Data and Network Security

(Master Degree in Computer Science and Cybersecurity)

Lecture 13

SAPIENZA
UNIVERSITÀ DI ROMA

Outline for today

- Recap last lecture
- Authentication methods

Outline for today

- Recap last lecture
- Authentication methods

Social Bots - Spam bots

(Semi-)automated accounts with usually harmful intention designed to mimic human behavior on social media platforms.

Social Bots - what they usually do/affect?

- Distortion of online discussion
- Manipulation of overall public opinion
- Lowering trust levels
- Misinformation amplification
- Possible economic implications
- Psychological implications

Social Bots - what they usually do/affect?

- Distortion of online discussion

By typically flooding platforms with repetitive or misleading content, spambots can push over the audiences certain viewpoints or narratives, artificially inflating their popularity or significance.

This distortion can make it difficult for users to discern genuine opinions and information from manipulated or fabricated content.

Social Bots - what they usually do/affect?

- Manipulation of overall public opinion

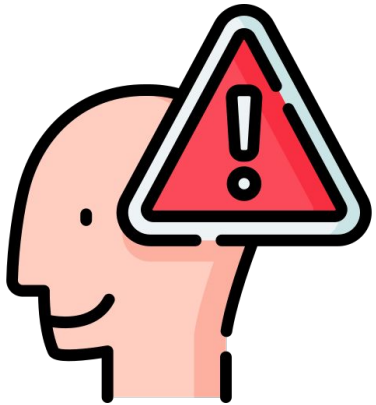
Engaging in coordinated campaigns to promote or discredit certain viewpoints, products, or political candidates this artificially generated content can alter public perceptions and attitudes, potentially impacting real-world outcomes such as elections or consumer behavior.

Social Bots - what they usually do/affect?

- Lowering trust levels

Aware users may become skeptical of the authenticity of social media engagement, leading to decreased trust in platforms and their content.

Fake accounts and automated interactions can tarnish the reputation of legitimate users and organizations, further eroding trust within online communities.



Social Bots - what they usually do/affect?

- Economic implications

Influencing consumer behavior and online market dynamics.
Spambots can inflate metrics such as:

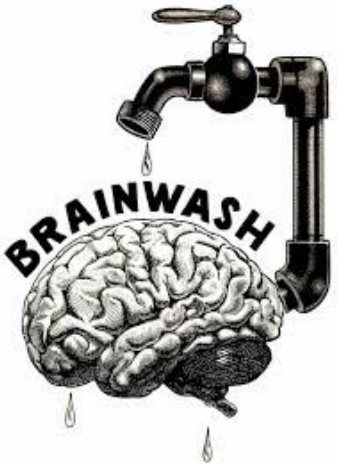
- likes, shares, and followers,

This can mislead advertisers/investors about the true reach and engagement of content of a “celebrity” which down the line with result in financial loss.

Social Bots - what they usually do/affect?

- Psychological implications

Exposure to manipulated or deceptive content may disrupt individuals' confidence/judgement in their ability to discern truth from false.



Social Bots - How to detect them?

Nowadays spambots have become indistinguishable from genuine accounts if analyzed one-by-one so:

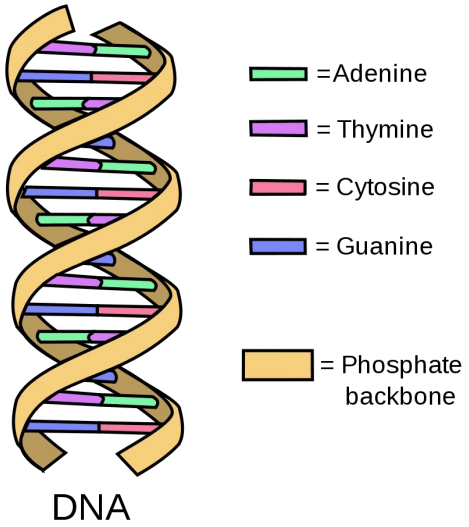
How about analyzing the online behavior of large groups of users, with the goal of detecting possible spambots among them?

Behavioral analysis

Behaviour - sequence of actions performed by an account.

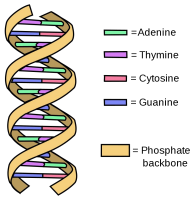
Inspiration from DNA:

- Each type of action is associated to a character (e.g., A, B, C)
- The online behaviour of an account is modeled as a sequence of characters according to the sequence of actions performed by that account



Spambot characterization

Automated accounts have similar DNA sequences.

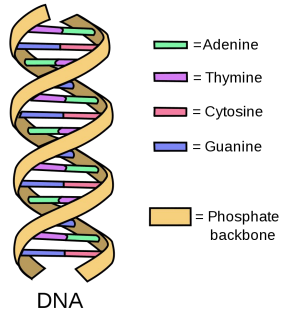


The most well-known and widely adopted analysis techniques are **sequence alignment** and **repetition/pattern elicitation.** One of the main goals of these techniques is to find commonalities and repetitions among DNA sequences.

Via an analysis of common sub-sequences and substrings it is possible to predict specific characteristics of the individual and to uncover relationships between different individuals.

Behavioral analysis

Behaviour - sequence of actions performed by an account.



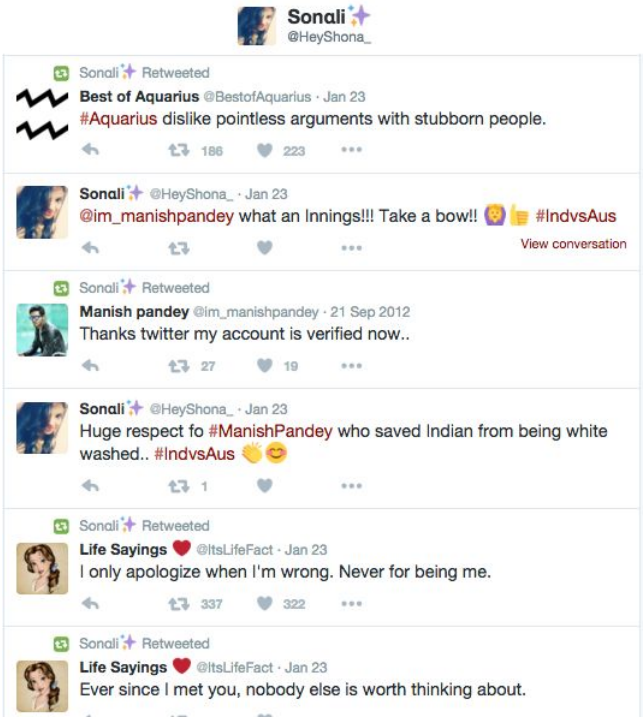
DNA

By drawing a parallel with biological DNA, we will see how to model users behaviors and interactions by means of strings of characters, representing the sequence of their actions.

Online actions-such as posting new content, replying to another user, following an account can be encoded with different characters, similarly to DNA sequences.

Digital DNA

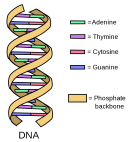
Encoding
T - tweet
R - retweet
P - reply



R
P
R
T
R
R

RPRTRR

Spambot characterization



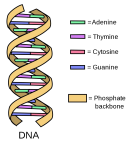
To create a digital DNA sequence all the possible actions are represented as a finite set of unique characters:

$$\mathbb{B} = \{B_1, B_2, \ldots, B_N\} \quad B_i \neq B_j \ \forall \ i, j = 1, \ldots, N \ \wedge \ i \neq j.$$

A digital DNA sequence is an ordered tuple, or row vector, of characters (i.e., a string) whose possible values are defined by the bases of its alphabet. A sequence s is defined as:

$$s = (b_1, b_2, \ldots, b_n) \quad b_i \in \mathbb{B} \ \forall \ i = 1, \ldots, n.$$

Spambot characterization

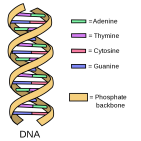


Different alphabets can model different granularities

$$\mathbb{B}^3_{content} = \left\{ \begin{array}{ll} \texttt{N} & \longmapsto \text{tweet contains no entities (plain text),} \\ \texttt{E} & \longmapsto \text{tweet contains entities of one type,} \\ \texttt{X} & \longmapsto \text{tweet contains entities of mixed types} \end{array} \right\}$$
$$= \{\texttt{N}, \texttt{E}, \texttt{X}\}$$

$$\mathbb{B}^6_{content} = \left\{ \begin{array}{ll} \texttt{N} & \longmapsto \text{tweet contains no entities (plain text),} \\ \texttt{U} & \longmapsto \text{tweet contains one or more URLs,} \\ \texttt{H} & \longmapsto \text{tweet contains one or more hashtags,} \\ \texttt{M} & \longmapsto \text{tweet contains one or more mentions,} \\ \texttt{D} & \longmapsto \text{tweet contains one or more medias,} \\ \texttt{X} & \longmapsto \text{tweet contains entities of mixed types} \end{array} \right\}$$
$$= \{\texttt{N}, \texttt{U}, \texttt{H}, \texttt{M}, \texttt{D}, \texttt{X}\}.$$

Longest common substring

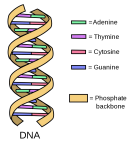


Observe the Longest substring between N sequences of digital DNA

…TRRRPRRTRRPRTPRPTPRRTRPR
…RPRTPTTRRRPRRTPRRRRTPPRP
…TTTRRRPRRRPRRTRTRPTRRRTP
…PRTRPRTPPPPRTPRRRRRPRRTR

Intuitively, users that share long behavioral patterns are much more likely to be similar than those that share little to no behavioral patterns.

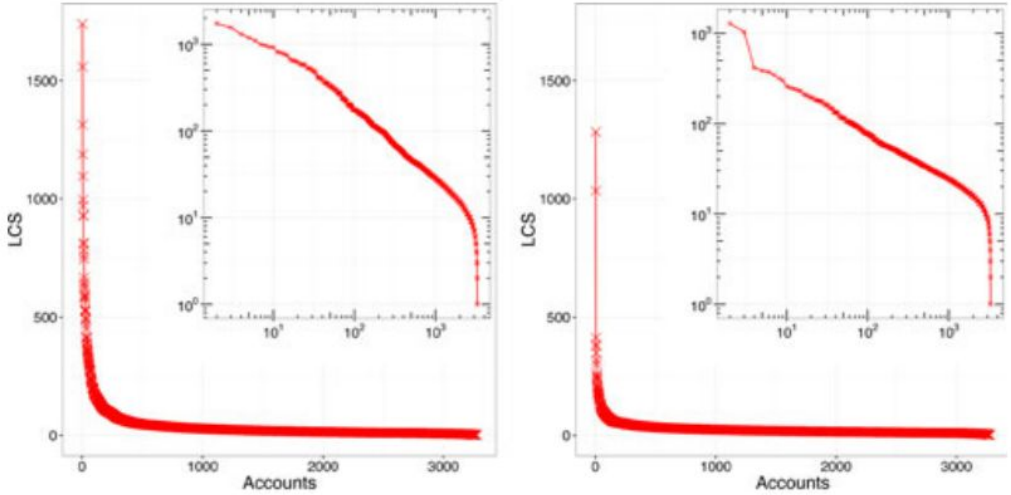
Longest common substring



Goal: find the LCS that is common to at least k of these strings:

$$A = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_M \end{pmatrix} = \begin{pmatrix} (b_{1,1}, b_{1,2}, \ldots, b_{1,n}) \\ (b_{2,1}, b_{2,2}, \ldots, b_{2,m}) \\ \vdots \\ (b_{M,1}, b_{M,2}, \ldots, b_{M,p}) \end{pmatrix}$$

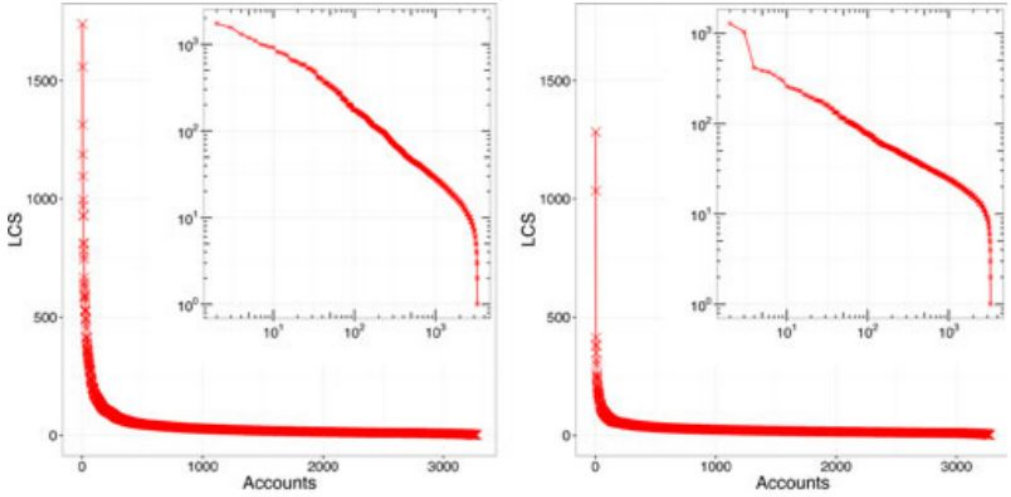
Longest common substring



(a) $\mathbb{B}^3_{type}$ alphabet. (b) $\mathbb{B}^3_{content}$ alphabet.

- x axis - the number of k accounts (corresponding to the k strings, used to compute LCS values)
- y axis - the length of the LCS common to at least k accounts.

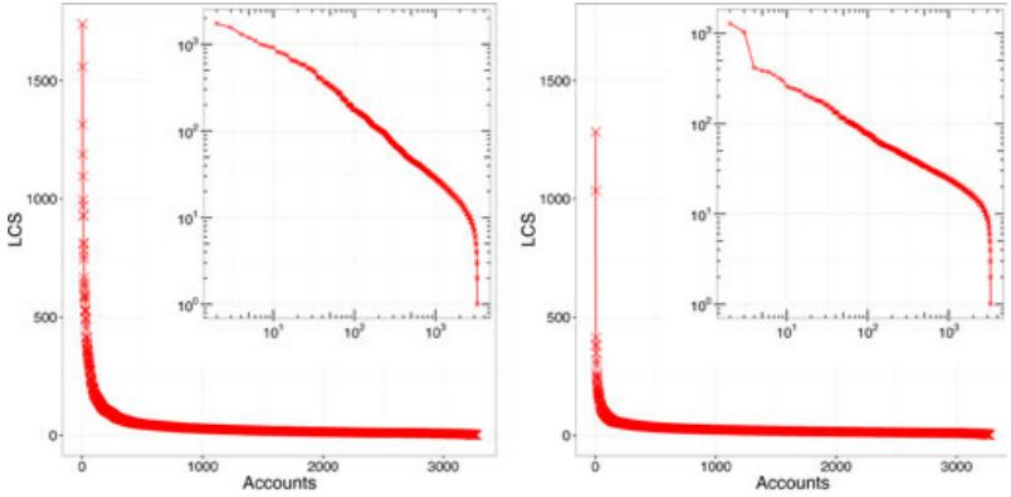
Longest common substring



(a) $\mathbb{B}^3_{type}$ alphabet. (b) $\mathbb{B}^3_{content}$ alphabet.

Each point in an LCS curve corresponds to a subset of k accounts that share the longest substring (of length y) among all those shared between all the other possible subsets of k accounts.

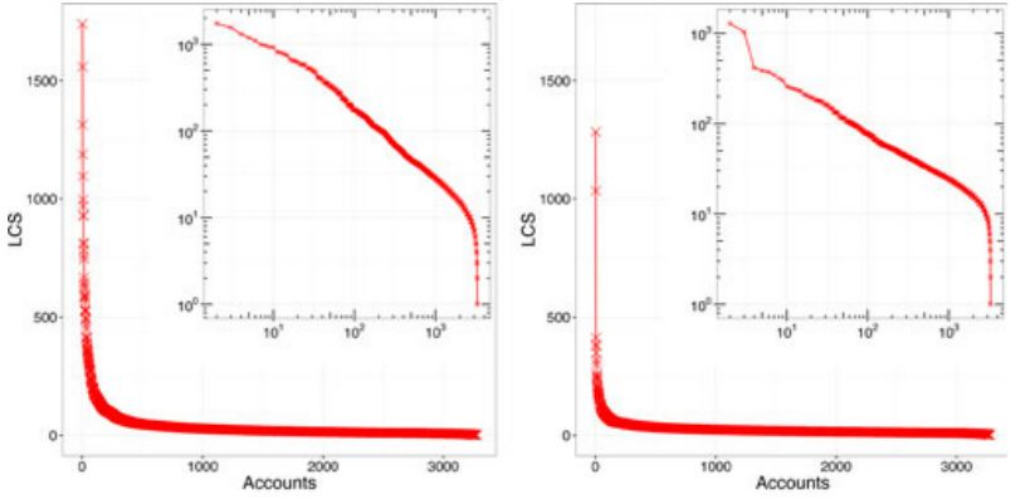
Longest common substring



(a) $\mathbb{B}^3_{type}$ alphabet. (b) $\mathbb{B}^3_{content}$ alphabet.

As the number k of accounts grows,
the length of the LCS common to all of them shortens.
LCS curves are monotonic non-increasing functions.

Longest common substring

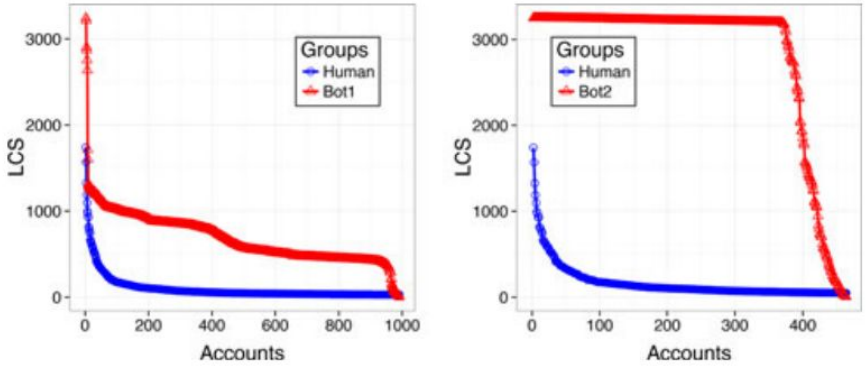


(a) $\mathbb{B}^3_{type}$ alphabet. (b) $\mathbb{B}^3_{content}$ alphabet.

Thus, it is more likely to find a long LCS among a few accounts rather than among large groups.

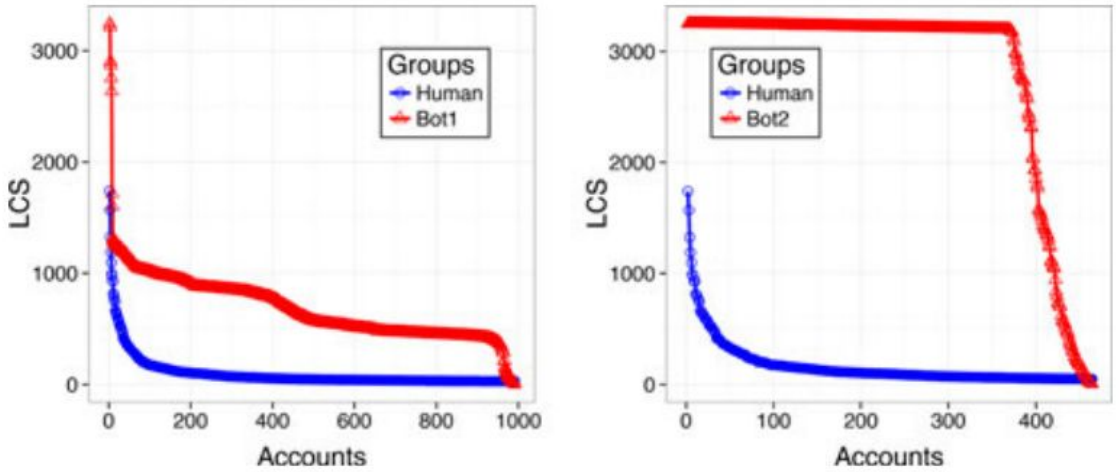
LCS curves are monotonic non-increasing functions.

Bots versus Humans



- LCS of both groups of spambots are rather long even when the number of accounts grows.
- Sudden drop of the LCS length when the number of accounts gets close to the group size for spambots
- Genuine accounts show little to no similarity-as represented by LCS curves that exponentially decay, rapidly reaching the smallest values of LCS length.

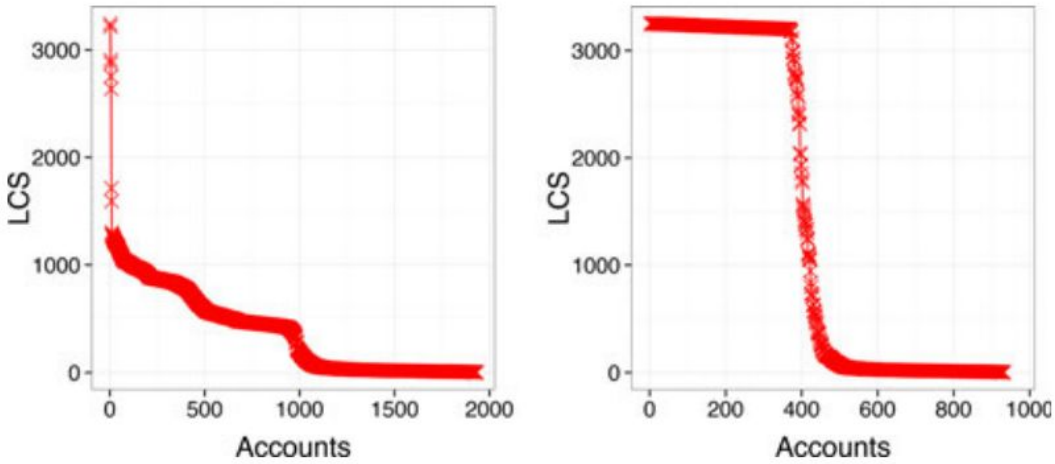
Bots versus Humans - Automation in distinguishing?



Consider high behavioral similarity as a proxy for automation and, thus, a high level of similarity among a large group of accounts might serve as an indicator for anomalous behaviors.

Heterogeneous Users - LCS

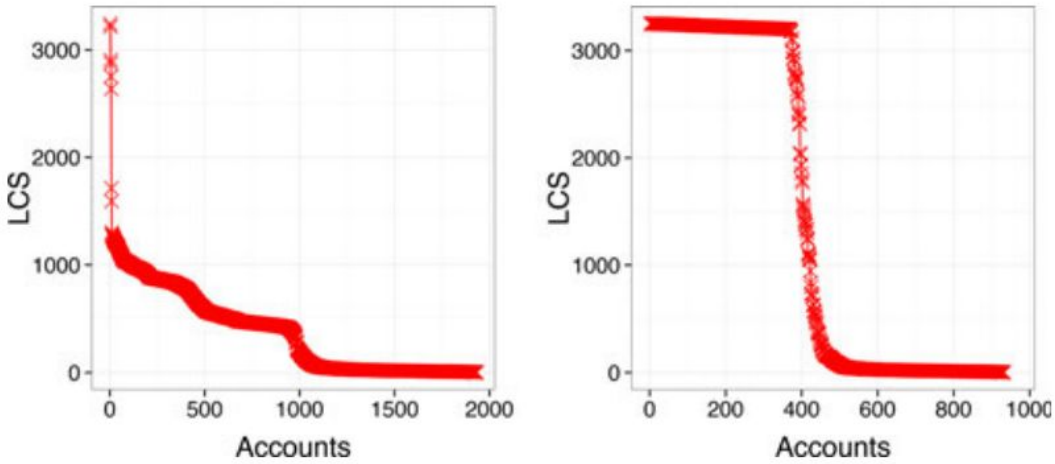
Mix various bot families and same quantity of normal users data.



LCS curves in both plots asymptotically reach their minimum value as the number of accounts grows.

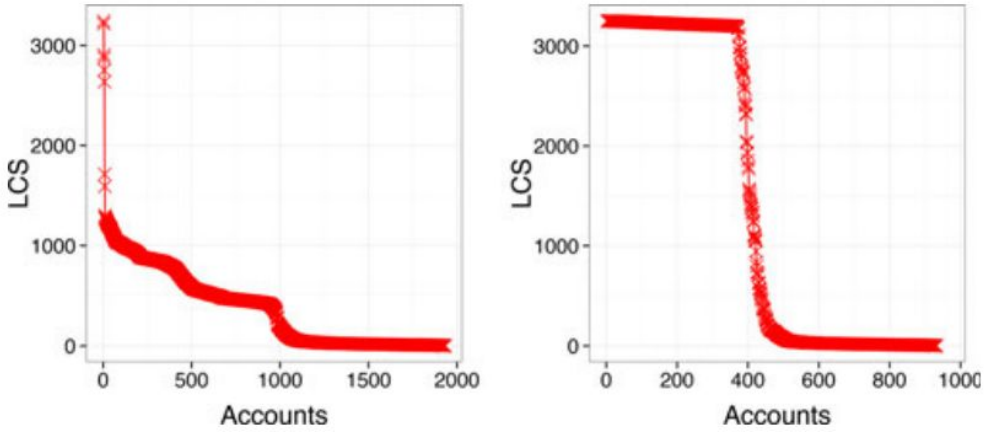
Heterogeneous Users - LCS

Mix various bot families and same quantity of normal users data.



We have a different behaviour compared to the individual groups

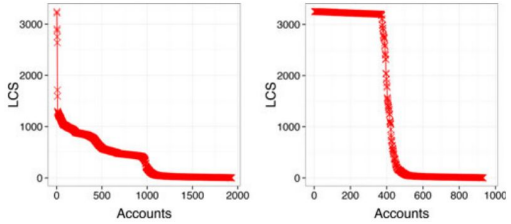
Heterogeneous Users - LCS



Observation: A trend seems to be dominant only until reaching a certain threshold. Then, a steep fall occurs and another possibly different-trend kicks in.

Observation: These portions of the LCS curves separated by the steep drops resemble LCS curves of the single groups of similar users

Heterogeneous Users - LCS



The steep drops of LCS curves separate areas where the length of the LCS remains practically unchanged, even for significantly different numbers of considered accounts.

These plateaux in LCS curves are strictly related to homogeneous groups of highly similar accounts. (multiple plateaux - multiple sub groups existing).

The steeper the drop in a LCS curve, the more different are the two subgroups of accounts split by that drop.

Slow and gradual decreases in LCS curves represent areas of uncertainty, where it might be difficult to make strong hypotheses about the characteristics of the underlying accounts.

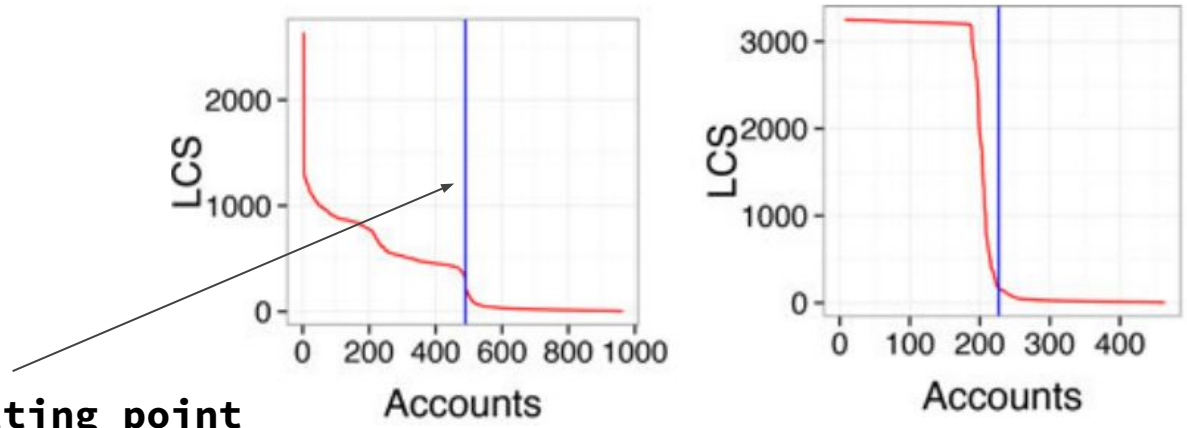
Supervised approach - case 1

A good division of the original set of users into several subgroups is one where all the users belonging to a given class are assigned to the same subgroup.

LCS curve of a heterogeneous group of users can be used as a splitting point to obtain two subgroups of more homogeneous users.

Supervised approach - case 1 - evaluation

Using a labeled dataset check all possible splitting point in the LCS curve of the training-set users and find the one that yields the best possible subgroup division. (Every point generates a different “classifier”)

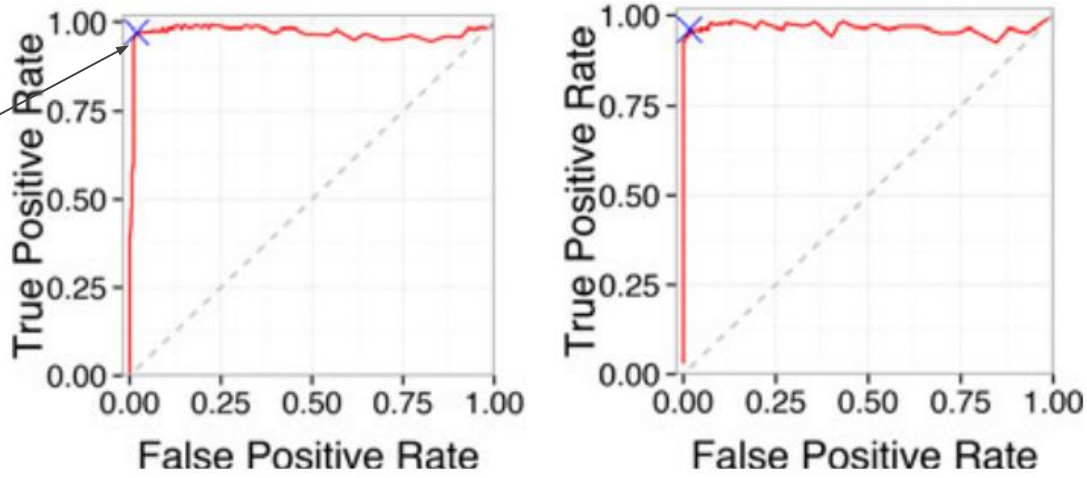


Best splitting point
Left – spambots
Right – genuine users

Supervised approach - case 1 - evaluation

Supervised – Cannot guarantee that the learned LCS value would still be effective when applied on a test-set different from the one used to derive the LCS value.

Best model



Outline for today

- Recap last lecture
- Authentication methods

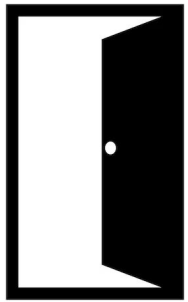
Authentication methods

Authentication is the process of verifying the identity of a user.

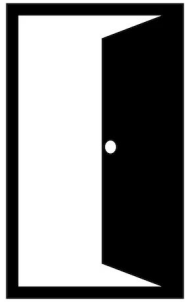
Objectives:

- Establish trust
- Prevent unauthorized access
- Protect sensitive information

Authentication methods - Actors



Authentication methods - Actors

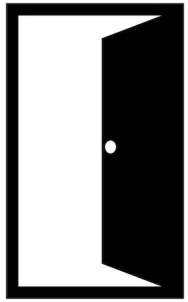


Claimant

Monitor

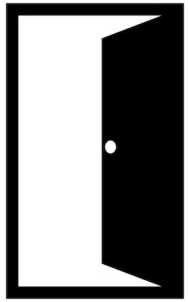
Information system

Authentication methods - Actors - Claimant



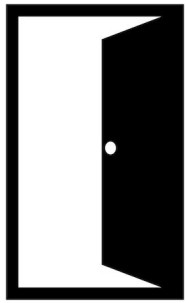
Claimant: entity that **authenticates** to the system, in order to use the services.

Authentication methods - Actors - Monitor



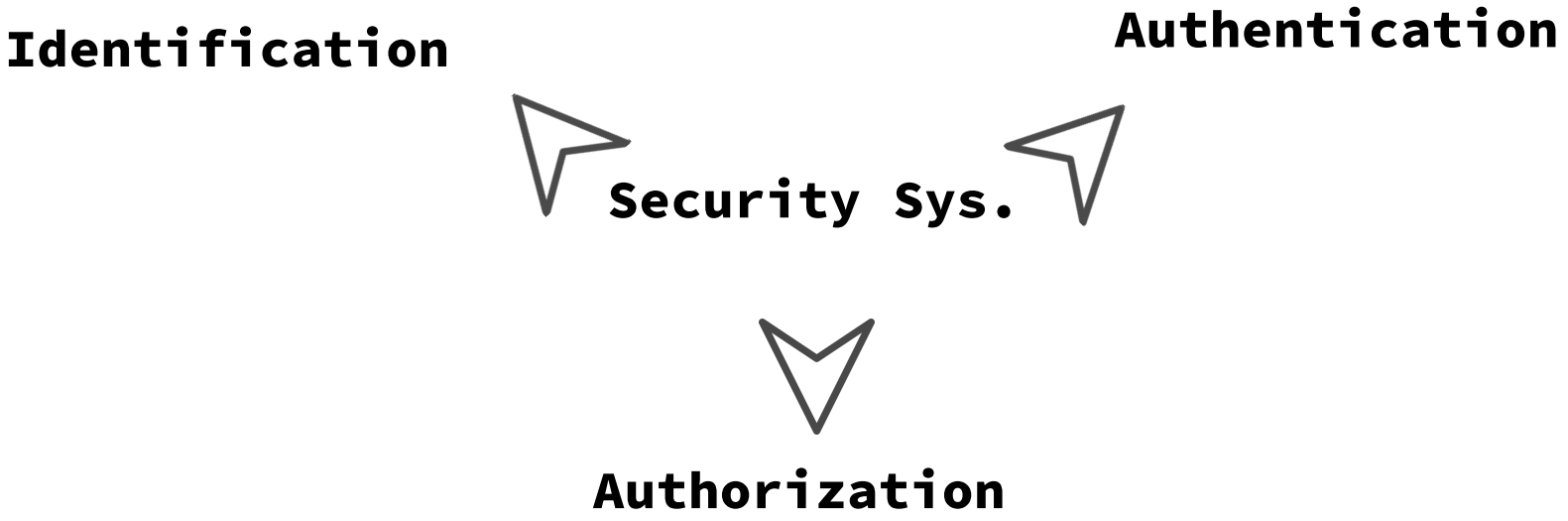
Monitor: the entity that provides an authentication service. It asserts the identity of a claimant and checks if it can grant him/her the use of the required service.

Authentication methods - Actors - Information System



Information System: provides services, such as an access to a computer account, an application, a door unlocking or a network printer, and will let the claimant use its services if the monitor correctly authenticated it

The core of a security system



The core of a security system

- Identification: Communicate your identity to the IS
- Authentication: Proof given by a claimant to assert by a monitor that claimant really corresponds to the identity he provided.
- Authorisation: The granted privileges to the claimant

The core of a security system - another crucial thing

We need a way to connect the claimant and the monitor:

Channel: is a support of communication between the claimant and the monitor.

It can either be considered as:

- confidential,
- authentic,
- secure
- insecure.

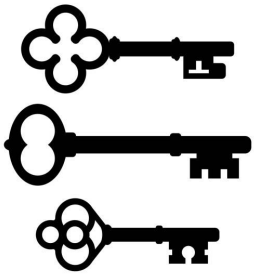
The core of a security system - another crucial thing

Channel is:

- **Confidential** - if it is interception resistant
- **Authentic** - if it is resistant to tampering
- **Secure** - resistant to both interception and tampering
- **Insecure** - obviously none

Authentication types - categories

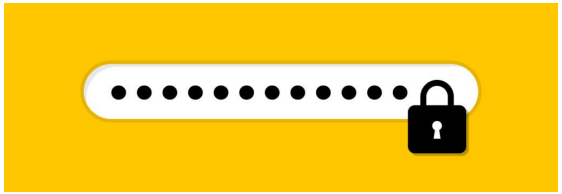
1. Ownership model (you own something)
    1. Physical keys
    2. Smart card
    3. NFC
    4. RFID
    5. Hardware-token

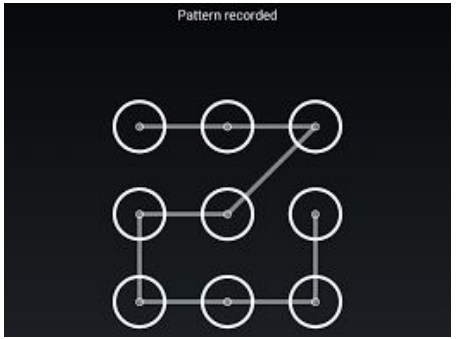


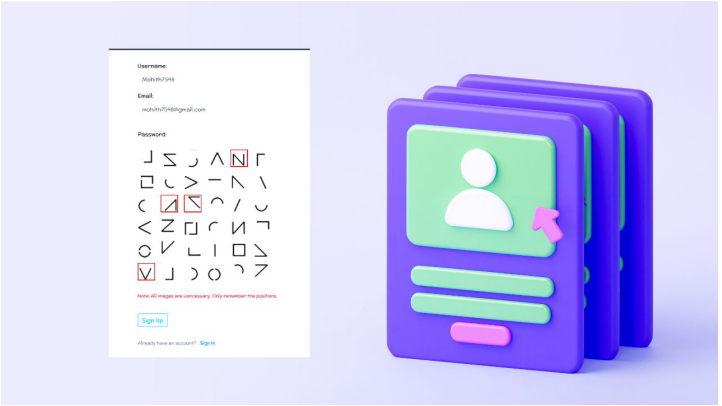
Authentication types - categories

2. Knowledge-based model:

    1. Passwords
    2. PIN code
    3. Lock pattern
    4. Graphical password
    5. Rhyme based
    6. Challenge response

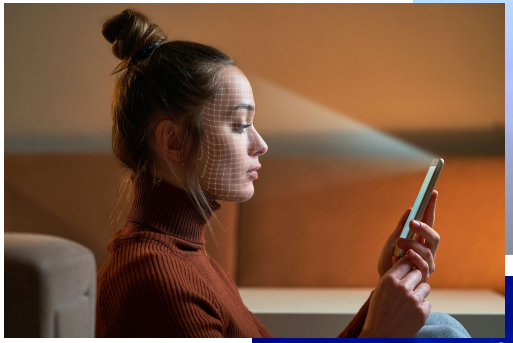


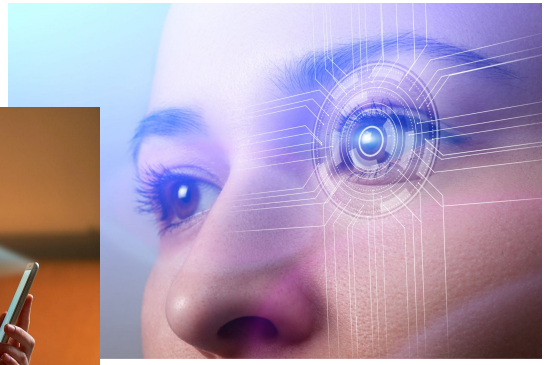


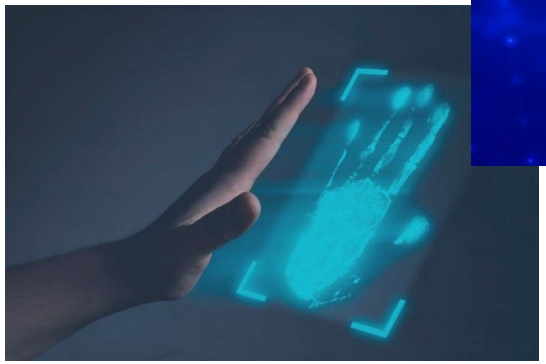


Authentication types - categories

3. Inherent-based model
    1. Fingerprints
    2. Palm
    3. Iris
    4. Voices
    5. Gestures
    6. Face





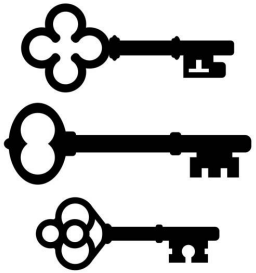


Authentication types - Issues/Drawbacks

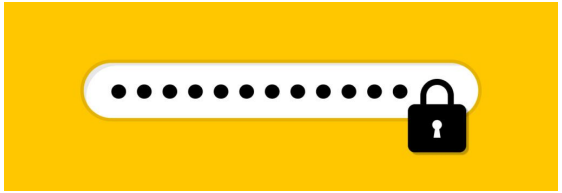
1. Ownership model (you own something)

    1. Losing it
    2. Token stealing
    3. High cost
    4. MITM attack
    5. Usability

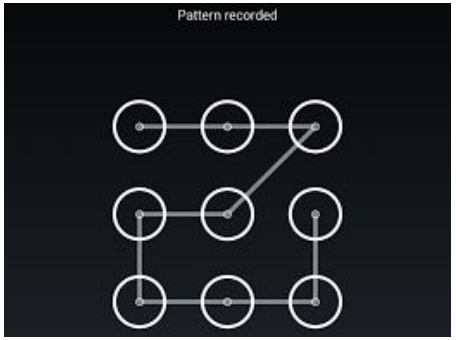
RFID



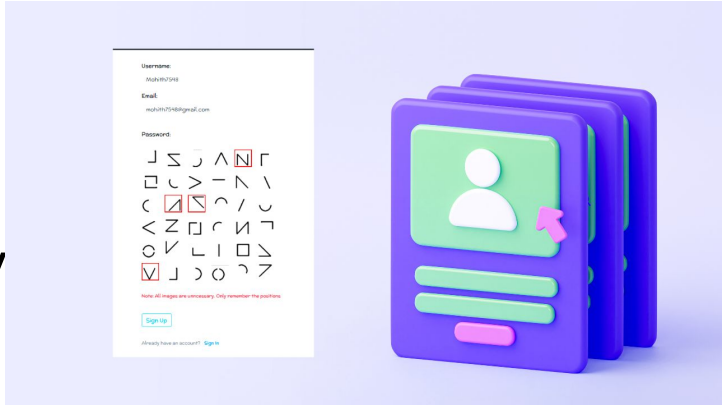
Authentication types - Issues/Drawbacks



2. Knowledge-based model:

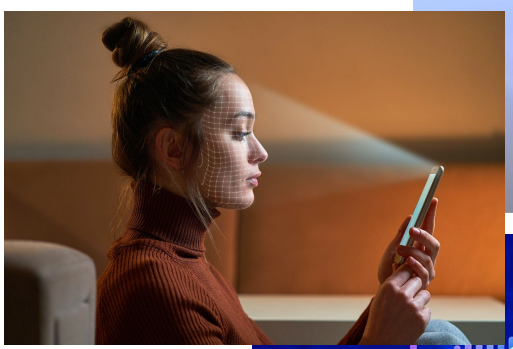


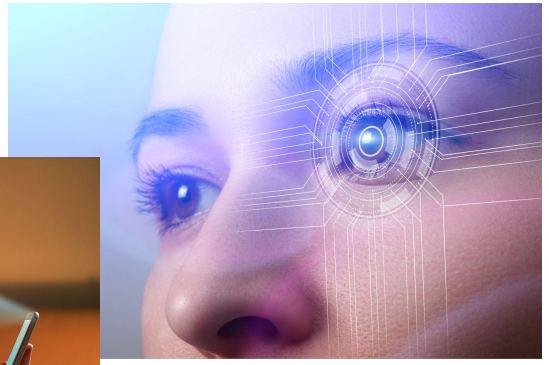
1. Keylogging
2. Shoulder surfing
3. Guessing
4. Brute force
5. Dictionary attacks
6. Screen Capturing
7. MITM attack
8. Memorization capability

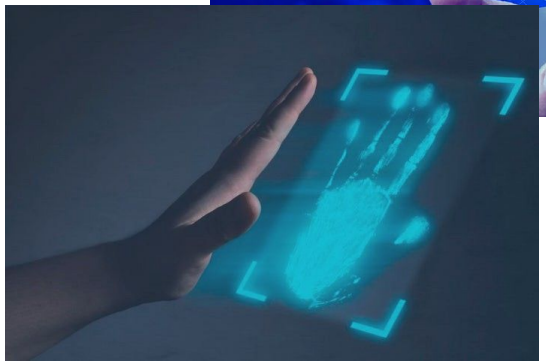


Authentication types - Issues/Drawbacks

3. Inherent-based model
    1. Forgery
    2. Replay
    3. MITM
    4. High costs
    5. Accuracy
    6. Medical (scars)
    7. Lighting conditions
    8. Clothes/Jewelry







Authentication types - Nowadays

Increase in cyber security threats

Authentication systems are being more empowered than before

(Non) Conventional attempts at multi-factor authentication

- Graphical centered systems
- Touch based
- EEG (electroencephalogram) based
- Web-based

Graphical centered systems

Various types of images or shapes are used as password.

- Images can be easily remembered by human than text.
- Human brains can process images easily.
- Utilizing images renders it resistant to dictionary attack, keylogger, social engineering etc.

Two types of graphical password techniques:

- Recognition based
- Recall based

Graphical centered systems

Two types of graphical password techniques:

- **Recognition based**
- **Recall based**

Recognition based - various images are presented to the user and from that user has to recognize the right images in a correct sequence.

Recall based - user has to reproduce something that he/she has created or selected during registration.

Graphical centered systems - some examples

Enrollment phase:

- Users choose some images from a set of 25 pictures
- And after that users are presented with 3 questions and they should select 3 points region-of-answer

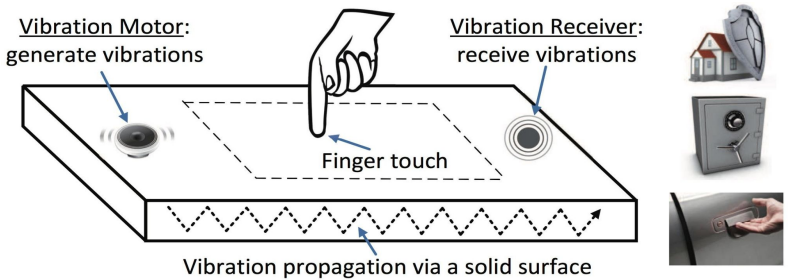
Authentication phase:

- Users should select correct images respectively in the first phase of authentication then they should select three regions of the preselected pictures as for the next step.

Graphical centered systems - some examples

Vibration and pattern:

To pass the authentication step, the user must select the same number of cells and feel the same number of vibration code as he had done in the registration phase.



Touch based systems

The user is authenticated based on the geometric properties of his drawn curves as well as his behavioral and physiological characteristics.

- No need to look at the screen

Touch based systems - one approach

- In the enrollment phase, a device owner uses one or multiple fingers to draw arbitrary geometric curves of his own choice (called a curve password) on his multi-touch screen.

- An authentication template is then created based on features extracted from his input, including:
  - **x-coordinate, y-coordinate, direction, curvature,**
  - **x-velocity, y-velocity,**
  - **x-acceleration, y-acceleration,**
  - **finger pressure, and hand geometry.**

Touch based systems - one approach

- **x-coordinate, y-coordinate, direction, curvature,**
  - relate to the something-you-know paradigm and can together
    accurately define the geometric characteristics of the drawn curve.
  - self-defined curves are easier for the user himself to remember and reproduce but difficult for attackers to guess.

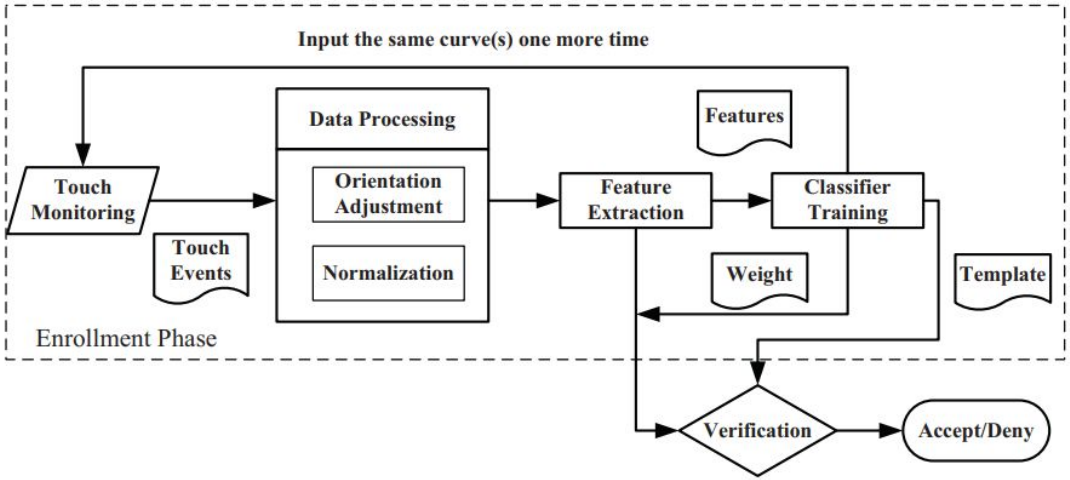
Touch based systems - one approach

- **x-velocity, y-velocity, x-acceleration, y-acceleration, finger pressure, and hand geometry.**
  - Something-you-are paradigm:
    - features correspond to the user’s behavioral and physiological characteristics.
    - They are nearly impossible for attackers to infer and forge even if they may manage to know the correct curve password.

Touch based systems - one approach

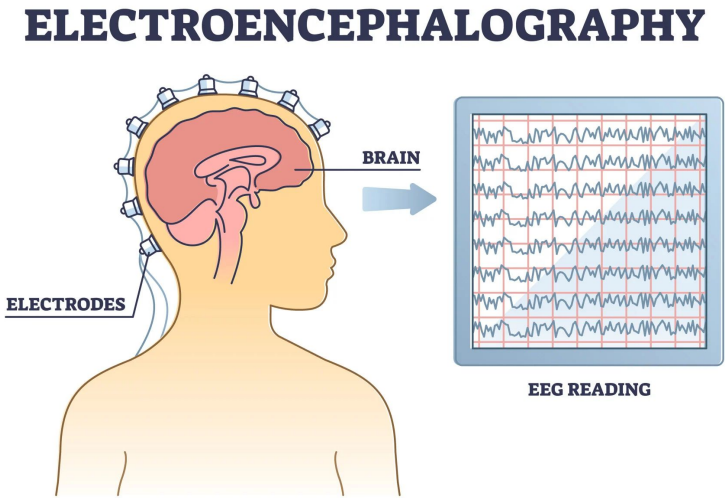
Authentication phase:

- anyone attempting to unlock the mobile device needs to draw on the multi-touch screen, from which a candidate template is extracted.
- If the candidate and authentication templates match, the user is allowed in.



EEG (electroencephalogram) based systems

An **electroencephalogram** (EEG) is a test that measures electrical activity in the brain using small, metal discs (electrodes) attached to the scalp. Brain cells communicate via electrical impulses and are active all the time, even during sleep. This activity shows up as **wavy lines** on an EEG recording.



EEG (electroencephalogram) based systems

Purpose: Strengthen password based authentication to prevent possible reply attacks by demanding also a certain mental state of the user that is entering the password.

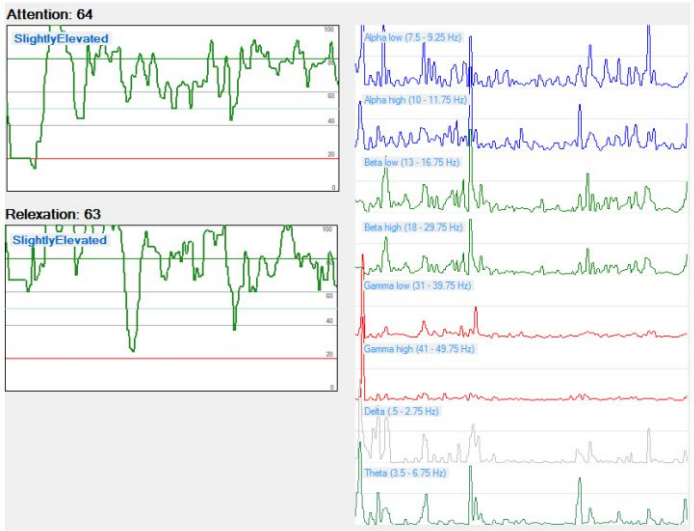
Possible issues:

- The main challenge of this research is recreating the brainwave. EEG as a biometric characteristic lacks consistency which depends on **stress**, **fatigue**, **medication**, **environment** (electrical equipment), etc. To cope with this, researchers often use some kind of **stimuli** to help in recreating the valid authentication EEG pattern (imagining letters, text, somebody you care etc).
- EEG devices are not cheap.

EEG (electroencephalogram) based systems

Authenticate with cheap EEG:
Cheap EEG devices are effective is recognising the mental state of the user.

- using **alpha** and **beta waves** to determine users' **relaxation** and **concentration**
- Can not measure effects caused by thought (**pass-thought**)

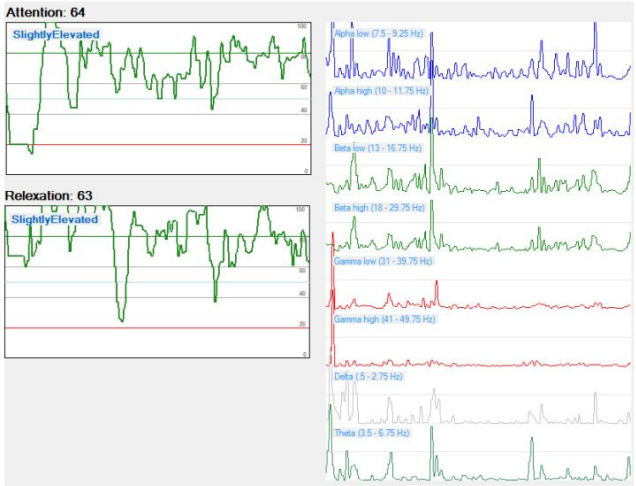


EEG (electroencephalogram) based systems

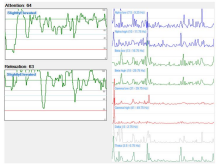
Authenticate with cheap EEG:

Cheap EEG devices are proved effective is recognising the mental state of the user

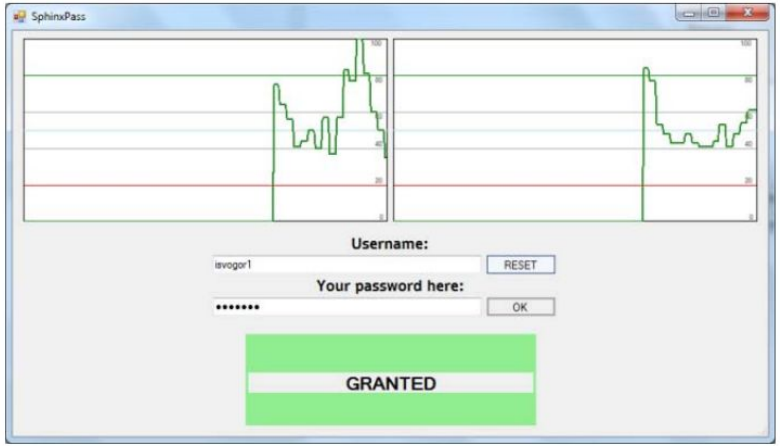
- using alpha and beta waves to determine users' relaxation and concentration
- Can not measure effects caused by thought (pass-thought)
- **EEG signals that it provides can be used as additional parameters for a password**



EEG (electroencephalogram) based systems

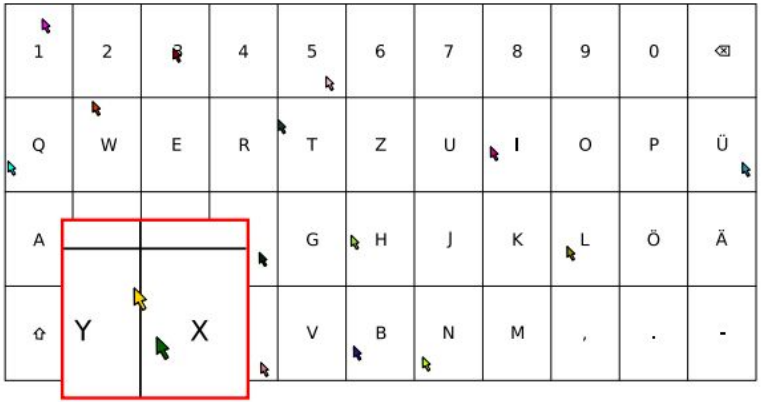


- The password will be divided into shorter segments, and for each segment, the user will select the mental state which must be matched to a segment in order for the full password to be accepted.
- User will be able to change the mental state by trying to concentrate on some problem or relaxing
- Neurofeedback, where user can see his mental state, and act to change it.



Web based systems - strengthening authentication

On-screen keyboards are often used to minimize the possibility of losing the password due to keyloggers and other malicious software. For instance, this is commonly used by online banking websites. They enforce the use of virtual keyboards or keypads to input the secret credentials.
While being more secure against keyloggers and the like, this approach is highly vulnerable to shoulder surfing attacks, that is, an attacker observing the input from a nearby position.



Reading Material

1. Authentication methods: Link-1, Link-2, Link-3