HPI Hasso Plattner Institut
Digital Engineering • Universität Potsdam

Hot Topics in Secure Identity Research

Winter Semester 2020/2021

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

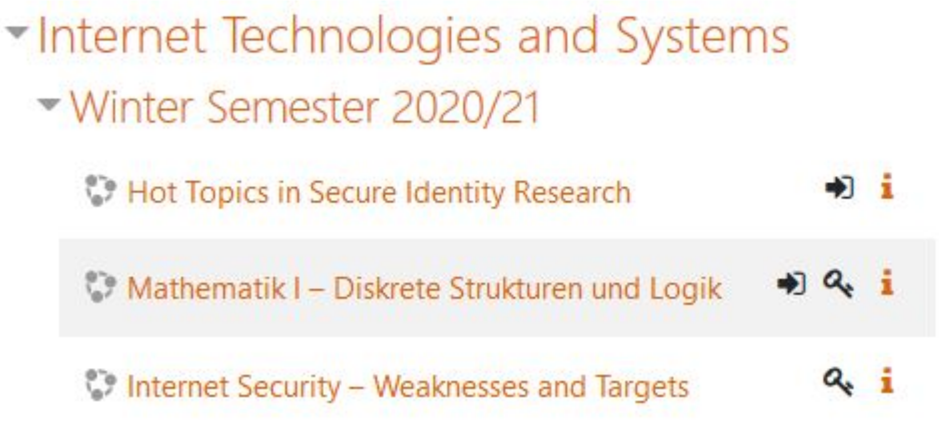
Winter Semester 2020/2021

Topics - Summary

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

MOODLE

https://hpi.de/friedrich/moodle/course/view.php?id=132



Internet Technologies and Systems

Winter Semester 2020/21

- Hot Topics in Secure Identity Research
- Mathematik I – Diskrete Strukturen und Logik
- Internet Security – Weaknesses and Targets

For questions during the presentation:
uni-potsdam.zoom.us/j/65841318030 (Passcode: 23646388)

Hot Topics in Secure Identity Research

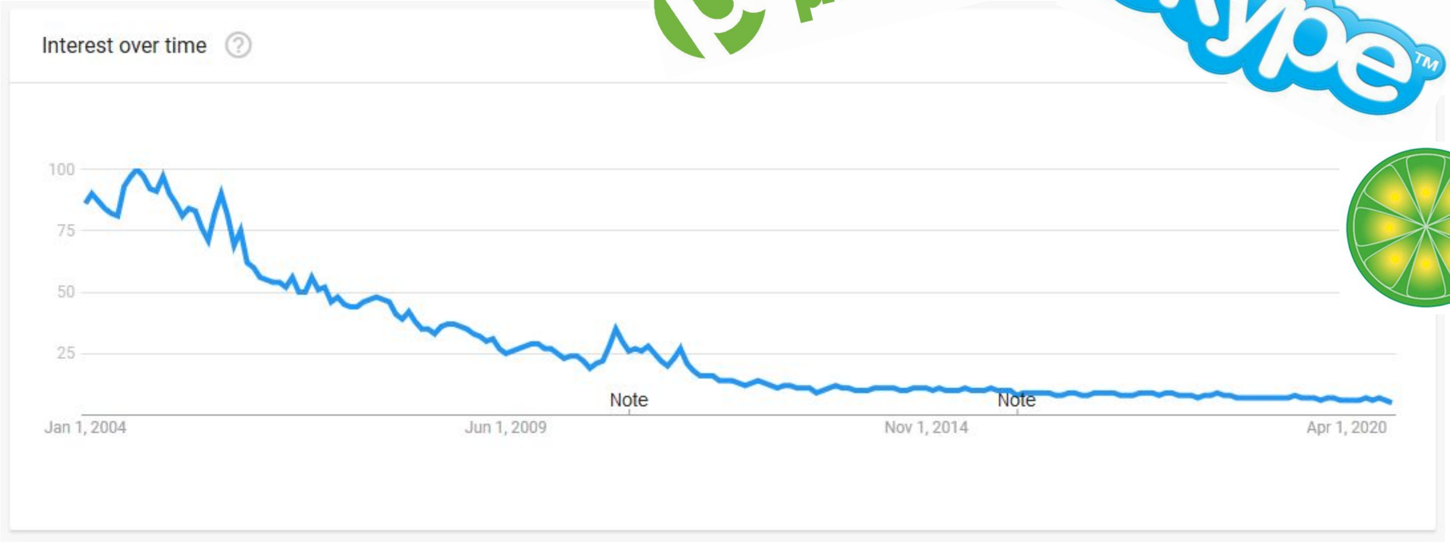
Analysing P2P Network Participants

Alexander Mühle

alexander.muehle@hpi.de

P2P Networks

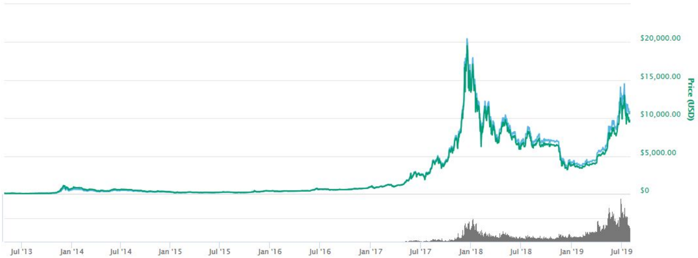
- Peer-to-Peer… who cares?



HPI – Secure Identity Lab
Cryptocurrencies

- Bitcoin
  - First published 2008
  - 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 [0]

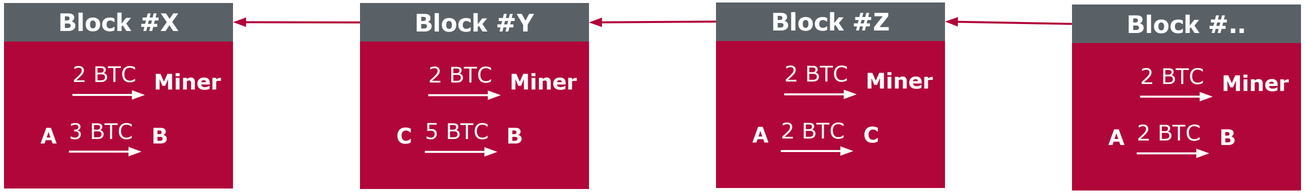
coinbase



[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.

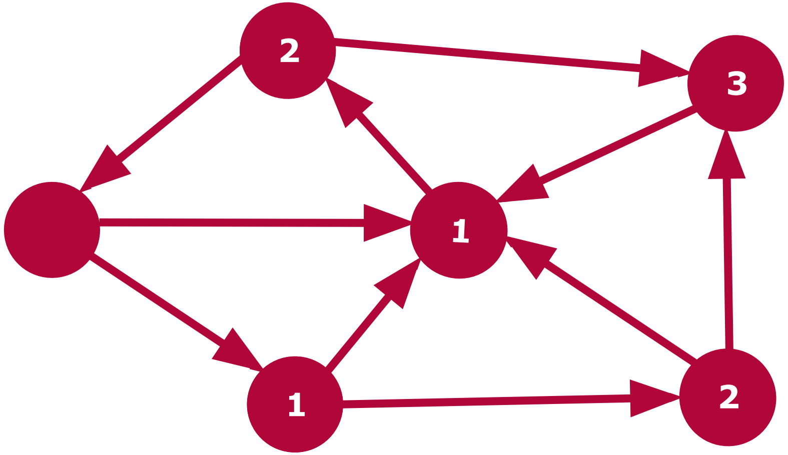
HPI – Secure Identity Lab
Transactions and Blocks

- **Distributed ledger** of financial data
- Participants publish **transactions** to the network
- **Miner** gather multiple transactions into a **block**
- Blocks are linked in a **chain**
- In order to publish a Block some **work** (Proof of Work) has to be done
- The chain with the **most work** is seen as the **consensus**
- Miner get a **reward** for new Blocks



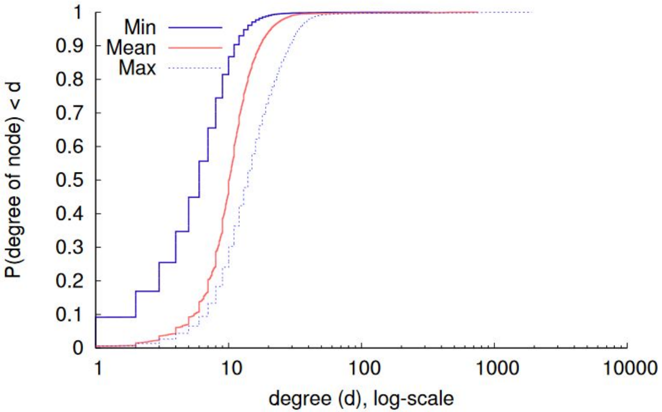
Peer-to-Peer Message Exchange

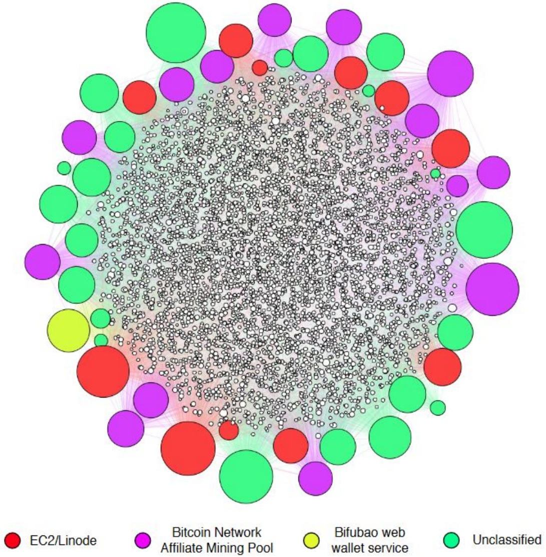
- Messages are propagated like **Gossips** (or structured approaches)
- New messages are sent to one's peers



Bitcoin: Peer-to-Peer Message Exchange

- Most nodes have between **active 7-12 Neighbours** [1]
- Some nodes are very well connected
  - Miners
  - Exchanges

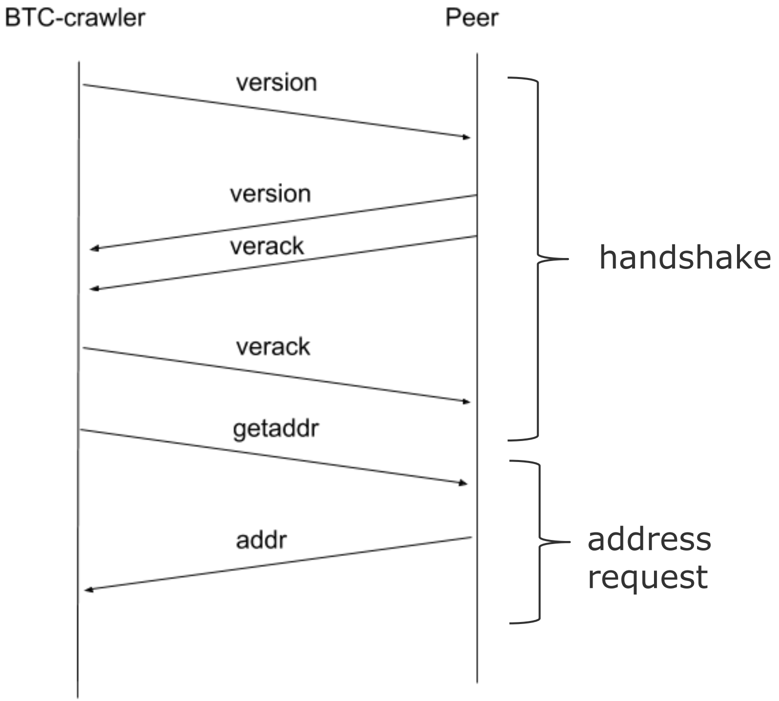




EC2/Linode
Bitcoin Network Affiliate Mining Pool
Bifubao web wallet service
Unclassified

[1] Miller, Andrew, et al. "Discovering bitcoin’s public topology and influential nodes." *et al* (2015).

HPI – Secure Identity Lab
Bitcoin: Peer Discovery



- 1000 addresses per message
- 2500 addresses per request
- Addresses are selected at random for each request
- Reference Implementation has maximum of 20480 addresses
- $\left(1 - \frac{1}{\frac{20480}{2500}}\right)^x$

Hot Topics in Secure Identity Research

Alexander Mühle

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.

HPI – Secure Identity Lab
Project: Analysing Network Data

Topic 1: Fingerprinting Bitcoin peers

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

HPI – Secure Identity Lab
Project: Analysing Network Data

Topic 2: Crawling and Analysing Ethereum peers

- How can we get the most complete view of Ethereum node peer databases?
- Can we estimate node age of Ethereum peers reliably?
  - □ **Peer database**
  - □ Kademlia buckets
- Analyse collected information (building on the existing network crawler) for uniqueness using Spark/Zeppelin
- Evaluate and test your approach in the real Ethereum network

HPI – Secure Identity Lab
Project: Goals

What will we do if you choose one of these topics

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

Bring Your Own Ideas

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

HPI Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Hot Topics in Secure Identity Research

Behavioral Authentication

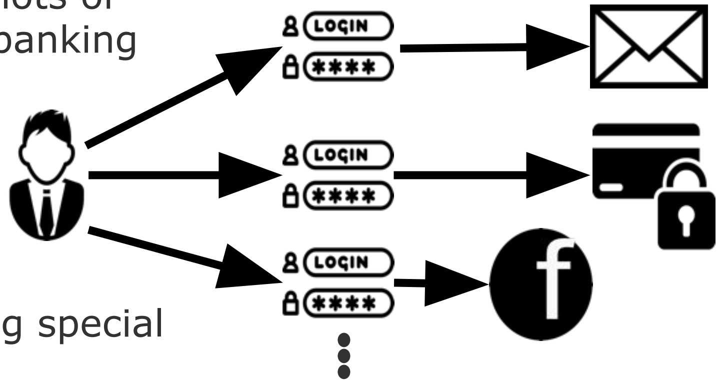
Eric Klieme
eric.klieme@hpi.de

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

Hasso Plattner Institut
Digital Engineering · Universität Potsdam

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)

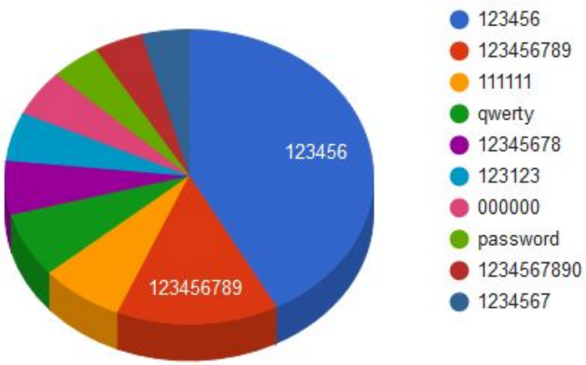
- Complex passwords hard to remember, use a much simpler
- Same passwords for different services

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%)

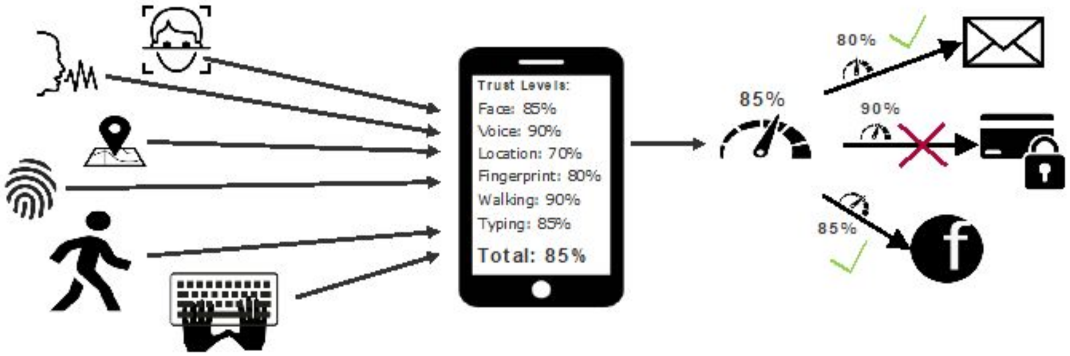


Distribution of top 10 leaked passwords

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

Alternative: Analyze user‘s behaviour continuously in the background based on **behavioral biometrics**

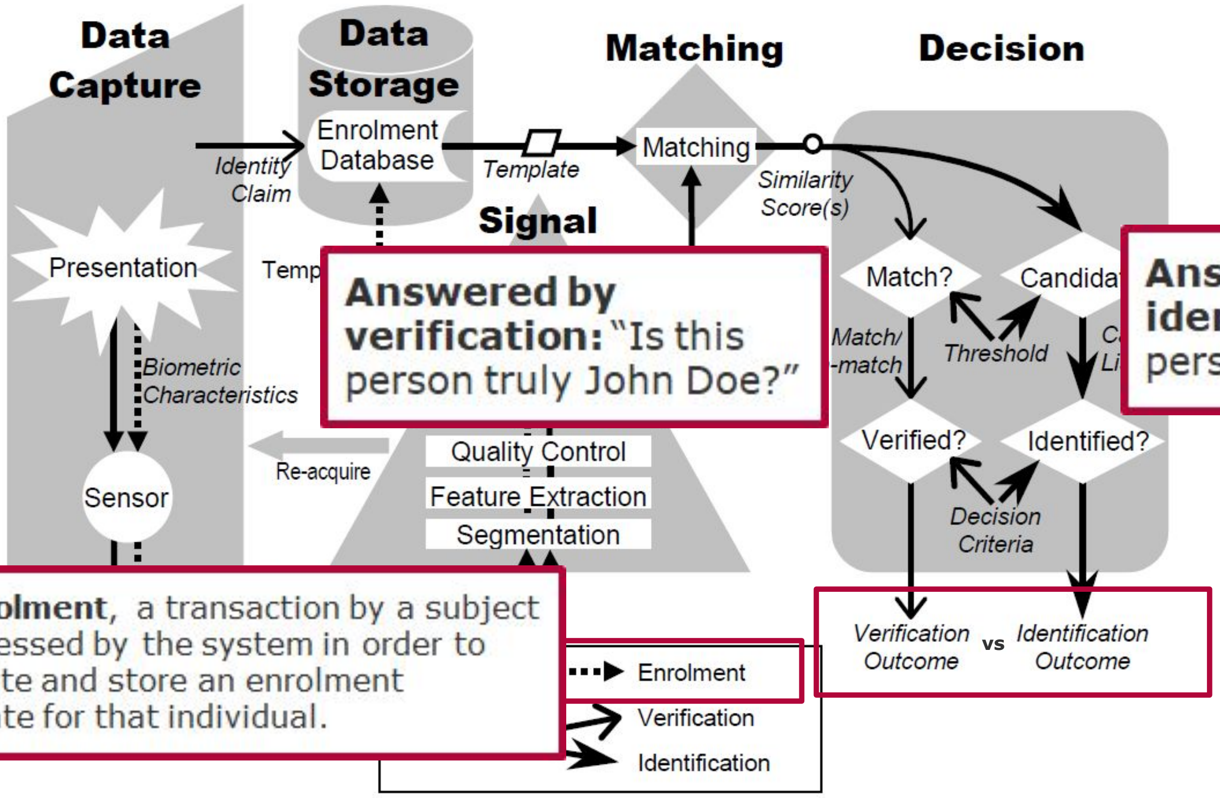
- Access sensors on devices around users
  - Devices that are “always” near-by, e.g. (smartphone, wearables)
  - Devices that are used, e.g. notebooks, pcs
- Sample behaviour via sensors, e.g. accelerometer, microphone, APIs
- Use biometric systems to create templates of users based on behaviour and 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

Answered by verification: "Is this person truly John Doe?"

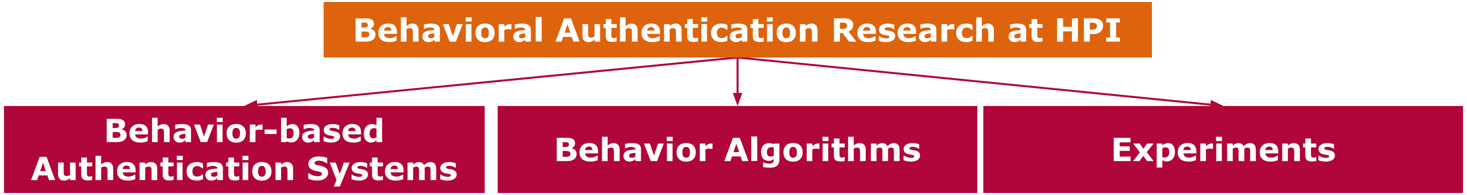
Answered by identification: "Is this person in the database?"

In **enrolment**, a transaction by a subject is processed by the system in order to generate and store an enrolment template for that individual.

Topic Presentation

Eric Klieme

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



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: **Experiments / Authentication Systems**
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”

- 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…)

Topic Presentation

Eric Klieme

Topic 1: **Experiments / Authentication Systems**
Behavioral Authentication Implementation Platform

- Contribution: „Proof-Of-Concept"
  - Search and Analyze a lot of related technologies for machine learning 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: **Experiments / Algorithms**
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 like? (Computer Vision)
  - How they approach it (Bluetooth Signal)
- “Basic” Prototype with extended door handle, data collection server, and user equipment exists
- **Project:** Finish prototype and do large scale user study and Machine Learning
  - Finish the prototype for robust data recording, participant tracking and more
  - Come up with a plan for a large-scale collection study, e.g. different offices, kitchens etc. - Although Covid-19 WiMis are in the Office
  - Finally apply ML in specific verification or identification scenarios

Topic Presentation

Eric Klieme

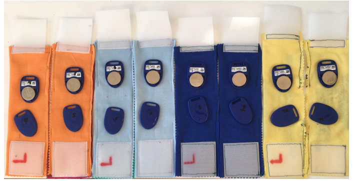
Topic 2: **Experiments / Algorithms**
Prototype “Evolution”

Version 1

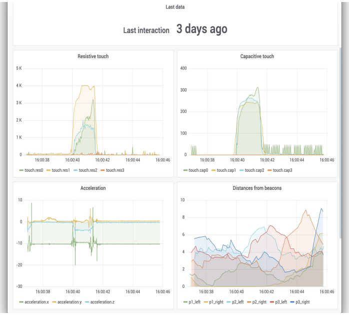
door handle with touch sensor

Version 2

door handle with touch sensor and accelerometer



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



Last data

Last interaction 3 days ago

data collection infrastructure with dashboard

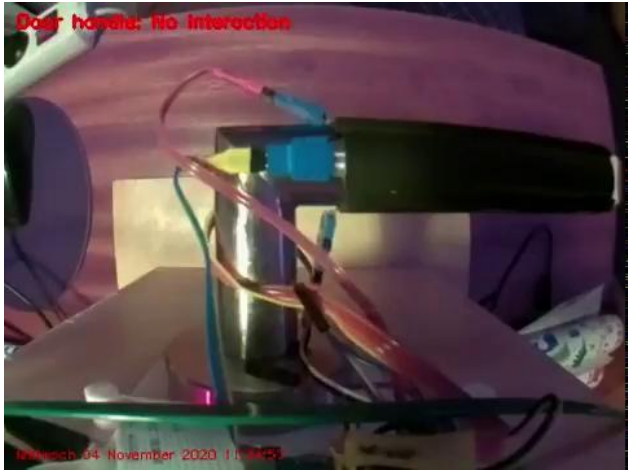
Topic Presentation

Eric Klieme

Topic 2: **Experiments / Algorithms**
Prototype “Evolution”

Version 3 : **Goal** of this seminar

Version 2 **+ Computer Vision + X (your ideas)**



Topic Presentation

Eric Klieme

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

Topic 3: **Experiments**
Integrate users into labeling

- Data collection is hard for behavioral data
- Problem:
  - Participants should behave **as natural as possible** with all contexts (e.g. different states of mind, clothes, times during a day etc.)
  - **But:** Motion data is sampled by sensors and not very descriptive
    - Who can see what a person is doing just by seeing X-Axis of accelerometer of smartphone, for example?
  - ☐ **Until now**: Supervised collection of data with researchers paying attention to correct labelling in a specific scenario
    - Adds a lot of bias and becomes very complex for diverse contexts

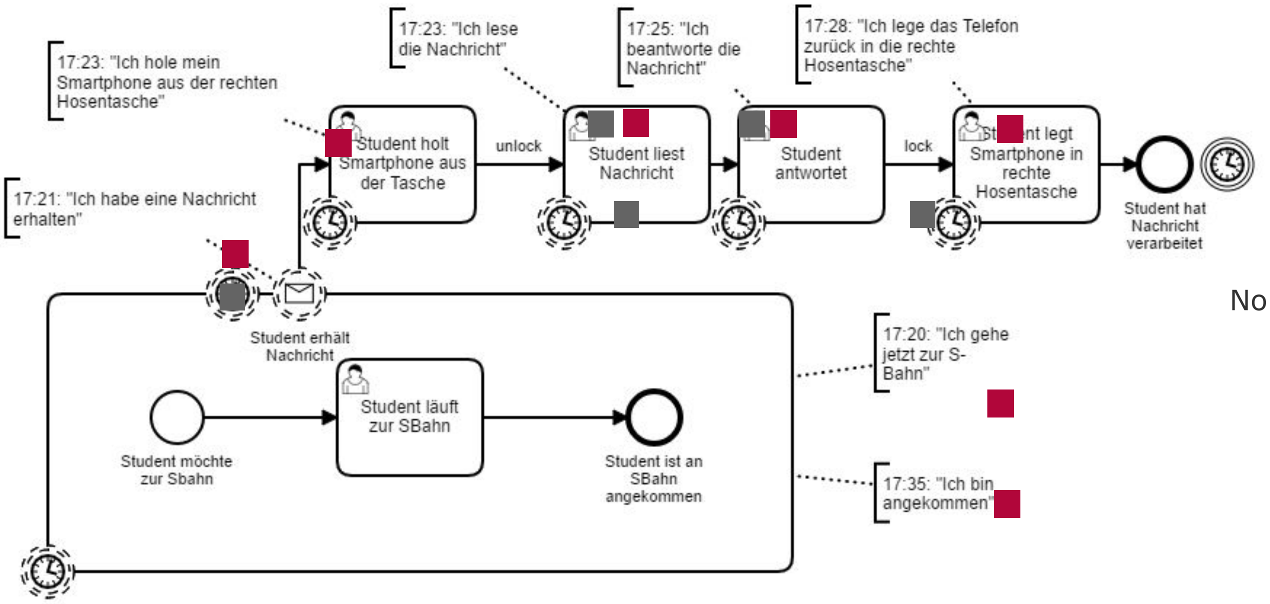
- **Idea:** Let the user help us
  - They label their own data here and then
  - Self-labeling used in psychology research

Topic Presentation

Eric Klieme

Example: A user interacts with the smartphone on the way to public transport

Access to information sources to sample behavior can be quite difficult


17:23: "Ich hole mein Smartphone aus der rechten Hosentasche"
17:23: "Ich lese die Nachricht"
17:25: "Ich beantworte die Nachricht"
17:28: "Ich lege das Telefon zurück in die rechte Hosentasche"
17:21: "Ich habe eine Nachricht erhalten"
Student holt Smartphone aus der Tasche
unlock
Student liest Nachricht
Student antwortet
lock
Student legt Smartphone in rechte Hosentasche
Student hat Nachricht verarbeitet
Student erhält Nachricht
Student möchte zur Sbahn
Student läuft zur SBahn
Student ist an SBahn angekommen
17:20: "Ich gehe jetzt zur S-Bahn"
17:35: "Ich bin angekommen"

No access possible (subconsciousness)

🟥 Student, implicit

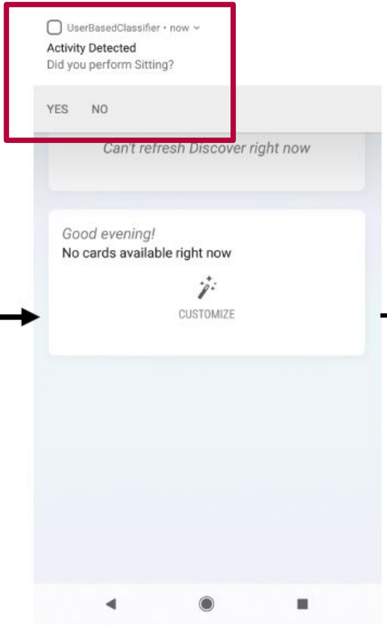
⬛ App, explicit

Access limited, sometimes API but not available everywhere

Idea: Basic activity detection in background, push notifications for user-request to annotate

Possible flow?!

The model classifies an activity of the user (e.g. sitting)



□ UserBasedClassifier • now ⌄

Activity Detected

Did you perform Sitting?

YES NO

Can't refresh Discover right now

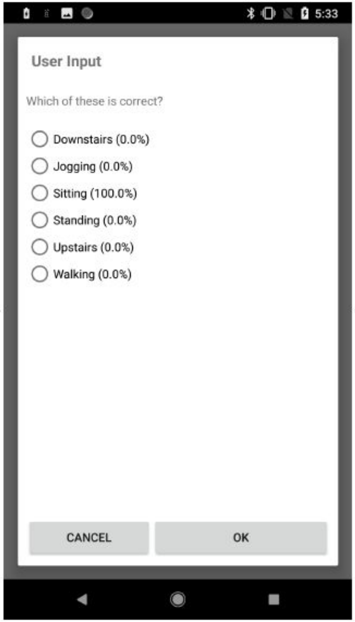
Good evening!

No cards available right now

CUSTOMIZE

Notification

User gets notified about the activity. He can say whether the class is correct. If it is incorrect and he wants to correct it, he can tap on the classification.



5:33

User Input

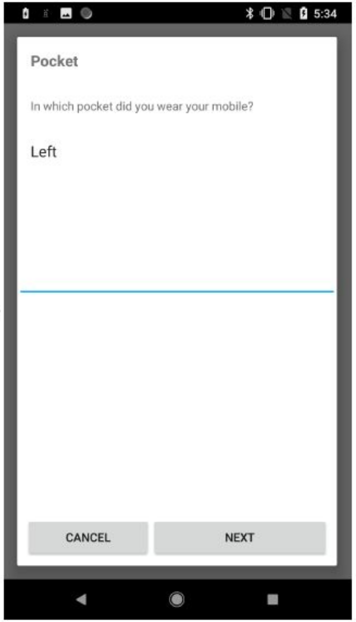
Which of these is correct?

○ Downstairs (0.0%)
○ Jogging (0.0%)
○ Sitting (100.0%)
○ Standing (0.0%)
○ Upstairs (0.0%)
○ Walking (0.0%)

CANCEL OK

Correction

User inputs to the system the class that would be correct.



5:34

Pocket

In which pocket did you wear your mobile?

Left

CANCEL NEXT

Context

User gives context information.

Topic 3: User-supported large-scale data collection application

- Contribution: „Proof-Of-Concept / User Study"
  - Related Work analysis
  - App and Backend Engineering, Algorithms
    - first PoC exist based on own app + Aware Experiment Platform Backend (runs within HPI)*
  - Evaluation in a user study, mostly Usability / Feasibility of approach
    - **Should perfectly work with COVID-19** - since we want any person to take part individually

- Nice-To-Have Skills
  - Java/Kotlin, Machine Learning, Tensorflow
  - Strong communication skills, creativity
  - Interest in conducting studies

* https://awareframework.com/

Topic Presentation

Eric Klieme

Behavioral Authentication

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 ;)

HPI Hasso Plattner Institut
Digital Engineering · Universität Potsdam

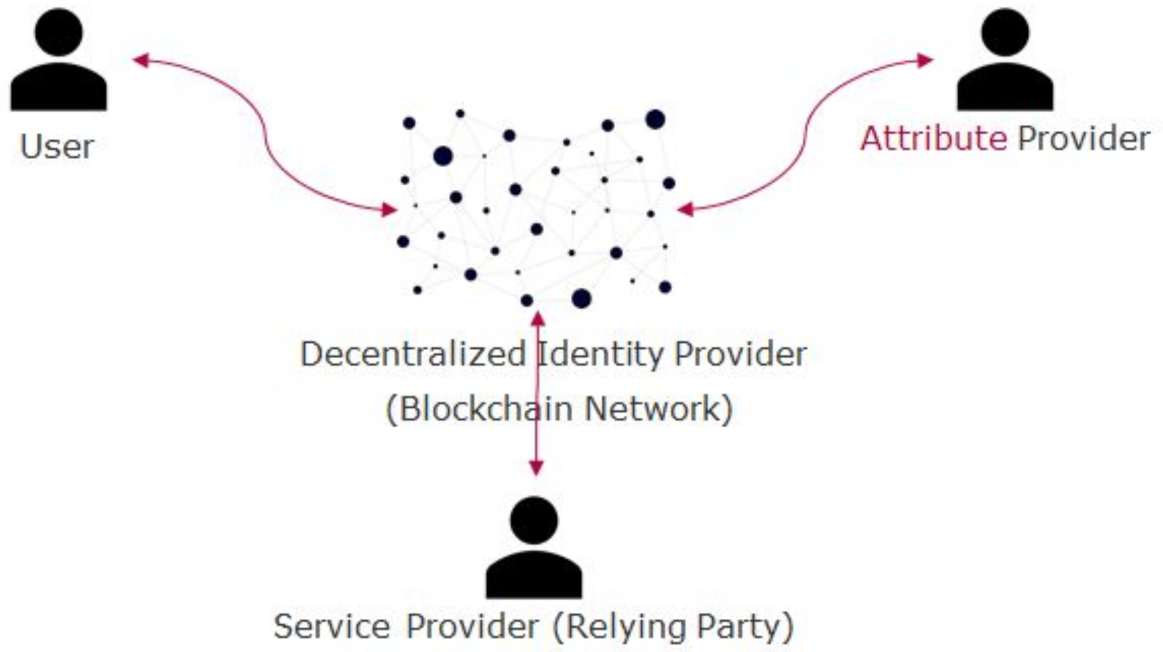
Hot Topics in Secure Identity Research

Self-Sovereign Identity Ecosystem

Andreas Grüner
andreas.gruener@hpi.de

Self-Sovereign Identity Overview

Self-sovereign Identity: "*individual control across any number of authorities*" (by C. Allen)


User
Attribute Provider
Decentralized Identity Provider
(Blockchain Network)
Service Provider (Relying Party)

Self-Sovereign Identity Solutions

u·port
BLOCKCHAIN HELIX™
HYPERLEDGER INDY
SELFKEY
evernym
HYPERLEDGER ARIES
ShoCard
JOLOCOM
civic
KYC-CHAIN
bitID
BLOCKSTACK
sovrin
identity for all

and many more ...

Project 1: Interoperability of SSI Solutions and Networks

- Existence of a myriad of SSI solutions and networks following different implementation approaches
  - uPort – Smart contracts on Ethereum
  - Jolocom – Smart contracts on Ethereum
  - Sovrin (Hyperledger Indy/ Aries) – Dedicated set of blockchains
  - Blockchain Helix/ Civic/ SelfKey

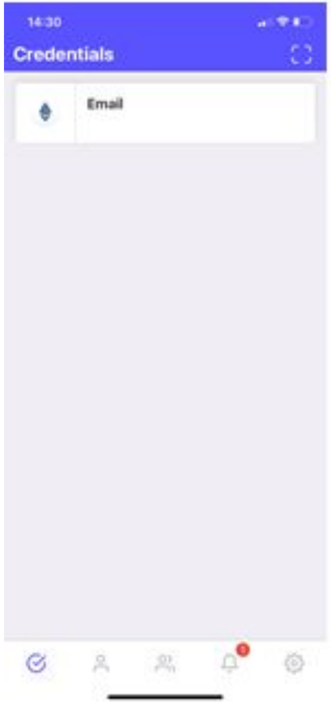
- What means "interoperability" and which level exists?
- Which concepts and approaches for interoperability exists or could be developed?
- What are the advantages and disadvantages of these approaches?

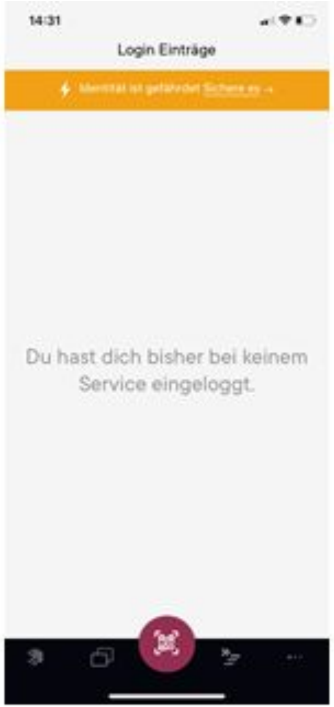
Project 2: Usability of Identity Wallets/ End User Agents

- 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.

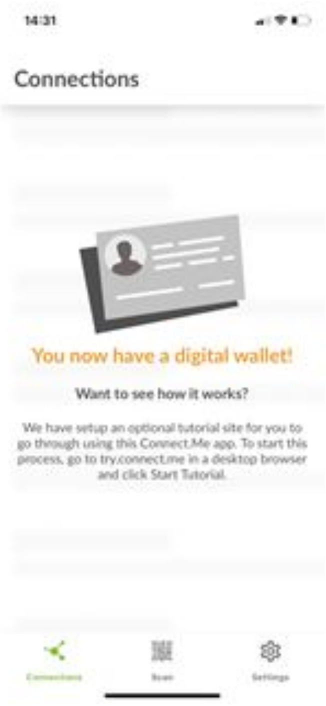
- What is usability? How is usability measured?
- What are core functionalities of an identity wallet/ agent?
- How is the usability of major solutions?
- What are deficiencies/ improvements for the major solutions?

Project 2: Usability of Identity Wallets/ End User Agents





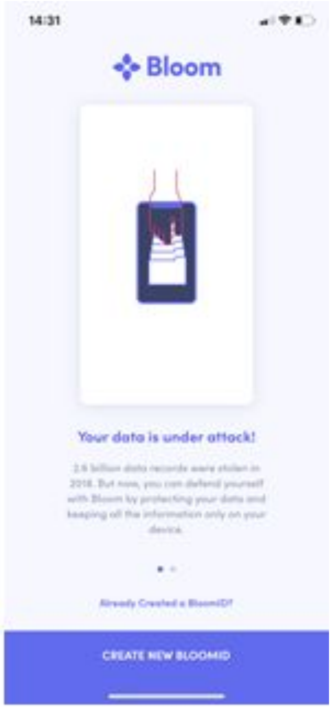
Du hast dich bisher bei keinem Service eingeloggt.



You now have a digital wallet!

Want to see how it works?

We have setup an optional tutorial site for you to go through using this Connect.Me app. To start this process, go to try.connect.me in a desktop browser and click Start Tutorial.



Your data is under attack!

2.6 billion data records were stolen in 2018. But now, you can defend yourself with Bloom by protecting your data and keeping all the information only on your device.

Project 3: Systematization of the SSI Ecosystem

- The current SSI ecosystem is overwhelmingly confusing. Manifold actors, projects, initiatives, standardization bodies, governmental groups, communities, companies, universities and research groups exists.
- Different SSI solutions, blockchains, protocols, frameworks, interoperability products and further aspects are developed

- What actors, initiatives and projects exists?
- What do these actors develop? What are their interests?
- How can they be arranged in a taxonomy/ classification?

Project 4: Security Analysis Methodology for SSI

- SSI and blockchain are emerging concepts
- Identity management is on the forefront of security and must be secure itself
- Formal security analysis approaches exists to improve security posture of systems

- What is the general SSI architecture? What are the components?
- Which existing security analysis methodologies (e.g. Attack Trees) can be applied?
- What would be a SSI specific security analysis methodology?

Project 4: Security Analysis Methodology for SSI

- Mühle, A., Grüner, A., Gayvoronskaya, T., Meinel, C.: A Survey on Essential Components of a Self-Sovereign Identity. Computer Science Review. 80-86 (2018)
- Kupferberg, M.: Blockchain-Based Identity Management: A Survey from the Enterprise and Ecosystem Perspective. IEEE Transactions on Engineering Management. 2019
- Universal Resolver. Online: https://github.com/decentralized-identity/universal-resolver
- Hyperledger Indy. Online: https://www.hyperledger.org/projects/hyperledger-indy
- Hyperledger Aries. Online: https://www.hyperledger.org/projects/hyperledger-aries
- C. Lundkvist, R. Heck, J. Torstensson, Z. Mitton, M. Sena: UPORT: A PLATFORM FOR SELF-SOVEREIGN IDENTITY. Online: http://blockchainlab.com/pdf/uPort_whitepaper_DRAFT20161020.pdf
- uPort. Online: https://www.uport.me
- Decentralized Identifiers (DIDs) v1.0. Online: https://www.w3.org/TR/did-core/
- Verifiable Credentials Data Model 1.0. Online: https://www.w3.org/TR/vc-data-model/

HPI Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Hot Topics in Secure Identity Research

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 would consider *lecture time* as *working time:*
  - 02.11.2020 – 12.02.2021 (~14 weeks)
  - presentation at the end of that time, documentation deadline shortly after
- 180h/14 weeks => 13h work a week per student
  - ~ **1,5 days of working** for the seminar only **per week**
  - A group of four students ~ 52h (7-8 PD) **per week**
- Although calculation mostly holds theoretically, rule of thumb for our expectations during progress meetings

Timeline (approx)

02.11.2020	Official first lecture / meeting, Q & A session
08.11.2020	**Submission of interest**
13.11.2020	Topic Assignment / Discussion
16.11.2020	**Start working**
21.12.2020	Intro + Related Work documented *
10.01.2021	**Idea Presentation / Amazing Prototype**
17.01.2021	Approach documented *
14.02.2021	Evaluation + Conclusion documented *
21.02.2021	**Final Presentation**
28.02.2021	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

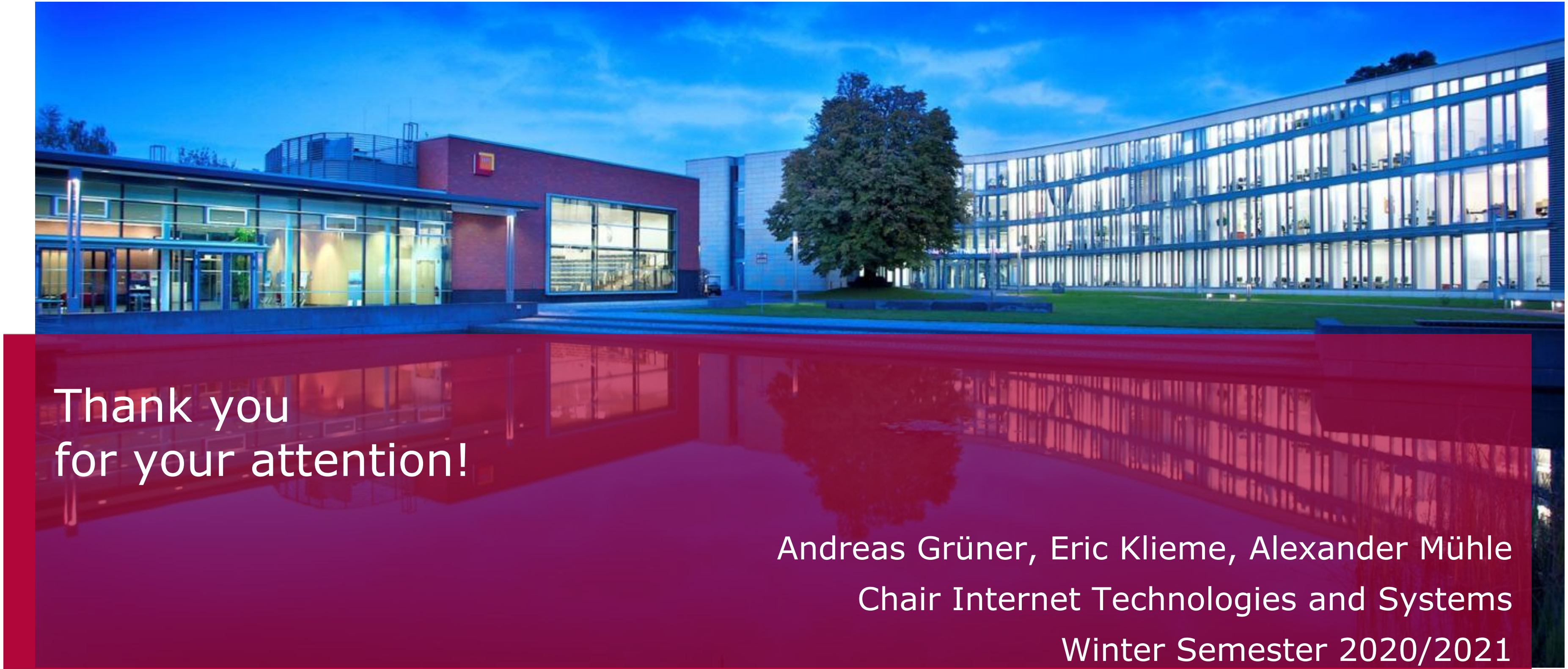
Evaluation

- 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://hpi.de/friedrich/moodle/course/view.php?id=132

- Until 8th of November (Sunday!)
  - Ranked topic selection of top 3 choices



Thank you for your attention!

Andreas Grüner, Eric Klieme, Alexander Mühle

Chair Internet Technologies and Systems

Winter Semester 2020/2021