do in the field of behavioral authentication in general?

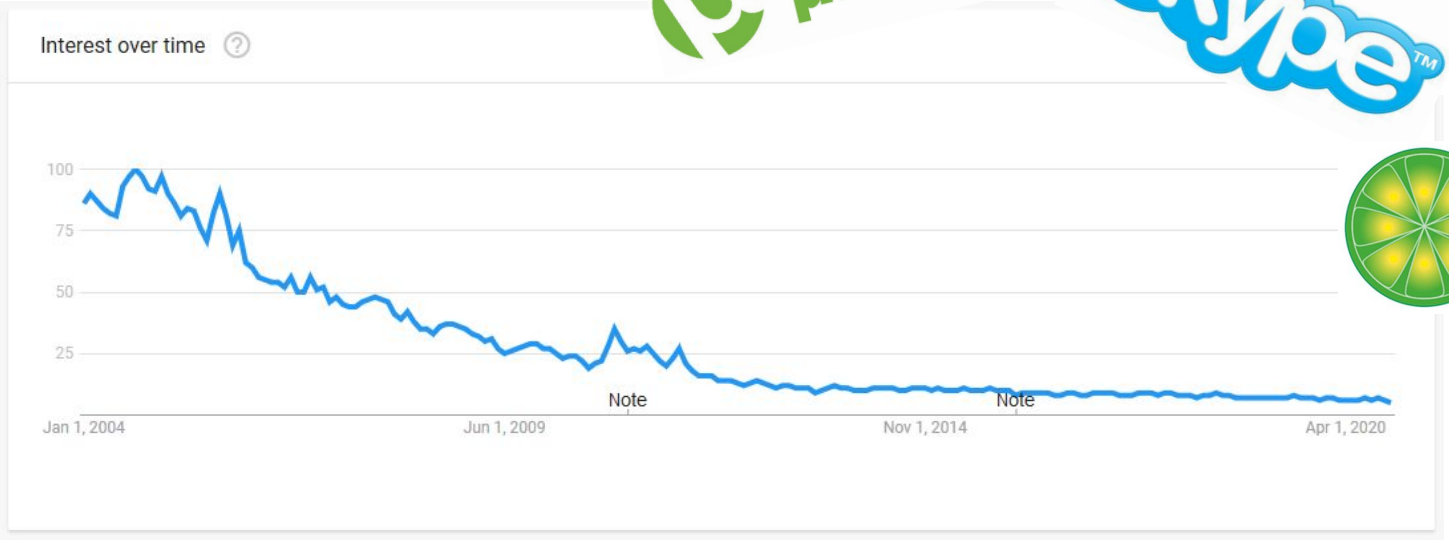
The door handle started as a student's idea as well ;)

Secure Identity Lab
P2P Networks

Alexander Mühle
alexander.muehle@hpi.de

P2P Networks

- Peer-to-Peer… who cares?



µTorrent®

skype™

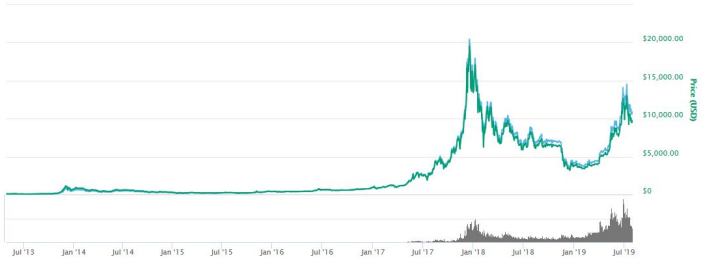
LimeWire

Topic Presentation
Alexander Mühle

Cryptocurrencies

- Bitcoin
  - Digital Cash ⇒ Pseudonyms only
  - Gained broad public awareness in 2017 through speculation
  - Drug trade, money laundering and cybercrime
  - Illegal activity as much as $72 Billion p/a [0]

coinbase

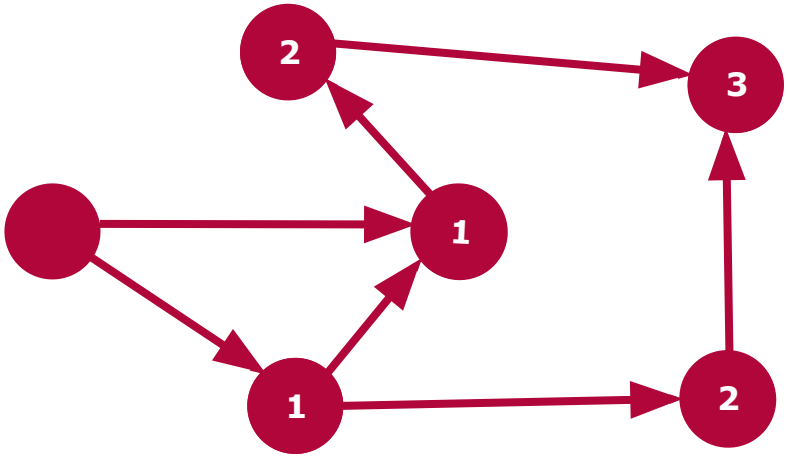


[0] Foley, Sean, Jonathan R. Karlsen, and Tālis J. Putniņš. "Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies?." *The Review of Financial Studies* 32.5 (2019): 1798-1853.

Topic Presentation
Alexander Mühle

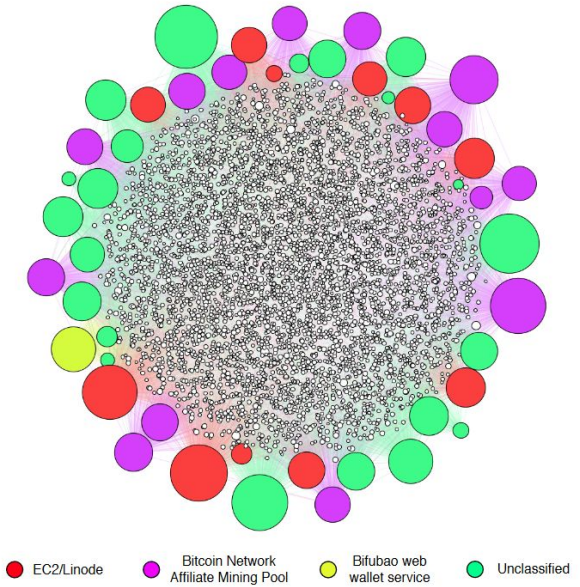
Peer-to-Peer Message Exchange

- Messages are propagated like **Gossips**
- New messages are sent to one's peers (typically 5-7 active neighbours)



Peer-to-Peer Neighbour Discovery

- Neighbour discovery done by iterative address requests
- Bitcoin clients keep track of ~20.000 peers



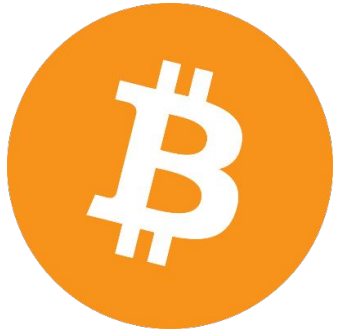
It’s not What You Know but Who You Know

“When systems are large and individual nodes only gain random knowledge of part of the network, their traffic can be detected by uniqueness of the information they have learnt”

[2] G. Danezis and R. Clayton, “Route Fingerprinting in Anonymous Communications,” in *Sixth IEEE International Conference on Peer-to-Peer Computing (P2P’06)*, Sep. 2006, pp. 69–72, doi: 10.1109/P2P.2006.33.

Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Topic 1: Fingerprinting Bitcoin peers



- Can we track Bitcoin peers through the information we can gather on them?
  - **Peer database**
  - Handshake information
  - Offline time, ...
- Collect and analyse information (building on existing network crawlers) for uniqueness using Spark/Zeppelin
- Evaluate and test your approach in the real Bitcoin network

Project: Goals

What will we do if you choose this topic

- **Learn about** peer-to-peer message exchange and propagation
  - Gossip Protocols (Bitcoin…)
  - Peer tables and neighbour discovery
- **Program** network software (i.e python3)
- Basics of an **ETL process** (extract, transform, load)
  - Elasticsearch, Spark
- **Write a paper** with your advisor

Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Bring Your Own Ideas

Do you have other interesting ideas on what to do with collected network data?

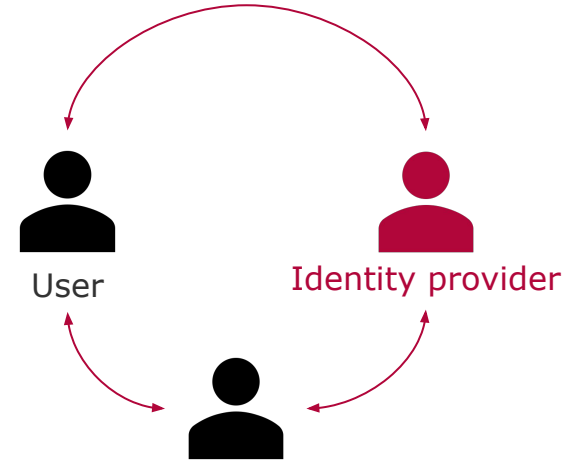
Secure Identity Lab

Self-Sovereign Identity

Andreas Grüner

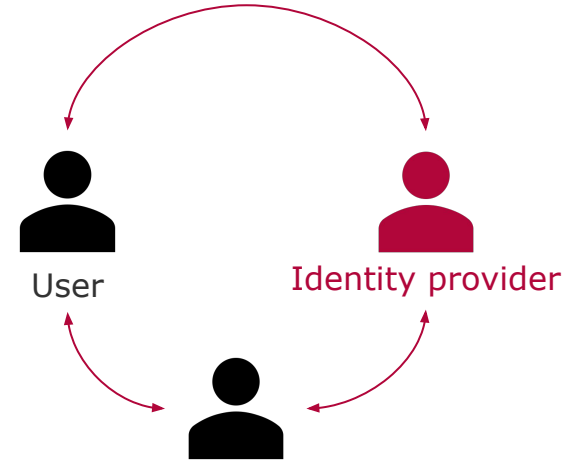
andreas.gruener@hpi.de

Identity Management
Main Actors



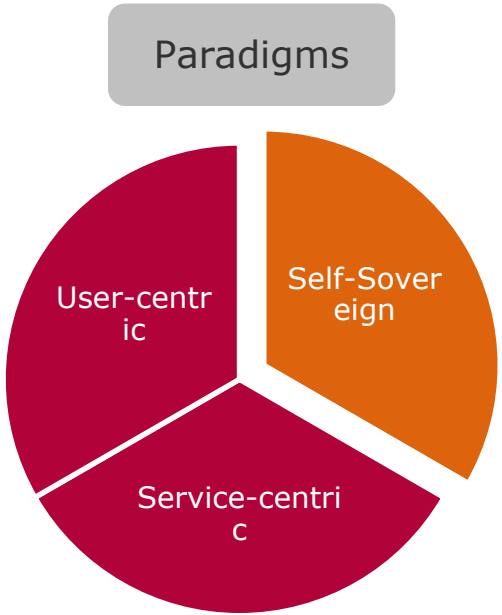
- Functions of the identity provider
  - Enrollment
  - Authentication
  - Authorization
  - Credential Management
  - Attribute Management and Verification

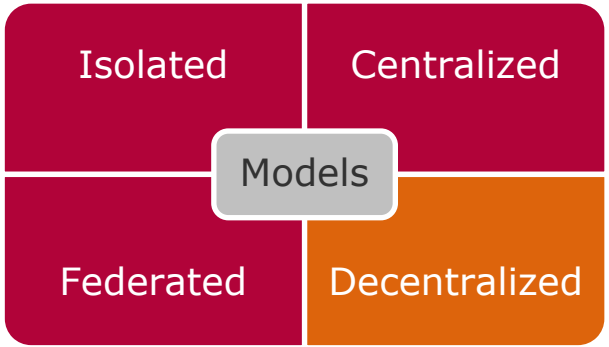
Identity Management
Identity Provider as a Trusted Third Party



- Drawbacks
  - Single point of failure
  - Control of digital identity and service
  - Profitable target
  - Endangered data privacy
  - Not trusted by everybody

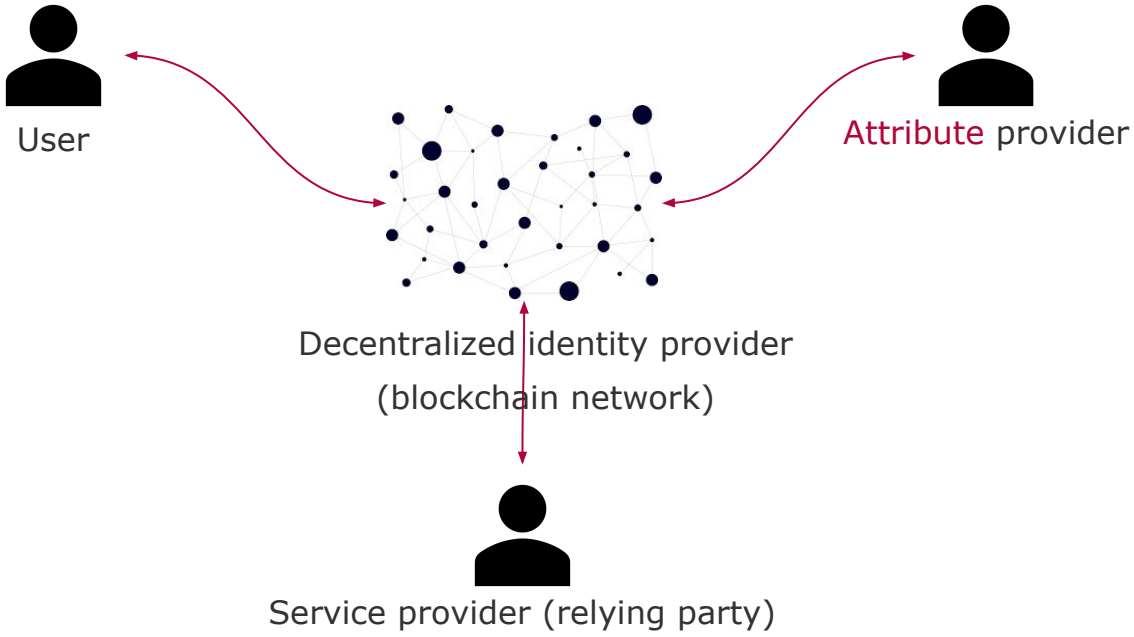
Identity Management
Paradigms and Models





Self-Sovereign Identity
Overview

Self-Sovereign Identity (SSI): *"individual control across any number of authorities"* (Allen)



Self-Sovereign Identity Solutions

Seminar Topics
Topic 1 - Interoperability between Hyperledger Indy

- Interoperability belongs to Allen's principles of self-sovereignty
- Existence of a myriad of SSI solutions and networks following their individual governance models
  - VON establishes a HL Indy network in Canada
  - IDunion creates a HL Indy network for Germany

- How can one connect these HL Indy networks for interoperability?
  - Build a formal concept for communication. Which transactions need to be transmitted?
  - Implement and simulate the interoperability concept
  - What are the advantages and disadvantages of the approach?

Seminar Topics
Topic 2 – Identity Management Usage on the Internet

- Any modern web application (on the Internet) requires identity management to recognize and authorize users
- A variety of options for identity management (following the models) exists
  - Registration and account creation (isolated model)
  - Use of a third party (e.g. Facebook) identity provider (centralized/ federated model)
  - Integration of a SSI solution (decentralized model)

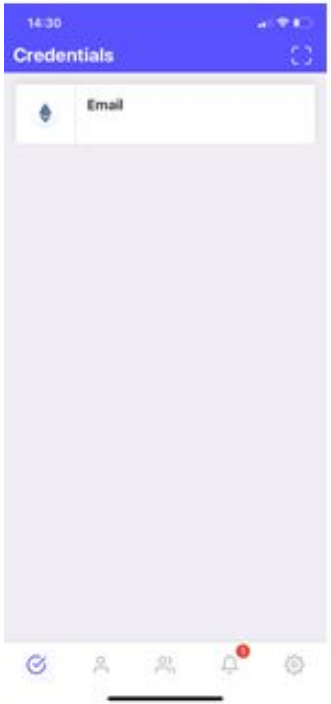
- Which identity management models are used on the Internet?
  - Automatically scan the internet for web applications that require a login
  - Classify the found pages according to the models
  - Evaluate the results (How often are SSI solutions already used? What are the major centralized identity providers? etc.)

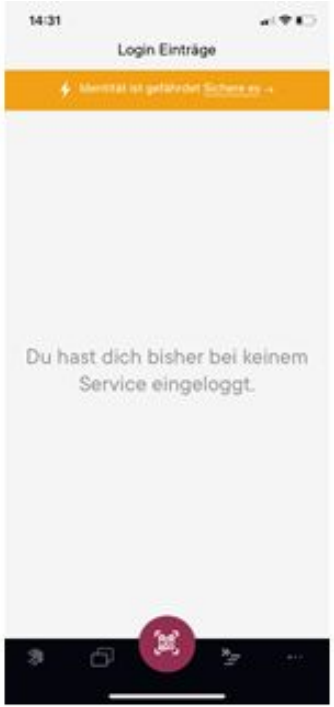
Seminar Topics
Topic 3 - Usability of Identity Wallets/ End User Agents (1/2)

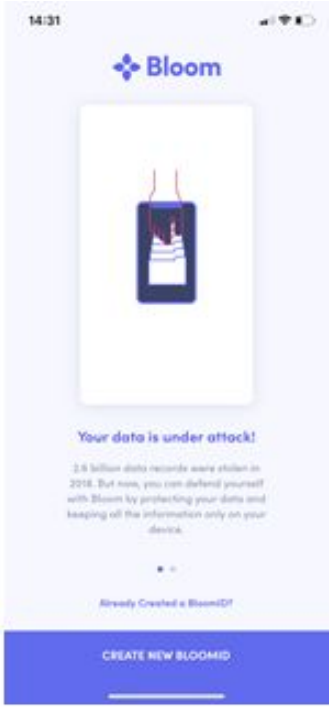
- SSI solutions provide an identity wallet or agent for end users to control and use their digital identity for interactions
- SSI and their identity wallets/ agents are a new topic for end users. Therefore, usability plays an important role.

- Are currently existing identity wallets usable?
  - What is usability? How is usability measured?
  - What are core functionalities of an identity wallet/ agent?
  - How is the usability of major solutions?
  - What are deficiencies/ improvements for the major solutions?

Seminar Topics
Topic 3 - Usability of Identity Wallets/ End User Agents (2/2)







Topic Presentation
Andreas Grüner

Seminar Topics
Topic 4 - Security of Identity Wallets/ End User Agents

- The goal of SSI wallet solutions is to store all our data / credentials for easy access...
- How secure is that “storage” actually?

- Possible tasks/questions could include:
  - How do current actors perform security for their wallets? (DAAD-Wallet, MIT-Wallet, Lissi / idUnion, …)
  - How is security performed in other critical applications? (E.g. Banking Applications, Google / Apple Pay, …)
  - How can secure storage environments such as Androids Trusty TEE be used to securely store “Wallet”-Contents?
  - How can recovery of “Wallet”-Contents be enabled in SSI Wallet Solutions?

Secure Identity Lab Seminar
Organization

Seminar Goals

Our goals for this seminar

- You should learn to dig into a specific topic and find a gap you want to fill with your group
  - Find and analyze related work
  - Define your own research question for the seminar
  - Understand and apply new technologies in a research context
- You should learn to self-organize your group work in a defined timeframe
- You should learn how to write a research paper
- You should learn how to communicate with your team / supervisors
  - If a problem occurs: Identify it, Talk about it (with us), Control / Fix it!

Workload

- This seminar will give you **6 ECTS** if you finish successfully
  - **1 ECTS ~ 30h** of work => In total, spending about **180h** is reasonable

- We offer different modes: *lecture time* vs *semester time:*
  - 12.04.2021 – 23.07.2021 (~14 weeks) => 13h work a week per student
  - 12.04.2021 – 20.09.2021 (~22 weeks) => 8h work a week per student
  - presentation at the end of that time, documentation deadline shortly after
  - ..after topic assignment we will find a mode matching all groups
- Although calculation mostly holds theoretically, rule of thumb for our expectations during progress meetings

Timeline (approx) for “lecture time” seminar

14.04.2021 Official first lecture / meeting, Q & A session

18.04.2021 Submission of interest

19.04.2021 Topic Assignment

28.04.2021 Idea Pitch / Get together

21.04.2021 Introduction to Academic Writing

05.05.2021 Intro + Related Work documented *

18.05.2021 Idea Presentation / Amazing Prototype

16.06.2021 Approach documented *

07.07.2021 Evaluation + Conclusion documented *

21.07.2021 Final Presentation

28.07.2021 Code Submission & Paper Submission

Provisional Dates!

usually, we will have a weekly meeting with each group to talk about progress, problems, etc.. Time & kind of meeting is negotiated individually

* … you will get a detailed review from us afterwards and then present to the other groups

Secure Identity Lab Seminar

Eric Klieme
Alexander Mühle
Andreas Grüner
Daniel Köhler

Evaluation

- Idea Presentation ~15%
  - Motivation, Related Work, Rough Approach / First Prototype
- Final Presentation ~25%
  - Idea Presentation + Full Approach, Evaluation, Discussion, Future Work
- Report ~30%
  - IEEE / ACM conference paper style
- Implementation ~20%
  - Readme, Logging/Tracing, Automation, Architecture / Code Docs etc.
- Communication ~10%
  - Meeting Organization / Protocols, Questions & Concerns, Problems, Active Discussion Requests etc.

Enrollment

- If you are interested (as a single person, as a team):
  - Enroll to Moodle
    - https://moodle.hpi.de/course/view.php?id=165
  - Pick a topic

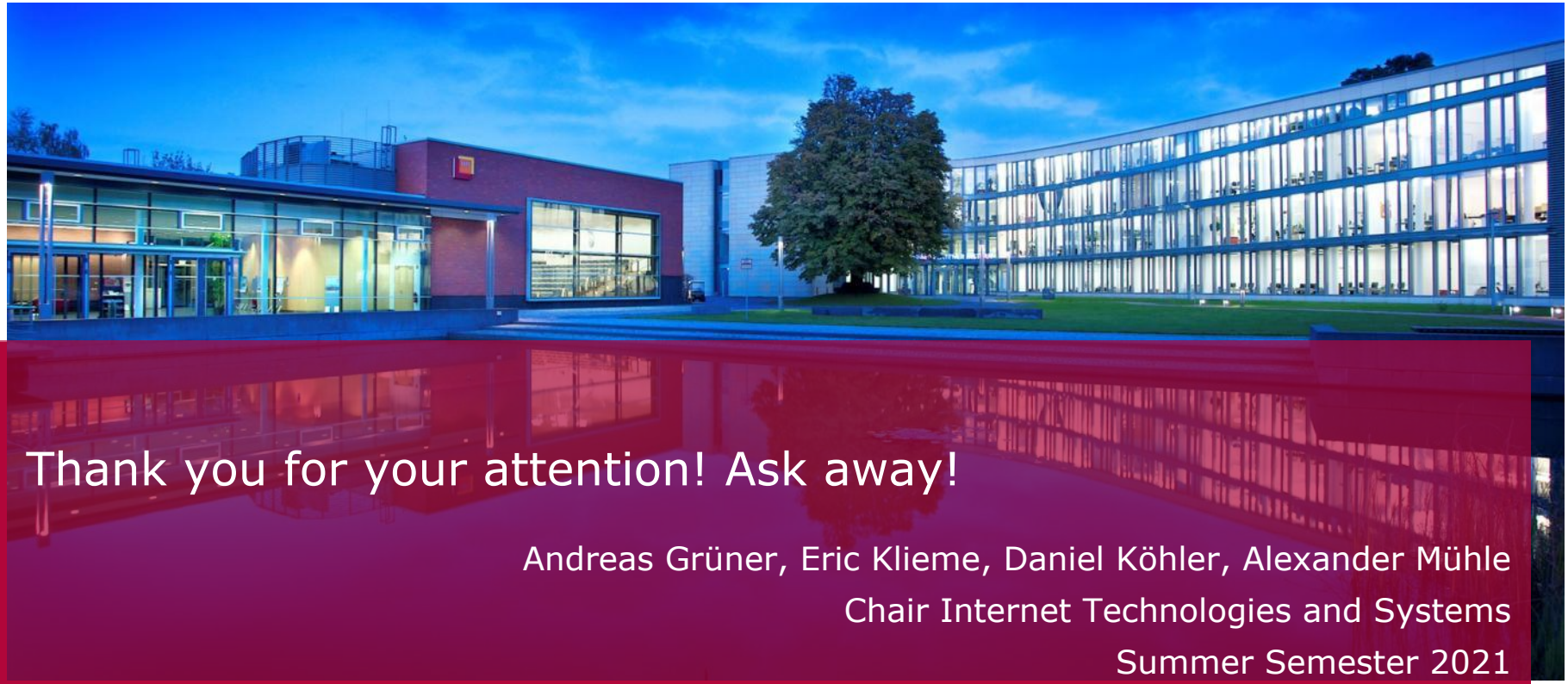
6 April - 18 April

Topics Selection

- Until 9th of April (Friday!)
  - belegung-master-2021@hpi.de

Secure Identity Lab Seminar

Eric Klieme
Alexander Mühle
Andreas Grüner
Daniel Köhler



Thank you for your attention! Ask away!

Andreas Grüner, Eric Klieme, Daniel Köhler, Alexander Mühle
Chair Internet Technologies and Systems
Summer Semester 2021