HPI Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Hot Topics in Secure Identity Research

Eric Klieme, Alexander Mühle, Muhammad Sukmana, Christian Tietz
Chair Internet Technologies and Systems

Summer Semester 2020

Background

- Entity: *A thing with distinct and independent existence*
  - Example: a person, a system, a group, etc.

- Identity: *The subset of an entity that distinguishes the entity in a given context*
  - Example: **You** are HPI students

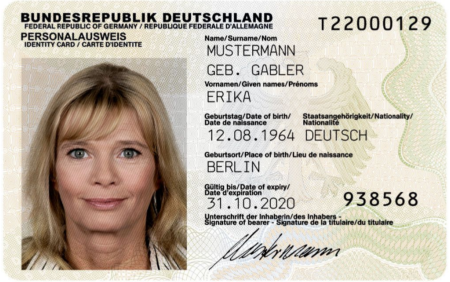
- Attribute: *The characteristic of an entity*
  - Example: Eye color, status, name

https://solutionscenter.nethope.org/webinars/view/digital-identity-in-the-humanitarian-space

Examples of Attribute Categories

- Static vs. Dynamic
  - Static = Rarely to never changes (e.g. name, address, biometrics)
  - Dynamic = Continuously changes (e.g. physical and logical behavior)

- Self-made vs. Assigned
  - Self-made = give to yourself (e.g. nickname, tattoo)
  - Assigned = given by others (e.g. roles, employment status)



BUNDESREPUBLIK DEUTSCHLAND
FEDERAL REPUBLIC OF GERMANY / REPUBLIQUE FEDERALE D'ALLEMAGNE
PERSONALAUSWEIS
IDENTITY CARD / CARTE D'IDENTITE

T22000129

Name/Surname/Nom
MUSTERMANN
GEB. GABLER

Vornamen/Given names/Prénoms
ERIKA

Geburtstag/Date of birth/ Date de naissance
12.08.1964

Staatsangehörigkeit/Nationality/ Nationalité
DEUTSCH

Geburtsort/Place of birth/Lieu de naissance
BERLIN

Gültig bis/Date of expiry/ Date d'expiration
31.10.2020

938568

Unterschrift der Inhaberin/des Inhabers - Signature of bearer - Signature de la titulaire/du titulaire

- Verified vs. Unverified
  - Verified = other entities approve this attribute (e.g. email with signature, Name with on national id card)
  - Unverified = not approved by others

Digital Identity

- Digital representation of an entity
- One entity has multiple digital identities
  - Private: Nickname, social media account
  - Government: Everything in your ID card
  - Business: Working, studying, LinkedIn
- Digital identity could be created by the entity or by third-party (e.g. from behavior analysis)
- Each digital identity of an entity can contain different attributes

Secure Identity

- No one other than the correct entity can use its (digital) identity(-ies)
- The access and the usage of identity needs to be secured by the entity, the identity provider, and the relying party
  - Identification and authentication
  - Authorization and access control
- Insecure identity could have bad impacts
  - Misused identity for unethical and illegal activities
  - Reputation & financial loss, unauthorized data access
- Therefore there are certain things that could help secure it:

Attribute and Privacy

- Anonymity and privacy in the internet is important
- However, the attributes in the digital identity could be used to identify the entity:
  - **Identifier**: *Attribute that could directly identify the entity*

    Example: name, ID number, gait
  - **Quasi-identifier:** *Attribute that is not enough to identify the entity on its own but combined with other identifiers could identify the entity*

    Example: gender, birth date, postal code, access to a website
- The attributes could allow third party to create and analyze digital identity about the entity (shadow profile), e.g. Facebook or Amazon

Topics - Summary

- In this seminar we will focus on three fields of secure identity research
  - Analyzing Bitcoin P2P network properties
  - Enforcing access control for enterprise file sharing system
  - Authenticating the users through their behavioural aspects
- If you have some ideas or topics on how to analyze, utilize, and secure the digital identity in various scenarios, feel free to submit it to us and we could incorporate it to the seminar

Hot Topics in Secure Identity Research

Exploring the Bitcoin P2P Network

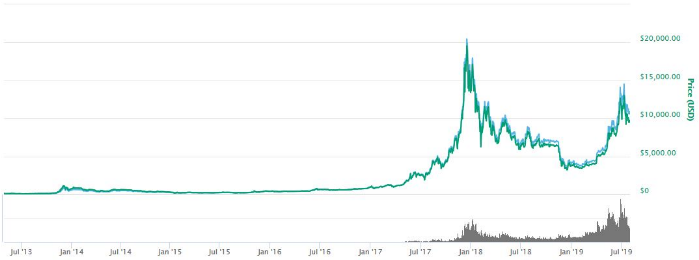
Alexander Mühle

alexander.muehle@hpi.de

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]

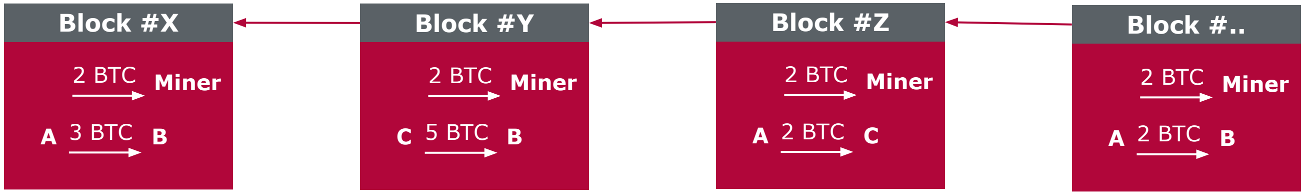
coinbase



[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.

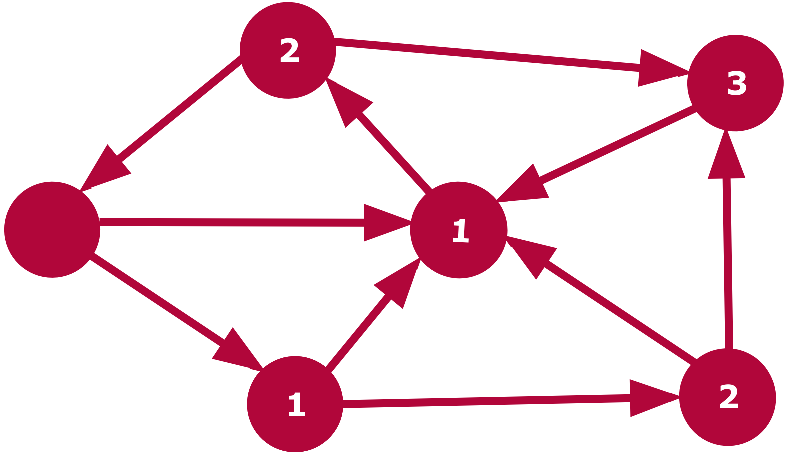
HPI – Secure Identity Lab
Bitcoin: Transactions and Blocks

- 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

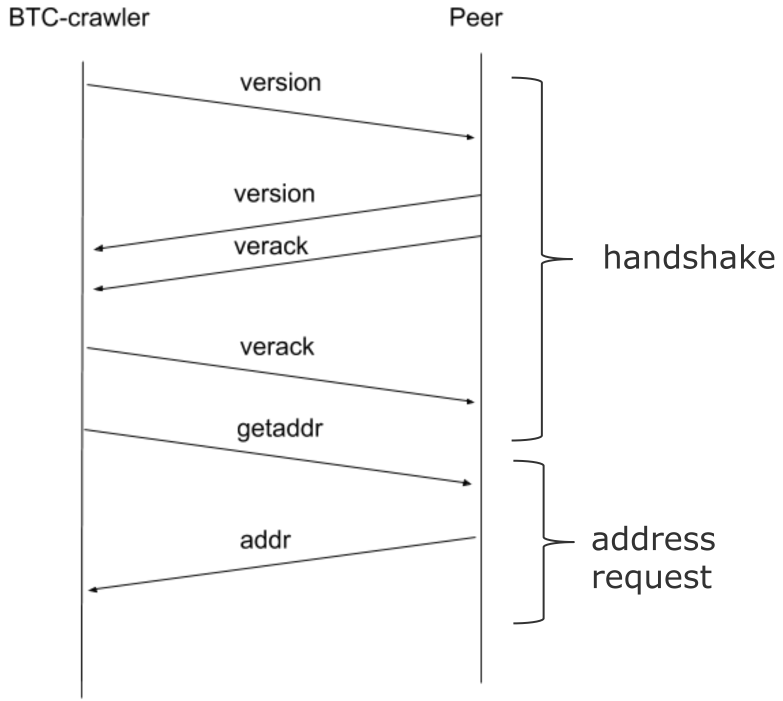


Bitcoin: Peer-to-Peer Message Exchange

- Messages are propagated like **Gossips**
- New messages are sent to one’s peers



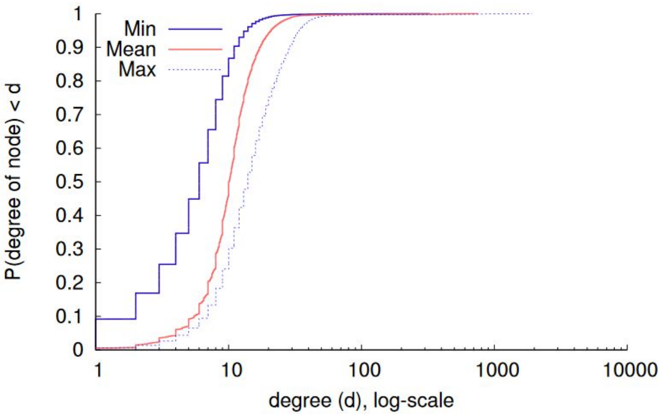
HPI – Secure Identity Lab
Bitcoin: Peer Discovery

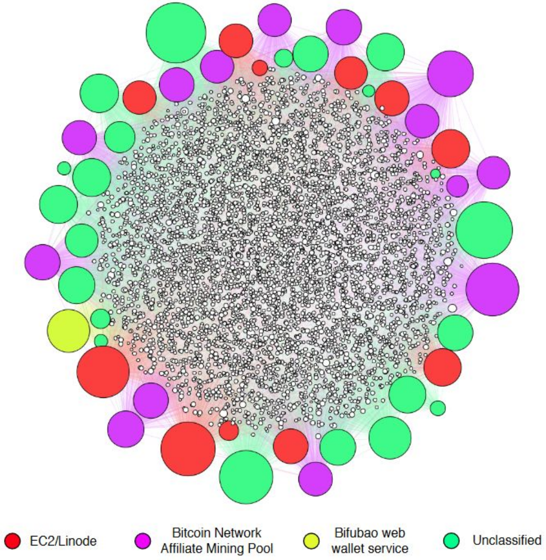


- 1000 addresses per message
- 2500 addresses per request
- Addresses are selected at random for each request
- Reference Implementation has maximum of 20480 addresses
- $(1-\frac{1}{\frac{20480}{2500}})^x$

Bitcoin: Peer-to-Peer Message Exchange

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

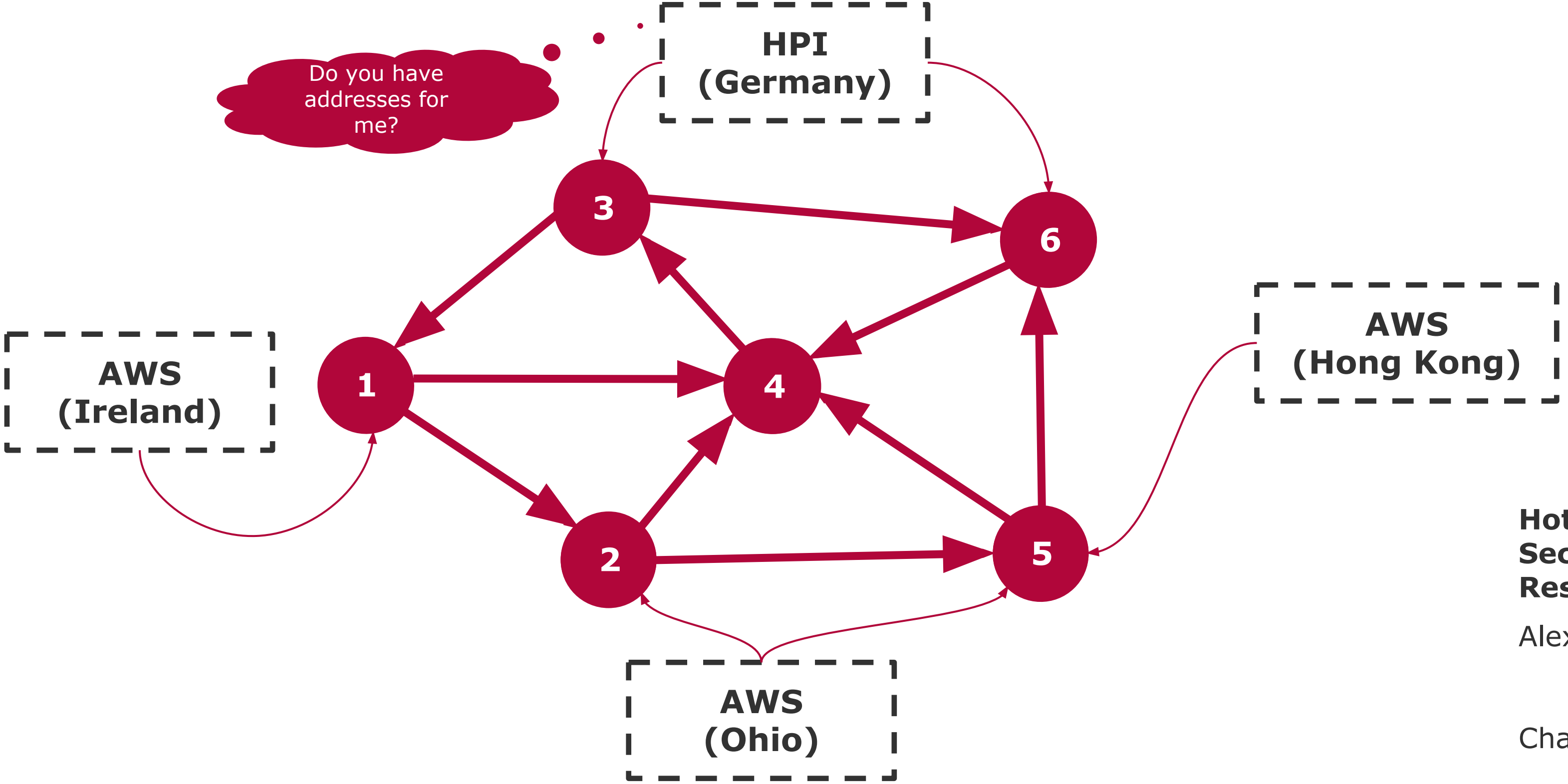



● EC2/Linode
● Bitcoin Network Affiliate Mining Pool
● Bifubao web wallet service
● Unclassified

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

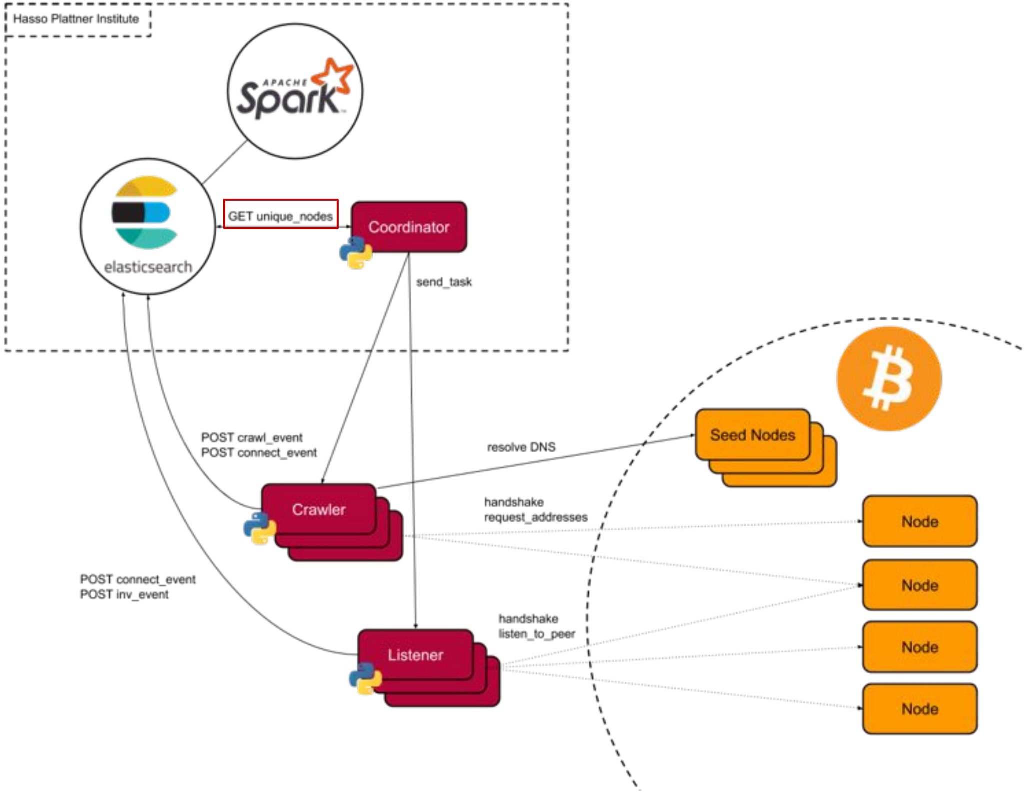
HPI – Secure Identity Lab

Project: Crawling Approach



HPI – Secure Identity Lab

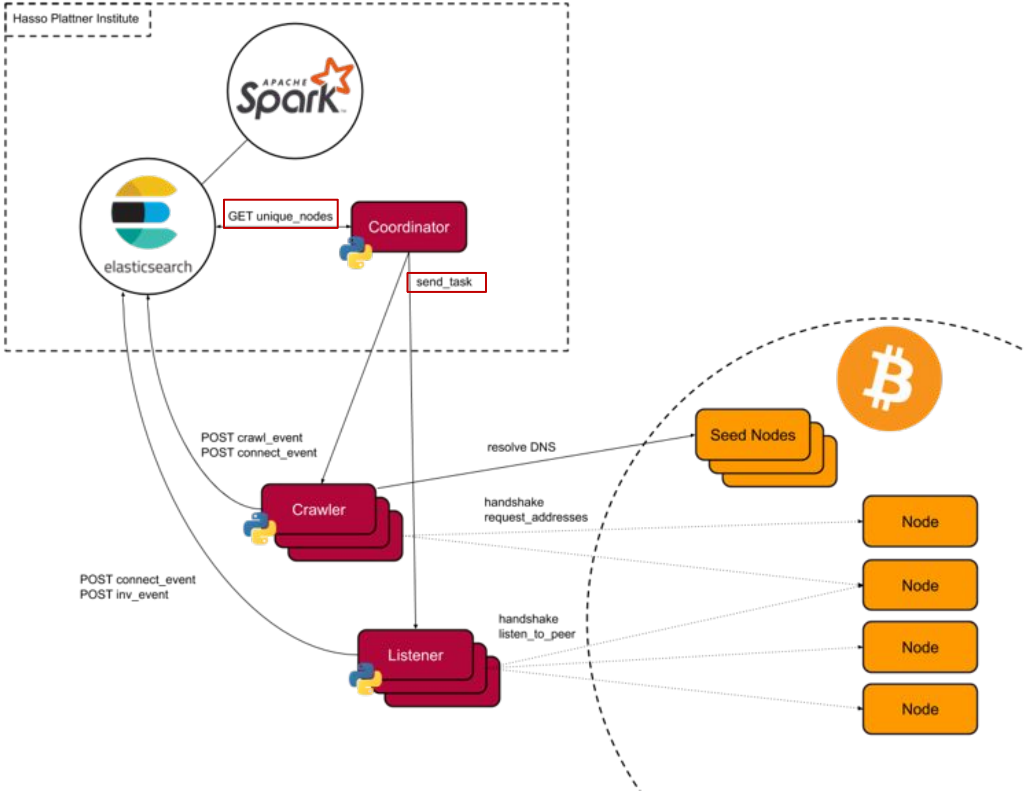
Project: Gathering Network Data



Hot Topics in Secure Identity Research

Alexander Mühle

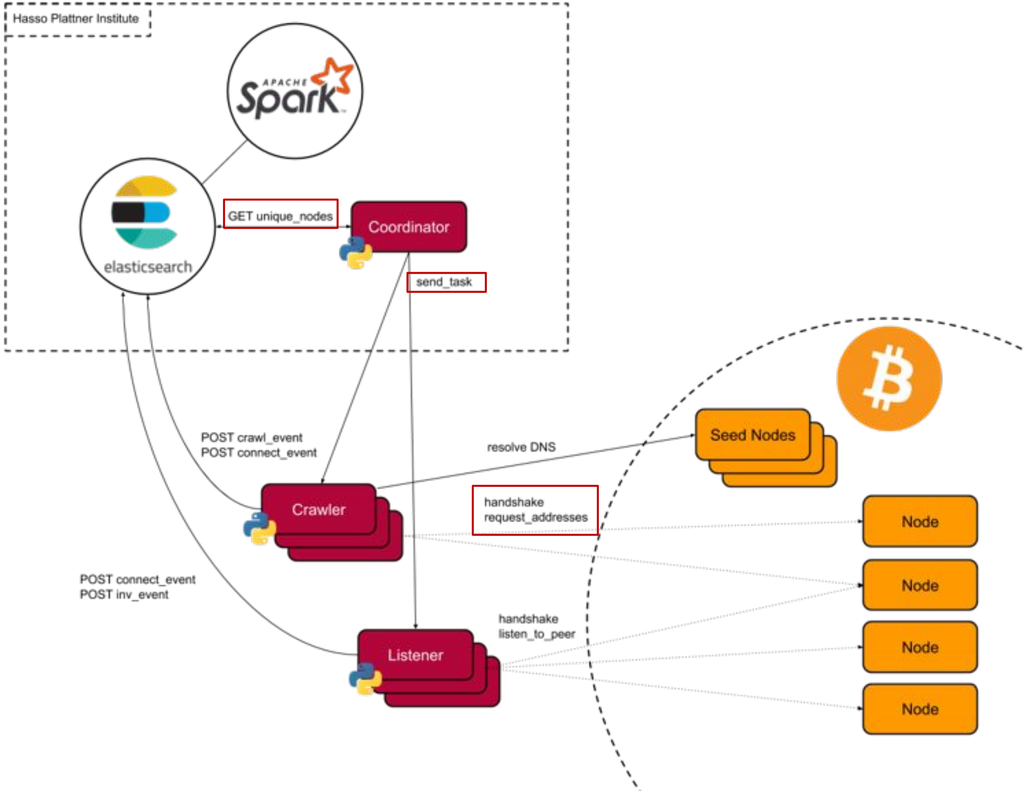
HPI – Secure Identity Lab
Project: Gathering Network Data



Hot Topics in Secure Identity Research

Alexander Mühle

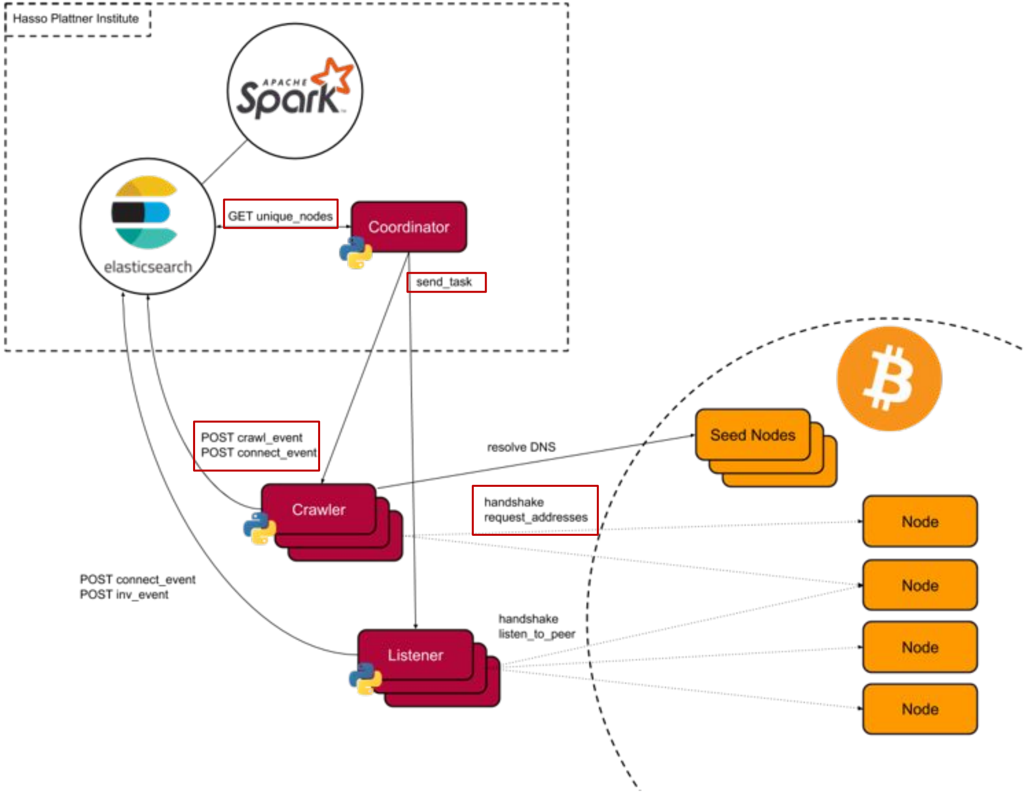
HPI – Secure Identity Lab
Project: Gathering Network Data



Hot Topics in Secure Identity Research

Alexander Mühle

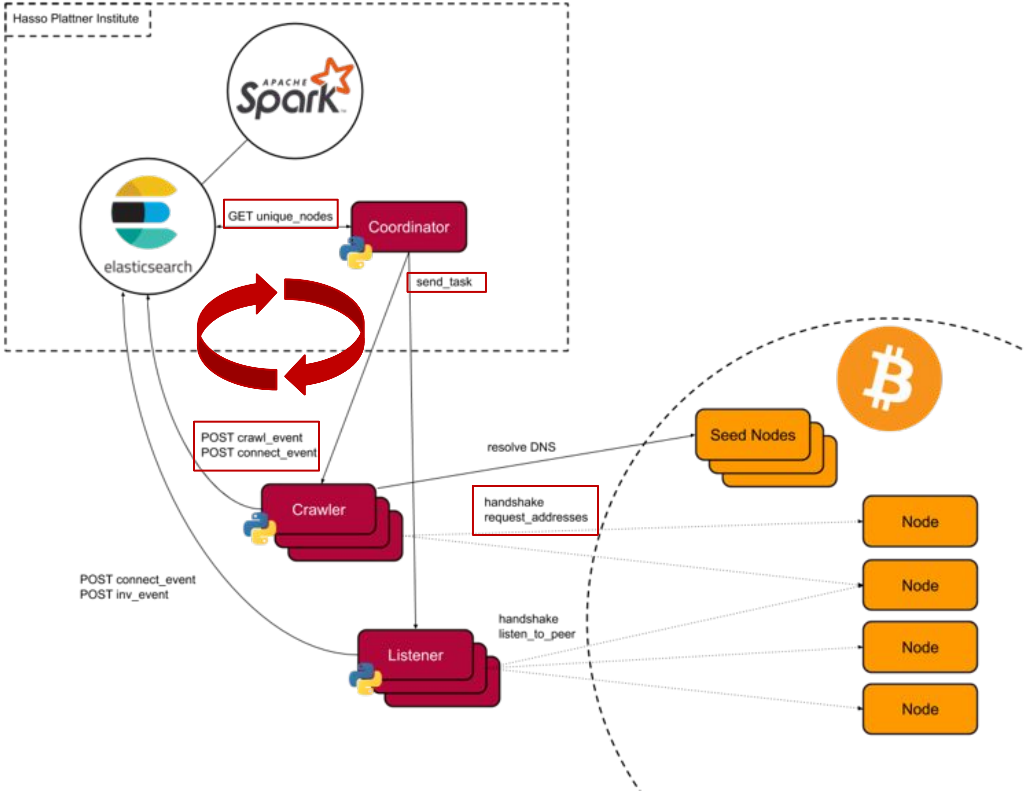
HPI – Secure Identity Lab
Project: Gathering Network Data



Hot Topics in Secure Identity Research

Alexander Mühle

Project: Gathering Network Data



HPI – Secure Identity Lab
Project: Analysing Network Data

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 crawlers) for uniqueness using Spark/Zeppelin

Further reading:

[0] Biryukov, Alex, and Ivan Pustogarov. "Bitcoin over Tor isn't a good idea." *2015 IEEE Symposium on Security and Privacy*. IEEE, 2015.

[1] Mastan, Indra Deep, and Souradyuti Paul. "A new approach to deanonymization of unreachable bitcoin nodes." *International Conference on Cryptology and Network Security*. Springer, Cham, 2017.

[2] Koshy, Philip. "CoinSeer: A telescope into bitcoin." (2013).

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

Project: Analysing Network Data

Characterise impact of Network Features

- Great Chinese Firewall
  - □ How does the Great Chinese Firewall impact miners and the Bitcoin ecosystem
  - □ Block propagation behaviour due to firewall bottleneck in peak times, content filtering…
- Analyse collected information (building on the existing network listeners)

Further reading:

[0] Decker, Christian, and Roger Wattenhofer. "Information propagation in the bitcoin network." *IEEE P2P 2013 Proceedings*. IEEE, 2013.

[1] Donet, Joan Antoni Donet, Cristina Pérez-Sola, and Jordi Herrera-Joancomartí. "The bitcoin P2P network." *International Conference on Financial Cryptography and Data Security*. Springer, Berlin, Heidelberg, 2014.

Optimise crawling strategy

- How can the crawling of the network be optimised to handle the natural churn (fluctuation of participants) of Peer-to-Peer Networks
  - □ Exponential back-off
  - □ Prioritising influential nodes
  - □ ...

Further reading:

[0] Deshpande, Varun, Hakim Badis, and Laurent George. "BTCmap: mapping bitcoin peer-to-peer network topology." *2018 IFIP/IEEE International Conference on Performance Evaluation and Modeling in Wired and Wireless Networks (PEMWN)*. IEEE, 2018.

[1] Donet, Joan Antoni Donet, Cristina Pérez-Sola, and Jordi Herrera-Joancomartí. "The bitcoin P2P network." *International Conference on Financial Cryptography and Data Security*. Springer, Berlin, Heidelberg, 2014.

Bring Your Own Ideas

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

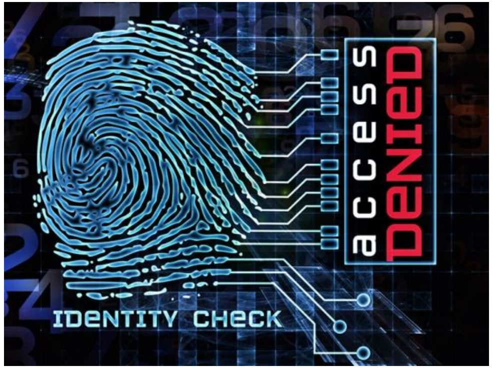
Hot Topics in Secure Identity Research

Enterprise File Access Control

Muhammad Ihsan Haikal Sukmana
muhammad.sukmana@hpi.de

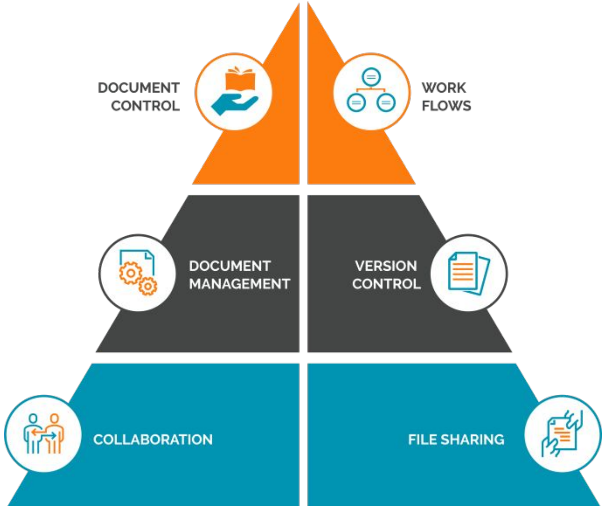
Authorization vs Access Control

- After an entity is authenticated, the system needs to ensure that the identity can only do certain actions to certain objects in the system
- A system needs to provide authorization and access control to ensure it
  - *Authorization*: Dictates what the identity can do in the system e.g. A Facebook user could not access other user‘s profile
  - *Access control*: The method to enforce the authorization policy e.g. Facebook‘s Settings option to deny access to non-friend user
- If a user does not have the correct privileges, the user is not authorized to do the actions in the system



Enterprise File Access Control

- Imagine that you are an enterprise file sharing solution system to provide secure file sharing for multiple companies
- File access control is important in the enterprise scenario
  - Ensure only authorized user could access the file
  - Should follow the enterprise`s hierarchical structure, such as the departments and employee`s roles
- If file access control is not properly configured and enforced, this could lead to confidential file unauthorized access and leakage that affect the company reputation or monetary loss



Topic Options

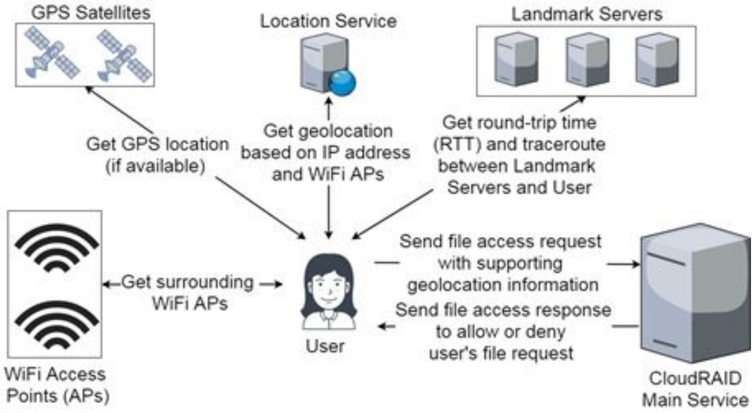
- We focus on improving the authorization and access control aspects for enterprise file sharing solution:
  - Location-based access control to ensure certain resource can be accessed in a particular location/region
  - Improving the flexibility and access control for secure inter-company file sharing system
  - Providing privacy-preserving access control and signature in the enterprise domain
  - Bring your own topics in the area of enterprise access control and authorization

Topic 1
Location-based Access Control

- Nowadays people can work anytime and anywhere thanks to Internet, especially right now due to Coronavirus
- File access control is important for secure file sharing but it is hard to enforce file access control where employees are not connected to office network
- Physical access control is still important to ensure that the file is accessed from a secure location/region
- Example: employee can only access the file in the certain region (e.g. only in Germany but not in China) with company can keep track the request and enforce necessary access control

Topic 1
Location-based Access Control

- Develop file access control model using device‘s geolocation information from GPS and Internet
  - Determine user‘s location using delay-based measurement techniques by utilizing the ping between the landmark servers and the client
  - Using multiple VMs available around the world
  - Improve the current algorithm to detect if user has fake geolocation using VPN, proxy, or fake GPS position
- Design and prototype implementation of the system


GPS Satellites
Location Service
Landmark Servers
Get GPS location (if available)
Get geolocation based on IP address and WiFi APs
Get round-trip time (RTT) and traceroute between Landmark Servers and User
Get surrounding WiFi APs
User
Send file access request with supporting geolocation information
Send file access response to allow or deny user's file request
WiFi Access Points (APs)
CloudRAID Main Service

Hot Topics in Secure Identity Research

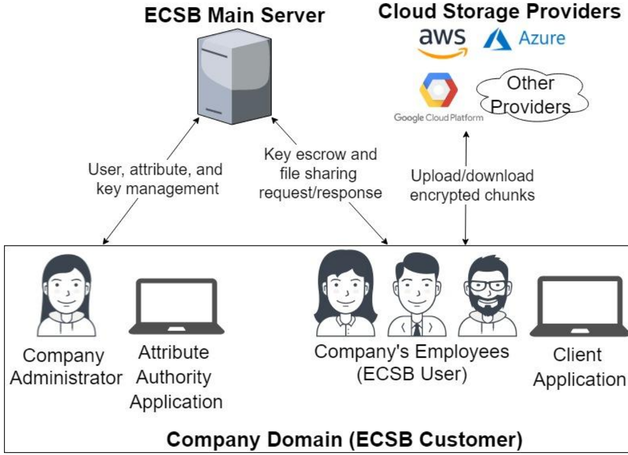
Muhammad Sukmana

Topic 1
Location-based Access Control

- Prerequisites & System Environment
  - Java/Python network programming language
  - Client & server architecture
  - Linux VMs from Amazon EC2
- References
  - https://link.springer.com/chapter/10.1007/978-3-030-15032-7_104
  - https://dl.acm.org/doi/pdf/10.1145/1028788.1028828
  - https://crpit.scem.westernsydney.edu.au/confpapers/CRPITV102Arif.pdf

Topic 2
Secure Inter-Company File Sharing

- Enterprise file sharing solution, such as Enterprise Cloud Storage Brokerage (ECSB), might need to provide secure file sharing between companies for collaboration
- However, some challenges need to be faced by ECSB in order to provide secure and efficient file sharing
  - □ Access control: ECSB needs to provide organizational-based access control to avoid cross-company data leakage
  - □ Flexibility: The file sharing in the system should be flexible following user`s file access restriction specification


ECSB Main Server
Cloud Storage Providers
aws
Azure
Google Cloud Platform
Other Providers
User, attribute, and key management
Key escrow and file sharing request/response
Upload/download encrypted chunks
Company Administrator
Attribute Authority Application
Company's Employees (ECSB User)
Client Application
Company Domain (ECSB Customer)

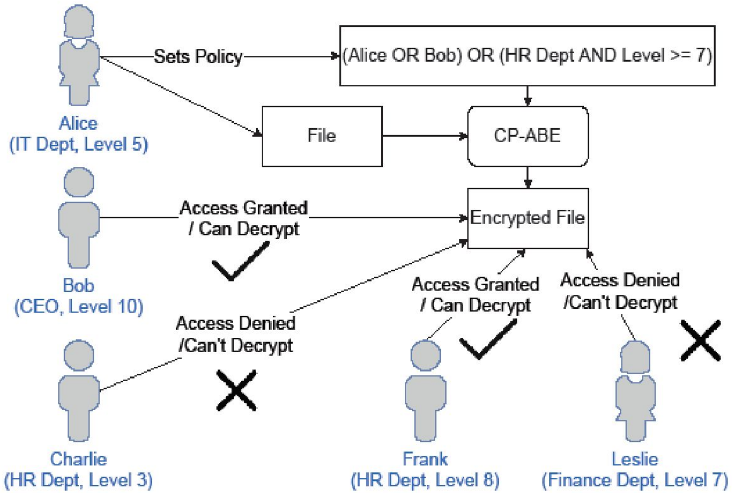
Hot Topics in Secure Identity Research

Muhammad Sukmana

Topic 2

Secure Inter-Company File Sharing

- We have developed a proof-of-concept ECSB based on an attribute-based encryption (ABE) scheme
- ABE is a new and experimental public key cryptography scheme that utilizes attributes as the keypair and the policy containing the attributes and logic gates
- Only authorized user with the correct attributes that fulfill the policy could decrypt the ciphertext
- The scheme has the flavour of multi-authority ciphertext-based policy ABE (CP-ABE)



Sets Policy
(Alice OR Bob) OR (HR Dept AND Level >= 7)
Alice
(IT Dept, Level 5)
File
CP-ABE
Encrypted File
Access Granted / Can Decrypt
Bob
(CEO, Level 10)
Access Denied /Can't Decrypt
Access Granted / Can Decrypt
Access Denied /Can't Decrypt
Charlie
(HR Dept, Level 3)
Frank
(HR Dept, Level 8)
Leslie
(Finance Dept, Level 7)

Muhammad Sukmana

Topic 2
Secure Inter-Company File Sharing

- Improve the file access control and file sharing functionality of our proof-of-concept enterprise file sharing system
  - Provide system-level organizational-based access control for each company while allowing the company to access some files of other companies
  - Improve on the policy structure of the ABE scheme to ensure flexible file sharing in the system, such as linear secret sharing scheme or non-monotonic access policy
- Design and prototype implementation of the system

Topic 2
Secure Inter-Company File Sharing

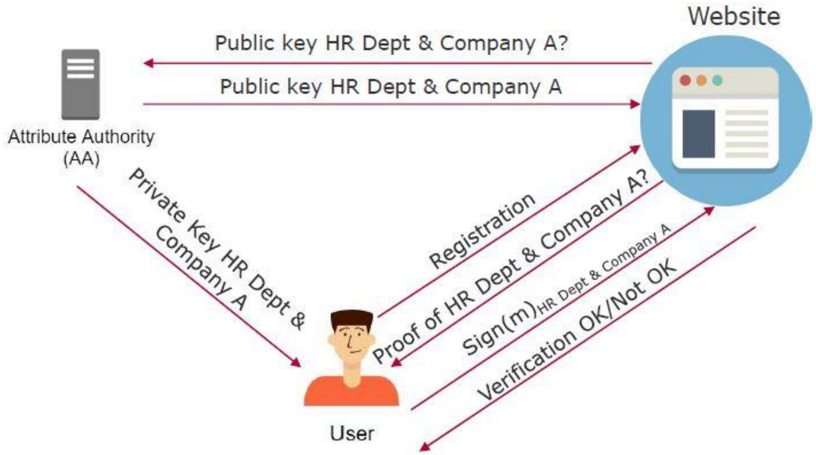
- Pre-requisites & System Environment
  - Java programming language
  - Docker
  - Cryptography and mathematics background (not necessary but appreciated)
- References
  - https://github.com/Anroc/PAD-TFDAC-MACS
  - https://eprint.iacr.org/2010/374.pdf
  - http://www-master.ufr-info-p6.jussieu.fr/2005/IMG/kordon/grar/sv-5.pdf

Topic 3
Privacy-Preserving Access Control and Signature

- In general, access control will require the information of the identity used by the entity in a certain system environment
- An entity could claim that it is authorized to do certain action in the system due to its characteristics (proof of possession) but it doesn't want to be known by other users
- However, the system or the attacker might be able to get or learn the information of the identity that could lead to unwanted scenarios, such as identity theft
- How would the system attest to the claim of the entity to enforce access control while preserving the privacy of the entity?

Topic 3
Privacy-Preserving Access Control and Signature

- One of the possible solutions is attribute-based signature (ABS)
- It utilizes the attributes to create the signature using the private key where other entities could attest to the claim by verifying the signature using the public key
- **Unforgeable**: Signature could not be forged against chosen predicate and message attacks
- **Perfect privacy**: Signature should not reveal nothing about the identity or attributes of the signer beyond the information in the claim


Website
Public key HR Dept & Company A?
Public key HR Dept & Company A
Attribute Authority (AA)
Private Key HR Dept & Company A
Registration
Proof of HR Dept & Company A?
Sign(m)$_{\text{HR Dept \& Company A}}$
Verification OK/Not OK
User

Topic 3
Privacy-Preserving Access Control and Signature

- Develop a proof-of-concept digital signature functionality for enterprise file-sharing system
  - A digital signature is generated once a file is created by the data owner
  - User could check if the file is generated by the authorized data owners responsible to provide non-repudiation
  - Force access control based on the signature
- Design and prototype implementation of the system

Topic 3
Privacy-Preserving Access Control and Signature

- Pre-requisites & System Environment
  - Java/Python programming language
  - Docker
  - Cryptography and mathematics backrgound (not necessary but appreciated)
- References
  - https://dl.acm.org/doi/pdf/10.1145/1755688.1755697
  - https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8279429

Hot Topics in Secure Identity Research

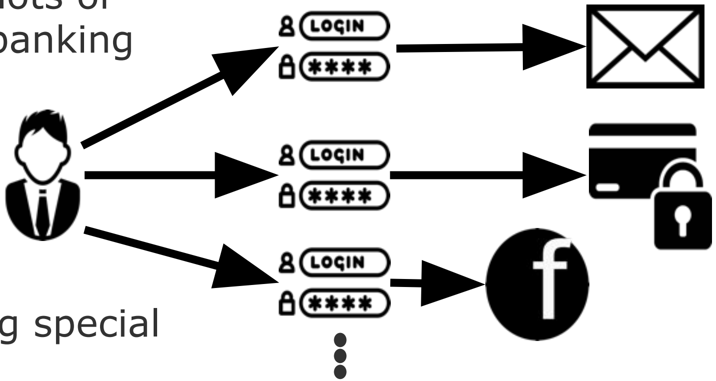
Behavioral Authentication

Eric Klieme & Christian Tietz
eric.klieme, christian.tietz@hpi.de

Traditional Username/Password authentication may not be the perfect solution for todays internet service usage

Problem

- A user has 25++ passwords on average and uses lots of different services in a range from social media to banking applications



Solution (in theory)

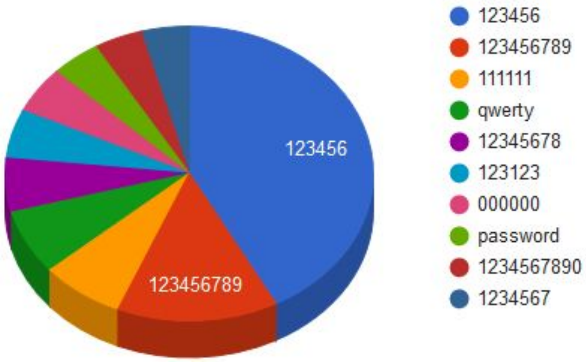
- Different password for every service
- Each password of a certain length, maybe including special letters
- Only remembered, not written down anywhere

Solution (assumed)

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

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 ~ 10.5 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/

Although there exist alternatives to „knowledge-based“ authentication methods based on objects or biometrics, most web services stick to username/password

Ways to identify a user

- By something he knows (knowledge-based)
- By something he carries or possesses (object-based)
- By personal characteristics (biometrics-based)

- More than one method/approach for one authentication process is referred to as multi-factor authentication

Web services used in daily life only apply knowlegde-based approach

- Hard to remember, simple passwords can be guessed/reengineered
- Once logged in, further authentication rarely implemented
- Once leaked, (multiple) services can be misused for fraud or worse

Object-based authentication is often part of multi-factor authentication but not that user-friendly

- Typical authentication method to access buildings, offices, flats etc.

Advantages

- Needs no remembering
- „Show and access"

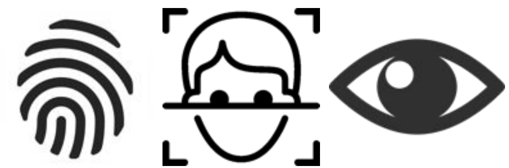
Disadvantages

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

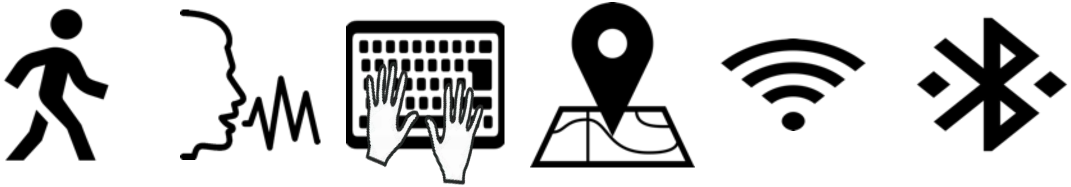
- Bank transactions or VPN access is often protected by multi-factor authentication with object-based (smartphone/token) and knowlegde-based (username/password) methods

Biometrics-based authentication methods include physical and behavioral aspects but collects very private data

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

Topic Presentation

Christian Tietz &
Eric Klieme

There are several requirements for an authentication mechanism to be a “good one”

- Easy-to-use
- Easy-to-remember
- Hard-to-guess/imitate
- Leak/Theft not possible
- Protects privacy
- Resistance against several attacks
- Approves identity in a continuous way

…to be continued

Username/password

Mail: foobar / Tm}_c5m2FgPjK%..

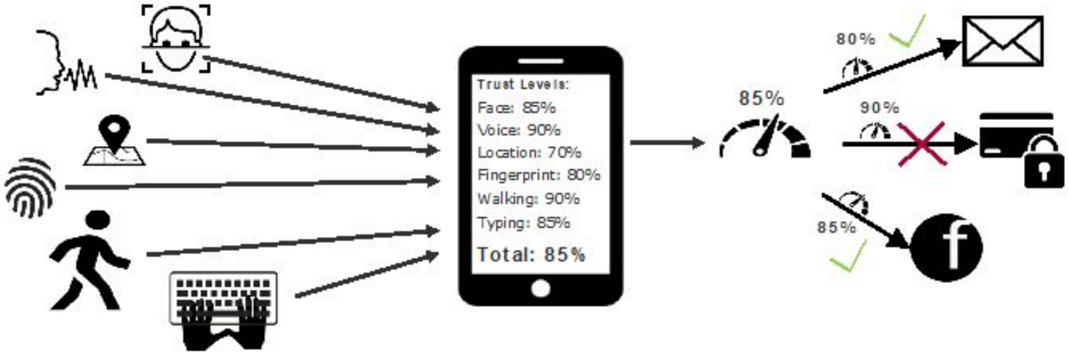
Facebook: foobar / dfTCbTV9%"6Dq&93

Topic Presentation

Christian Tietz &
Eric Klieme

Vision: Analyze user‘s behaviour continously in the background to build a *Trust Level* for (web) authentication

- Access sensors on devices the users already possess (smartphone, wearables) to apply different behavioural biometrics
- Aggregate values and build a trust level that can be used by services
- Due to the computational power of current and upcoming devices all calculations are done on the devices themselves and only the calculated trust level leaves the phone



Topic Presentation

Christian Tietz &
Eric Klieme

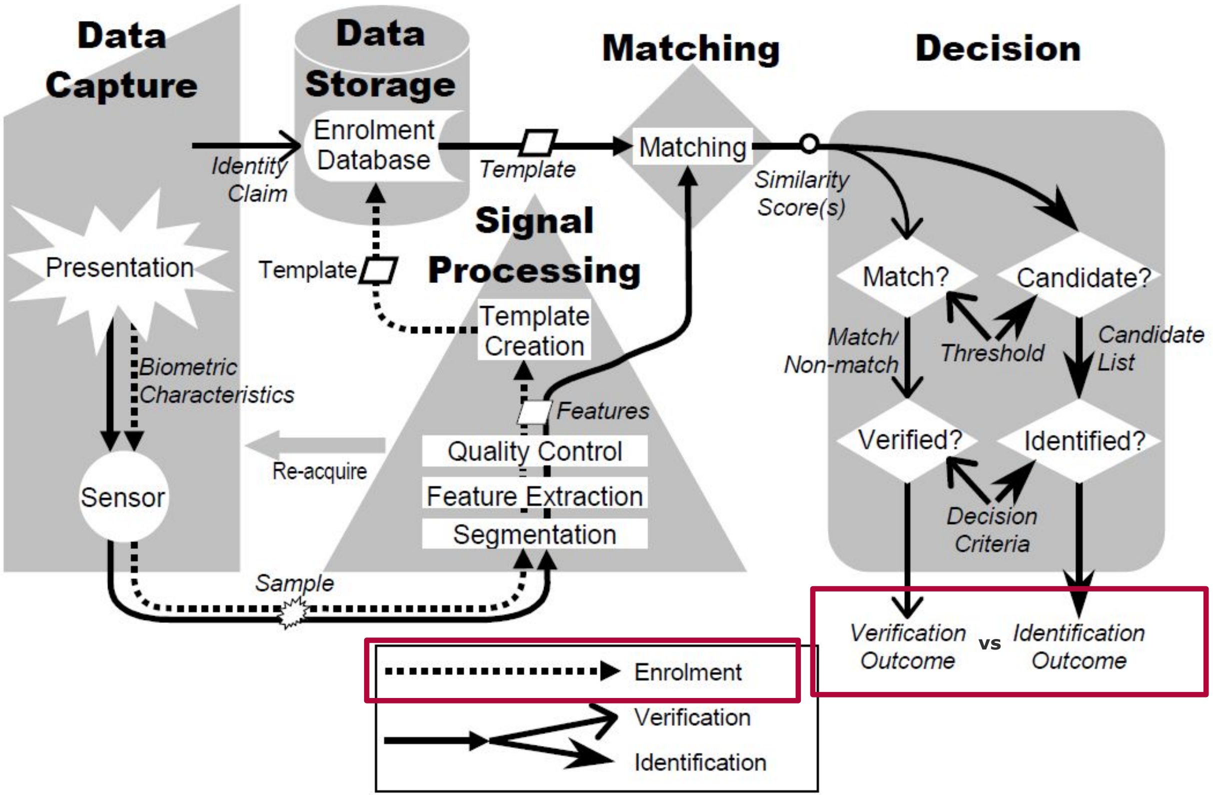
Trust level based behavioral authentication can provide solutions to many of the requirements

- **Easy-to-use** => Users need no additional interaction with services except allowing the transmission of the trust level. No additional tokens need to be brought, no passwords need to be remembered
- **Easy-to-remember** => User just have to walk, type etc. as usual.

- **Hard-to-guess/imitate** => Research showed that attacks on walking styles or typing behaviour are very difficult
- **Leak not possible** => Compared to traditional passwords no hash is stored. The recognition of the user is done on the phone and only the trust level is transmitted
- **Protects privacy** => Due to the strong computational power of the devices no private data has to leave the phone

- **Resistance against several attacks** => State-Of-The-Art Identity Management Protocols for Enrollment and Trust Level Submission to Web Service
- **Approves identity in a continuous way** => Sensor data like accelerometer data is provided constantly. Integrating different inputs like walking, bluetooth, wifi or typing at every point of time an authentication is possible

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

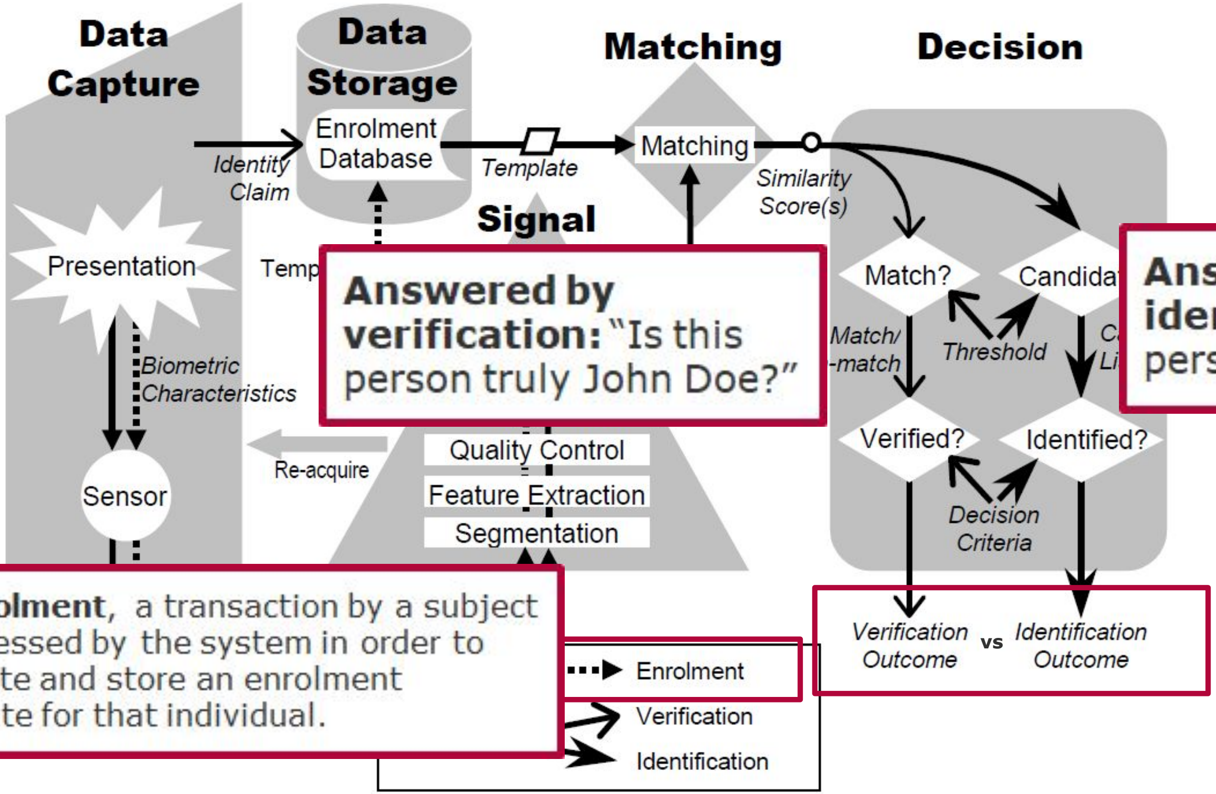


ISO 19795

Topic Presentation

Christian Tietz &
Eric Klieme

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



Answered by verification: “Is this person truly John Doe?”

Answered by identification: “Is this person in the database?”

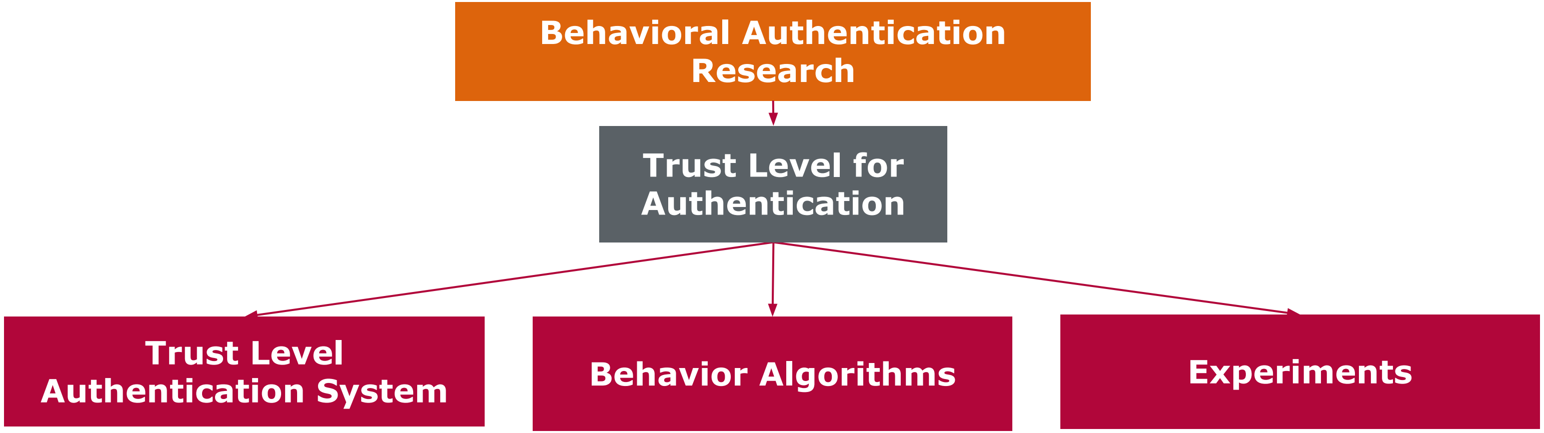
In **enrolment**, a transaction by a subject is processed by the system in order to generate and store an enrolment template for that individual.

ISO 19795

Topic Presentation

Christian Tietz &
Eric Klieme

Overview Research Area and Questions



Behavioral Authentication Research

Trust Level for Authentication

Trust Level Authentication System

Behavior Algorithms

Experiments

Trust Level Authentication System

- How can a trust level be calculated?
- How can a trust level be used for authentication and authorization?
- Which architectures are possible?

Behavior Algorithms

- Which behavioral data can be used to verify a persons identity?
- Which algorithms are suitable for the different areas of behavior?
- How can specific algorithms be optimized?

Experiments

- How can I effectively collect data in experiments?
- How can I verify my own approaches in the real world?
- What other datasets exists and how can I use them?

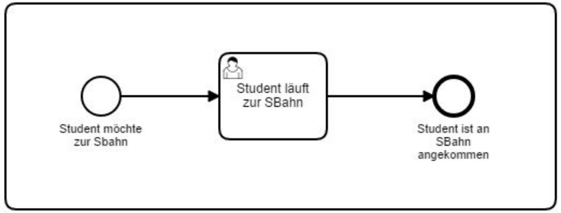
Project 1: **Experiments**
Integrate users into labeling

- 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
  - Proven from psychology research that they tell the truth

Why do we need to collect data?

Which data do we need?



1. Context: Current activity is walking

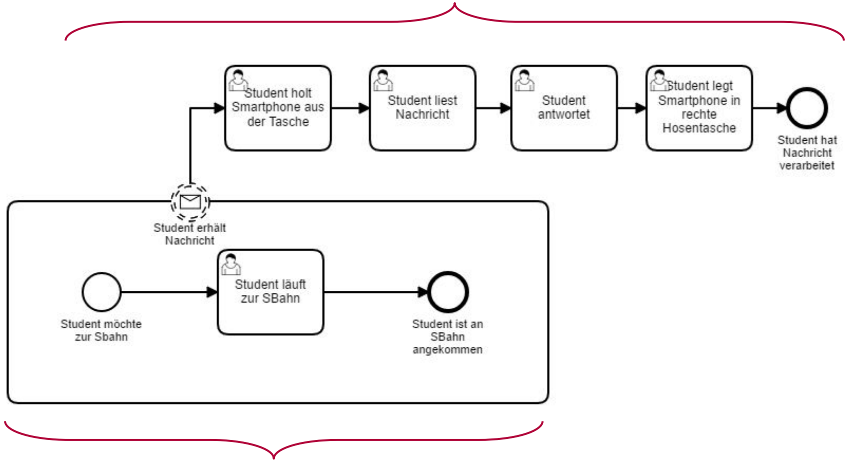
Topic Presentation

Christian Tietz &
Eric Klieme

The collection of data for improving behavioural authentication is difficult

Which data do we need?

2. Actions during the activity: During that activity I did this and that using the app on my smartphone after I pulled it out of my pocket


Student holt Smartphone aus der Tasche
Student liest Nachricht
Student antwortet
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

1. Context: Current activity is walking

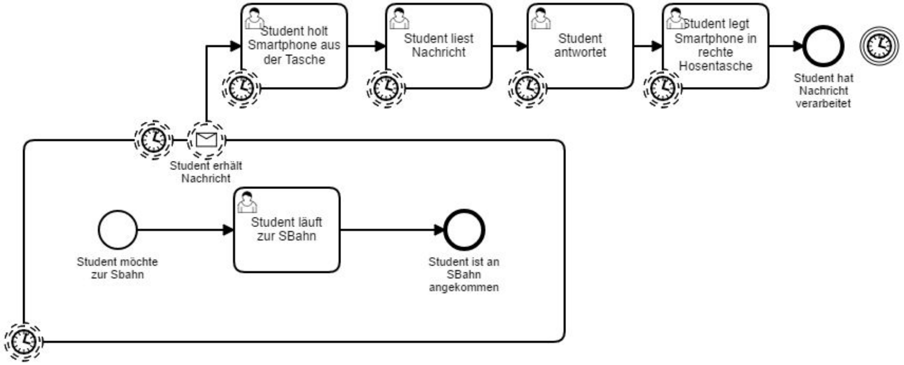
Topic Presentation

Christian Tietz & Eric Klieme

The collection of data for improving behavioural authentication is difficult

Which data do we need?

2. Actions during the activity: During that activity I did this and that using the app on my smartphone after I pulled it out of my pocket

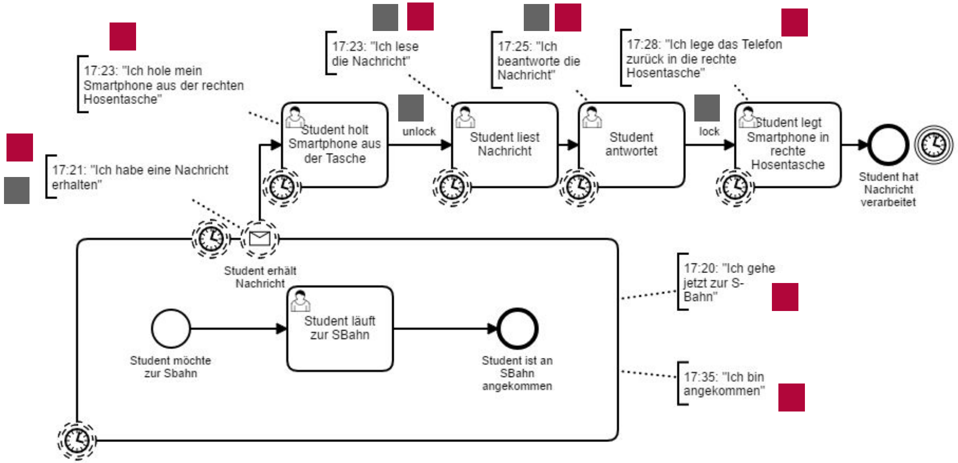

Student holt Smartphone aus der Tasche
Student liest Nachricht
Student antwortet
Student legt Smartphone in rechte Hosentasche
Student hat Nachricht verarbeitet
Student erhält Nachricht
Student möchte zur Sbahn
Student läuft zur SBahn
Student ist an SBahn angekommen

1. Context: Current activity is walking

3. Ordered, usually by timestamp

The collection of data for improving behavioural authentication is difficult

Information sources can be accessed only in a very limited way


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

No access possible (subconciousness)

🟥 Student, implicit

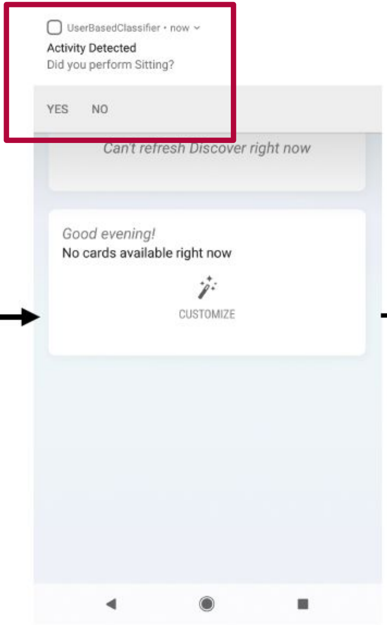
⬛ App, explicit

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?

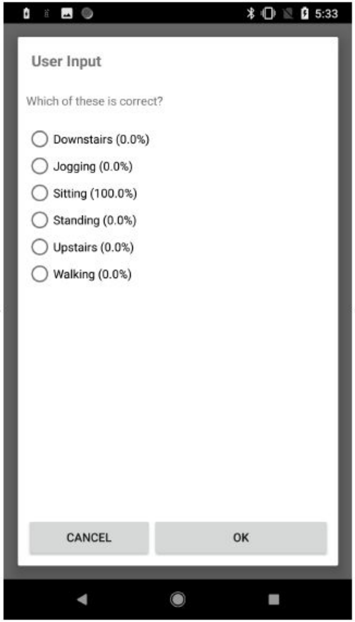
YES NO

Can't refresh Discover right now

Good evening!

No cards available right now

CUSTOMIZE



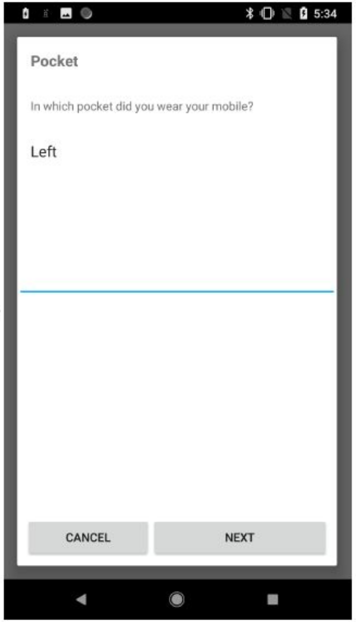
5:33

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



5:34

Pocket

In which pocket did you wear your mobile?

Left

CANCEL NEXT

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.

Correction

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

Context

User gives context information.

Project 1: 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

* https://awareframework.com/

Topic Presentation

Christian Tietz &
Eric Klieme

Project 2: **Behavior Algorithms**
Activities while Walking require robust activity detection and user verification systems

- Human style of walking – *gait* – is well researched
  - Activity detection
  - User identification and verification

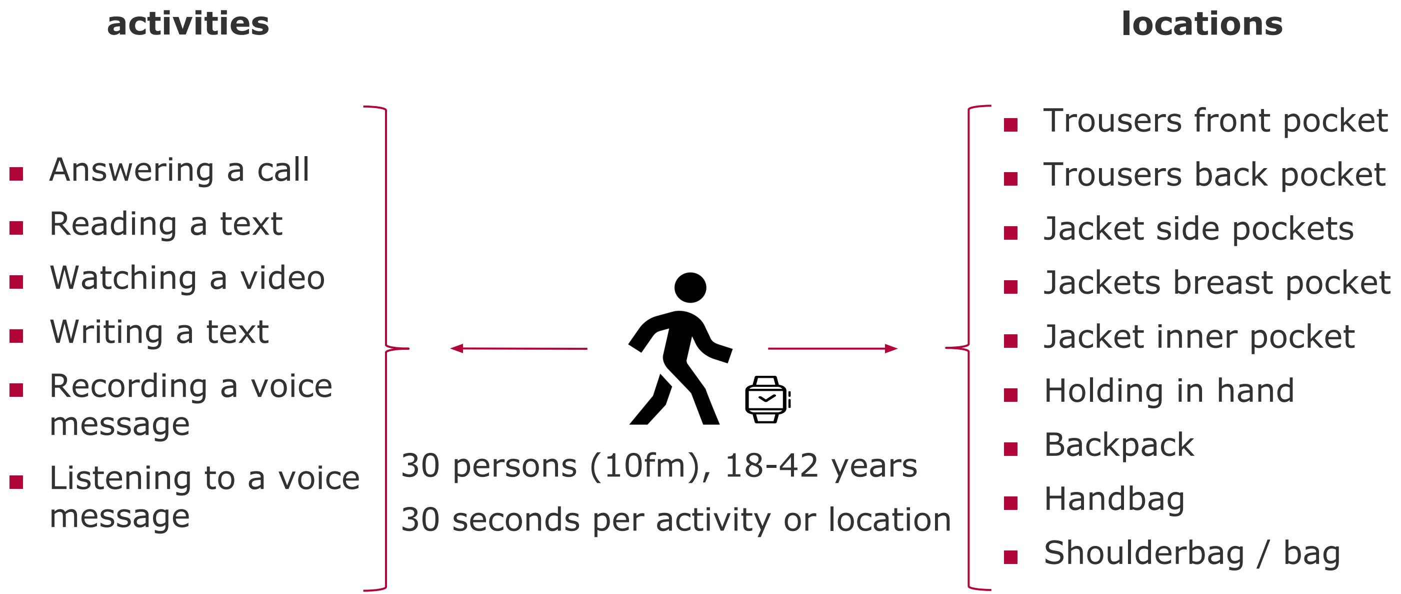
- SMOMBIELAND: People do a lot of different activities with the smartphone while walking ⇒ **S**martphone Z**ombie**

- ☐ Additional robustness requirements for any mechanism
  - Detect different activities while walking
  - Identify and verify users (dis)regarding the current activity

Topic Presentation

Christian Tietz &
Eric Klieme

No public data set available, so data was collected in a realistic scenario already



activities

- Answering a call
- Reading a text
- Watching a video
- Writing a text
- Recording a voice message
- Listening to a voice message

30 persons (10fm), 18-42 years

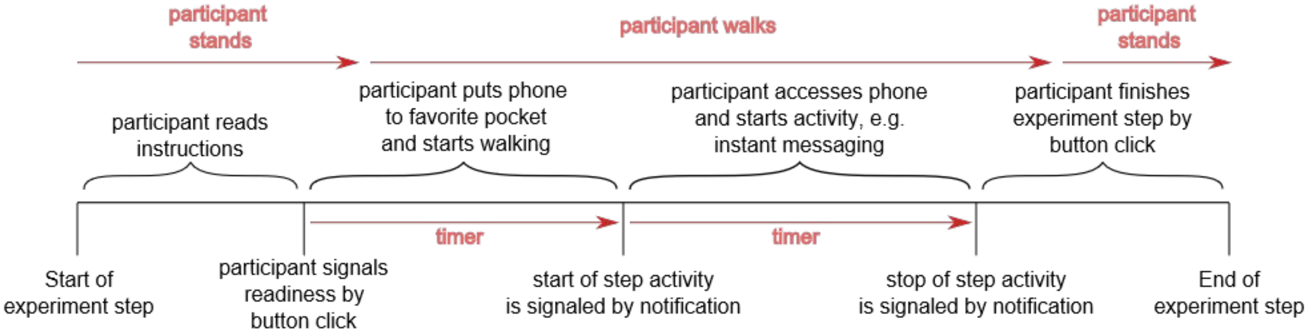
30 seconds per activity or location

locations

- Trousers front pocket
- Trousers back pocket
- Jacket side pockets
- Jackets breast pocket
- Jacket inner pocket
- Holding in hand
- Backpack
- Handbag
- Shoulderbag / bag

- Google Pixel Smartphone and Huawei Watch
- Specific collection app that recorded sensor data with 400Hz

Participants collected data by themselves in a semi-supervised recording scenario reducing bias



participant stands

participant walks

participant stands

participant reads instructions

participant puts phone to favorite pocket and starts walking

participant accesses phone and starts activity, e.g. instant messaging

participant finishes experiment step by button click

timer

timer

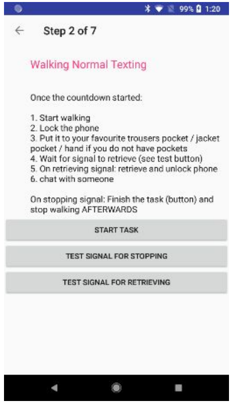
Start of experiment step

participant signals readiness by button click

start of step activity is signaled by notification

stop of step activity is signaled by notification

End of experiment step

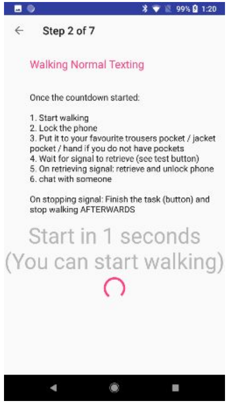


Walking Normal Texting

Once the countdown started:

1. Start walking
2. Lock the phone
3. Put it to your favourite trousers pocket / jacket pocket / hand if you do not have pockets
4. Wait for signal to retrieve (see test button)
5. On retrieving signal: retrieve and unlock phone
6. chat with someone

On stopping signal: Finish the task (button) and stop walking AFTERWARDS



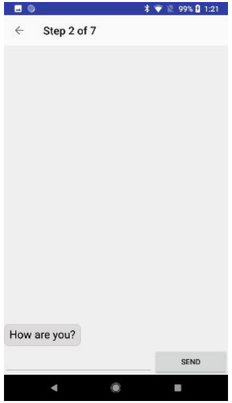
Walking Normal Texting

Once the countdown started:

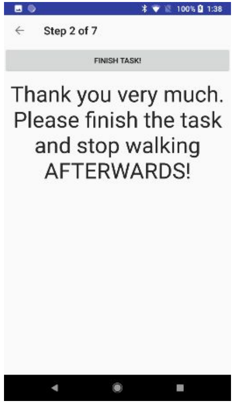
1. Start walking
2. Lock the phone
3. Put it to your favourite trousers pocket / jacket pocket / hand if you do not have pockets
4. Wait for signal to retrieve (see test button)
5. On retrieving signal: retrieve and unlock phone
6. chat with someone

On stopping signal: Finish the task (button) and stop walking AFTERWARDS

Start in 1 seconds
(You can start walking)



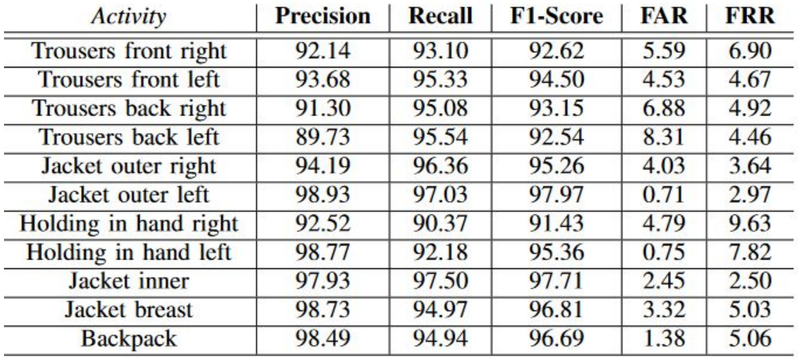
How are you?



Thank you very much. Please finish the task and stop walking AFTERWARDS!

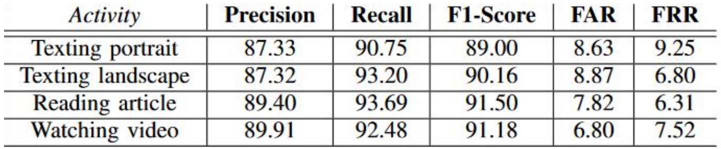
Accelerometer-only SVM performs best on a three-fold activity / location split with promising results for identity verification

Wearing only

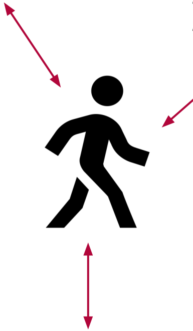


Activity	**Precision**	**Recall**	**F1-Score**	**FAR**	**FRR**
Trousers front right	92.14	93.10	92.62	5.59	6.90
Trousers front left	93.68	95.33	94.50	4.53	4.67
Trousers back right	91.30	95.08	93.15	6.88	4.92
Trousers back left	89.73	95.54	92.54	8.31	4.46
Jacket outer right	94.19	96.36	95.26	4.03	3.64
Jacket outer left	98.93	97.03	97.97	0.71	2.97
Holding in hand right	92.52	90.37	91.43	4.79	9.63
Holding in hand left	98.77	92.18	95.36	0.75	7.82
Jacket inner	97.93	97.50	97.71	2.45	2.50
Jacket breast	98.73	94.97	96.81	3.32	5.03
Backpack	98.49	94.94	96.69	1.38	5.06

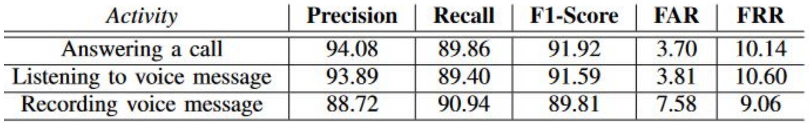
Screen attention



Activity	**Precision**	**Recall**	**F1-Score**	**FAR**	**FRR**
Texting portrait	87.33	90.75	89.00	8.63	9.25
Texting landscape	87.32	93.20	90.16	8.87	6.80
Reading article	89.40	93.69	91.50	7.82	6.31
Watching video	89.91	92.48	91.18	6.80	7.52



speech related



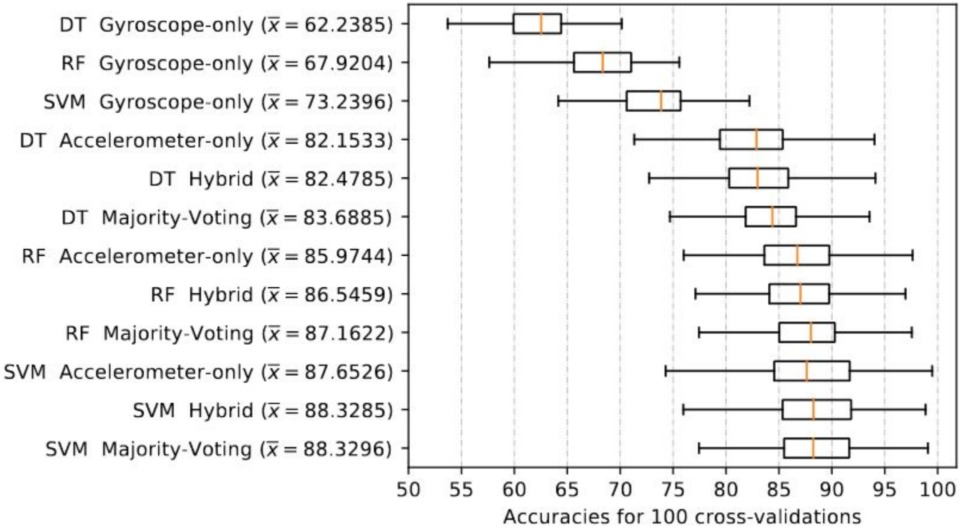
Activity	**Precision**	**Recall**	**F1-Score**	**FAR**	**FRR**
Answering a call	94.08	89.86	91.92	3.70	10.14
Listening to voice message	93.89	89.40	91.59	3.81	10.60
Recording voice message	88.72	90.94	89.81	7.58	9.06

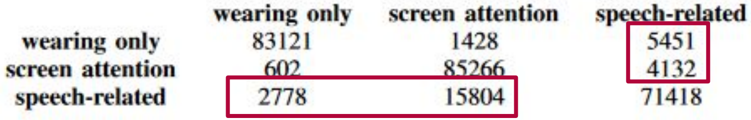
Topic Presentation

Christian Tietz &
Eric Klieme

For a real world deployment **subset detection** is sufficiently possible with a supervised SVM

- Prior to verification, we need to detect which activity is executed
- Different strategies for voice messages complicate detection
  - Recording / listening like phone call
  - Recording / listening like reading a text
  - Listening with the device speakers pointing to the ear





	wearing only	screen attention	speech-related
wearing only	83121	1428	5451
screen attention	602	85266	4132
speech-related	2778	15804	71418

confusion matrix

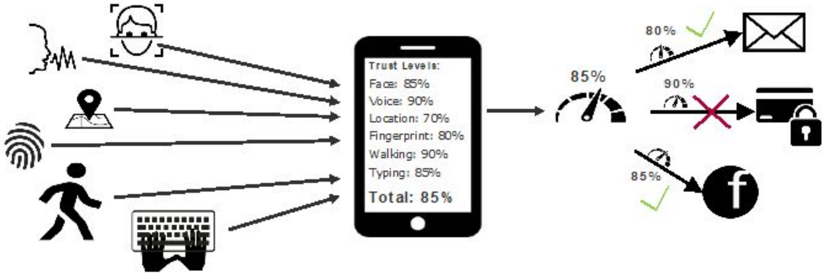
Project 2: Improving Human Activity Detection

- Question to Answer: Can we detect fine-grained human activities executed while walking with the smartphone?
  - Which pocket?
  - What specific activity (writing, watching etc.)
- Dataset: SMOMBIE data // Collect new data

- Contribution: „Specific improvement‟
  - Related Work analysis
  - Feature Engineering / Classifier Engineering / Experiments
  - Evaluation of different ML approaches / decision scheme
  - Basic Pipeline exists
- Nice-To-Have Skills
  - Python, Statistics, Machine Learning, Deep Learning
- Reference: https://ieeexplore.ieee.org/document/8455964

Project 3: Trust Level Network Performance

- Our Idea
  - Smartphone computes a trust level and
  - Sends it regularly to all logged-in web services
- This influences the battery life and network load



- Your Task: Measure the trust level performance
  - battery consumption, network speed and load, etc. of
  - 1 Smartphone send to 5 web services
  - 1 Smartphone send to 50 web services
  - 1 Smartphone send to 1 broker which distribute to web services
  - ...
- Skills learned:
  - Android, Kotlin/Java

Topic Presentation

Christian Tietz &
Eric Klieme

Behavioral Authentication

Bring Your Own Ideas

Do you have other interesting ideas on what to do with collected SMOMBIE data or in the field of behavioral authentication in general?

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:*
  - 28.04 - 31.07.2020 (~13 weeks)
  - presentation at the end of that time, documentation deadline shortly after
- 180h/13 weeks => 14h work a week per student
  - ~ **1,5 days of working** for the seminar only **per week**
  - A group of four students ~ 56h (7 PD) **per week**
- Although calculation mostly holds theoretically, rule of thumb for our expectations during progress meetings

Timeline (approx)

16.04.2020 Official first lecture / meeting, Q & A session

22.04.2020 Submission of interest

27.04.2020 Topic Assignment / Discussion

28.04.2020 Start working

04.06.2020 Intro + Related Work documented *

11.06.2020 Idea Presentation / Amazing Prototype

18.06.2020 Approach documented *

23.07.2020 Evaluation + Conclusion documented *

30.07.2020 Final Presentation

06.08.2020 Code Submission & Paper Submission

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 Docu etc.
- Communication 10%
  - Meeting Organization / Protocols, Questions & Concerns, Problems, Active Discussion Requests etc.

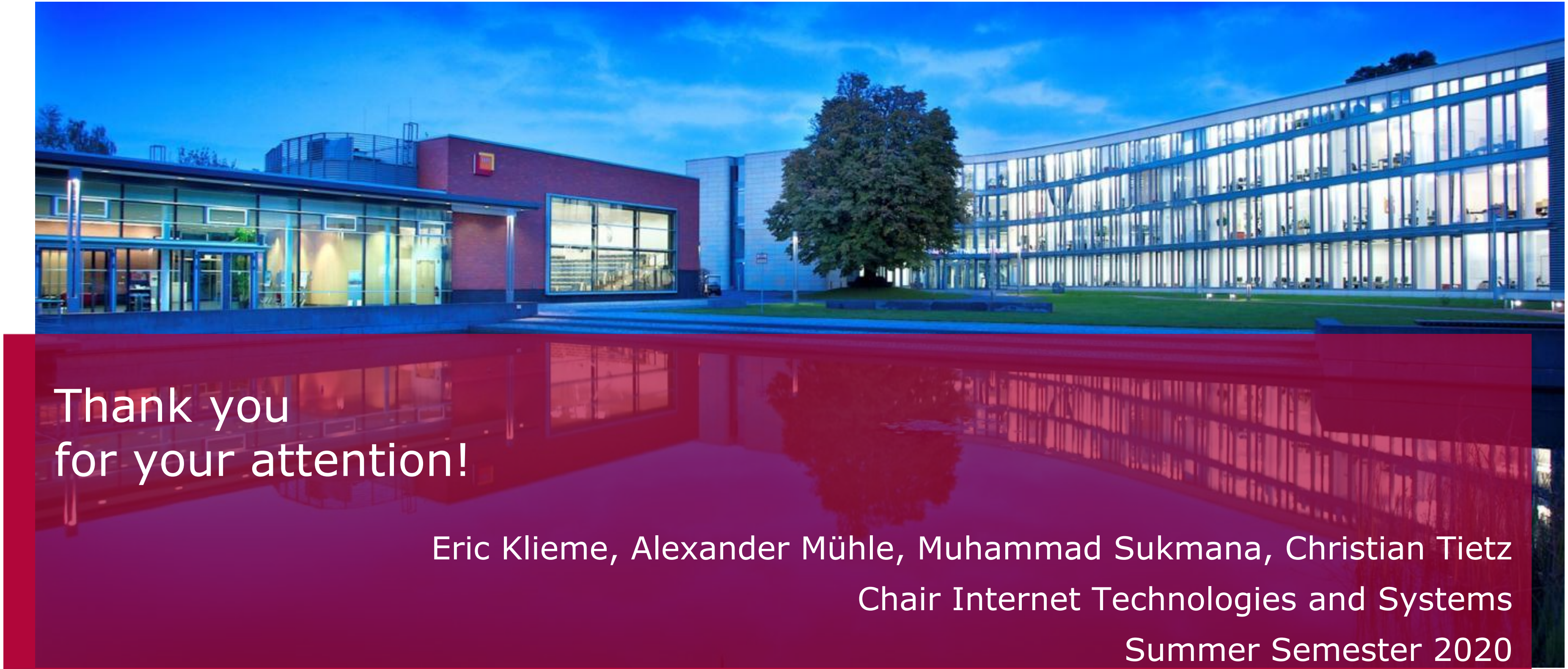
Enrollment

- If you are interested (as a single person, as a team):
  - Send a ranked list of your favorite three topics (and possible members)

- Per mail to
  - Eric.Klieme@hpi.de
  - Alexander.Muehle@hpi.de
  - Christian.Tietz@hpi.de
  - Muhammad.Sukmana@hpi.de

- Until 22nd of April (Wednesday!), **First Come First Serve**

HPI Hasso Plattner Institut
Digital Engineering · Universität Potsdam



Thank you
for your attention!

Eric Klieme, Alexander Mühle, Muhammad Sukmana, Christian Tietz
Chair Internet Technologies and Systems
Summer Semester 2020