On the Feasibility of Side-Channel Attacks with Brain-Computer Interfaces

Ivan Martinovic, Doug Davies, Mario Frank, Daniele Perito, Tomas Ros, Dawn Song

KAIST

Presented by Valentin Guittard

Questions

1. Are you listening?

2. Who is familiar with Brain-Computer Interfaces (BCI)?

3. And Electroencephalography (EEG)?

4. Who read this paper?

Content

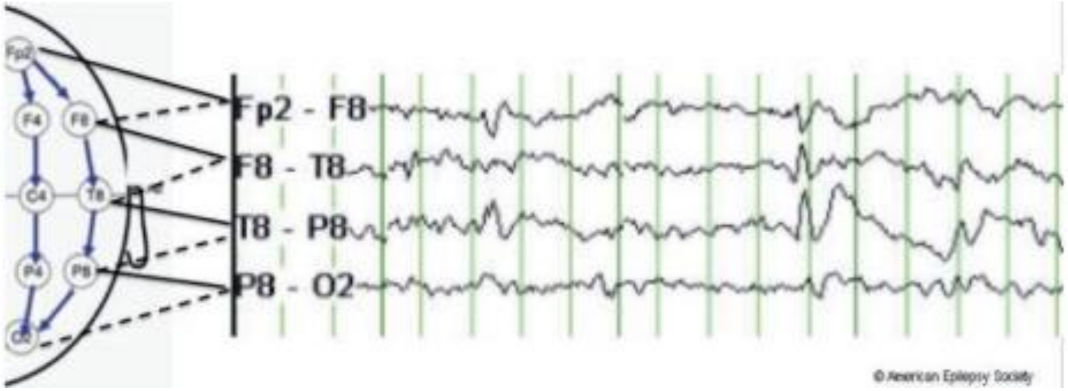
- **Introduction to BCI, EEG and ERP**
- ***"On the Feasibility of Side-Channel Attacks with Brain-Computer Interfaces"***
  - **Related Work**
  - **Experiments & Methodology**
  - **Results & Contribution**
- **Questions**

Brain-Computer Interface (BCI)

- Communication tool between **users** and **systems**
- **No external device** or **muscle** intervention
- Video games, hands-free keyboards, medicine…

Electroencephalography (EEG)

- Simple and **non-invasive**
- Records **electrical fields** produced by the **neuronal activity**
  - Millions of synchronized neurons
  - Captured by **14 scalp electrodes**
  - **Sample frequency 128-512Hz** typically

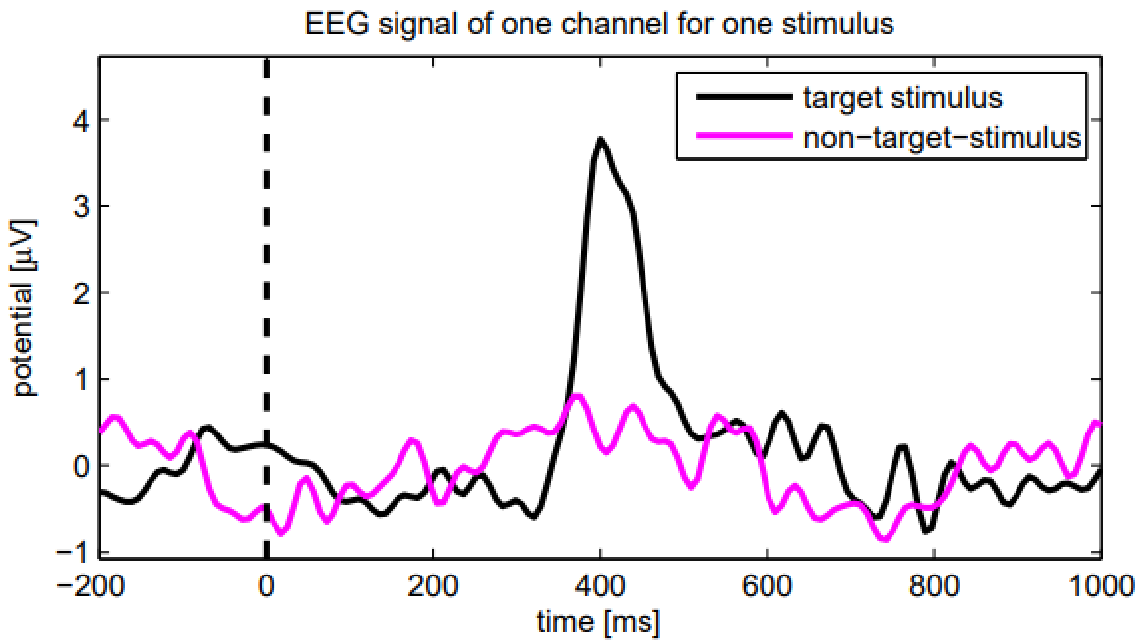


Event-Related Potential (ERP)

"An Event-Related Potential is detected as a pattern of voltage change after a certain auditory or visual stimulus is presented to a subject. Every ERP is time-locked to the stimulus"

Most prominent ERP: P300

- **Amplitude peak** in the EEG signal
- **300ms after** the stimulus
- Response to **target / personally meaningful stimuli**



On the Feasibility of Side-Channel Attacks with Brain-Computer Interfaces

Ivan Martinovic, Doug Davies, Mario Frank, Daniele Perito, Tomas Ros, Dawn Song

KAIST

Who?

Ivan Martinovic

Professor of Computer Science

University of Oxford, England

Doug Davies, Mario Frank, Daniele Perito, Dawn song

Professors at UC Berkeley, US

Tomas Ros

Cognitive neuroscientist

University of Geneva, Switzerland

Related works

1999

"Parametric person identification from the EEG using computational geometry"

Poulos et al.

2007

"Person Authentication Using Brain waves (EEG) and Maximum A Posteriori Model Adaptation"

Marcel et al.

2009

"A new approach for EEG feature extraction in P300-based lie detection"

Abootalebi et al.

2010

"Guessing What's on Your Mind: Using the N400 in Brain Computer Interfaces"

Van Vliet et al.

2012

2016

"Hacking the brain: brain–computer interfacing technology and the ethics of neurosecurity"

Marcello et al.

2017

"Side-Channel Attacks Against the Human Brain: the PIN Code Case Study"

Lange et al.

2023

"Brain-Hack: Remotely Injecting False Brain-Waves with RF to Take Control of a Brain-Computer Interface"

Armengol-Urpi et al.

BCI Devices

- **Consumer-grade** BCI devices
- Low-cost **EEG-based** BCI devices
- Software development kits provided

A MindSet device (NeuroSky)

An EPOC device (Emotiv Systems)

Threat model

Attacker:

- Malicious third-party developer

Goal:

- Retrieve personal information with no malware

Attacker assumptions:

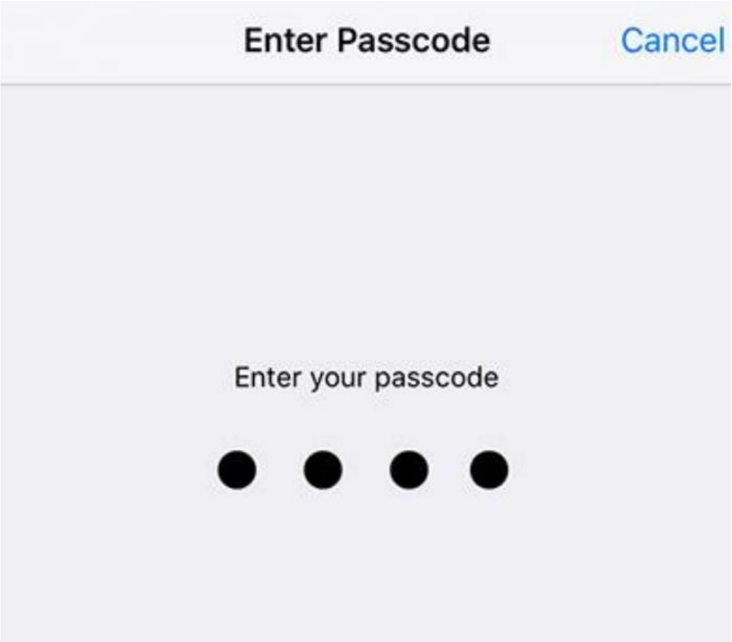
- Can read the EEG signal
- Can display text, images and video on a screen

Experiments

- **5 experiments**
- **≈ 90s** for each experiment
- **30** participants
- Three main steps:
  - ***Verbal explanation*** *of the task by the operator*
  - *On-screen* ***message*** *for 2 seconds*
  - ***Images flashed*** *in random order for the duration of the experiment*

Experiment 1 – PIN Code

1. “Choose and memorize a **random 4-digit PIN**”
2. “Enter **first digit** at the end of the experiment”



Experiment 2 – Bank Information

1. Show **logos** of different banks
2. Show images of **debit cards**

What is the name of your bank?

Bank logos

Debit cards

Experiment 3 – Month of Birth

1. Flash the **names of the months** randomly

When were you born?

Experiment 4 – Face Recognition

Do you know any of these people?

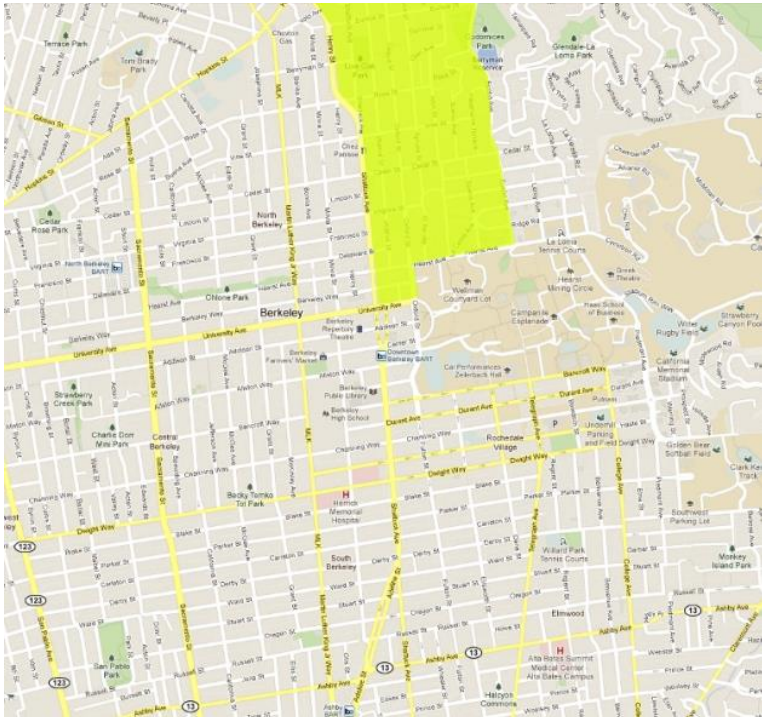
10 unknown persons

Barack Obama

Experiment 5 – Geographic Location

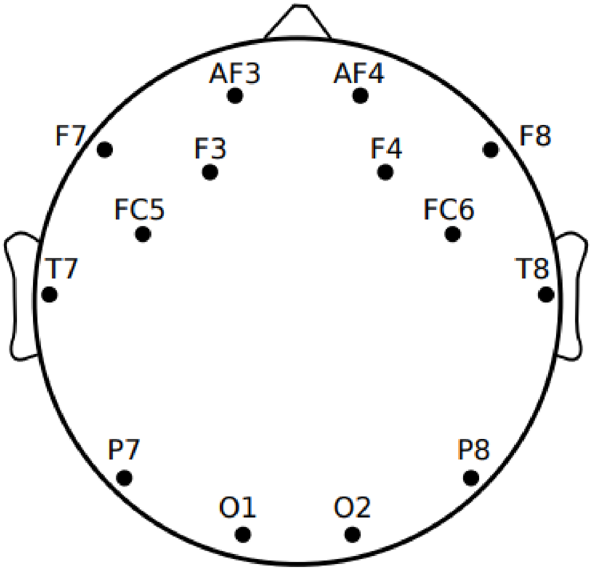
1. Show a map with **different highlighted zones**

Where do you live? Count the occurrences



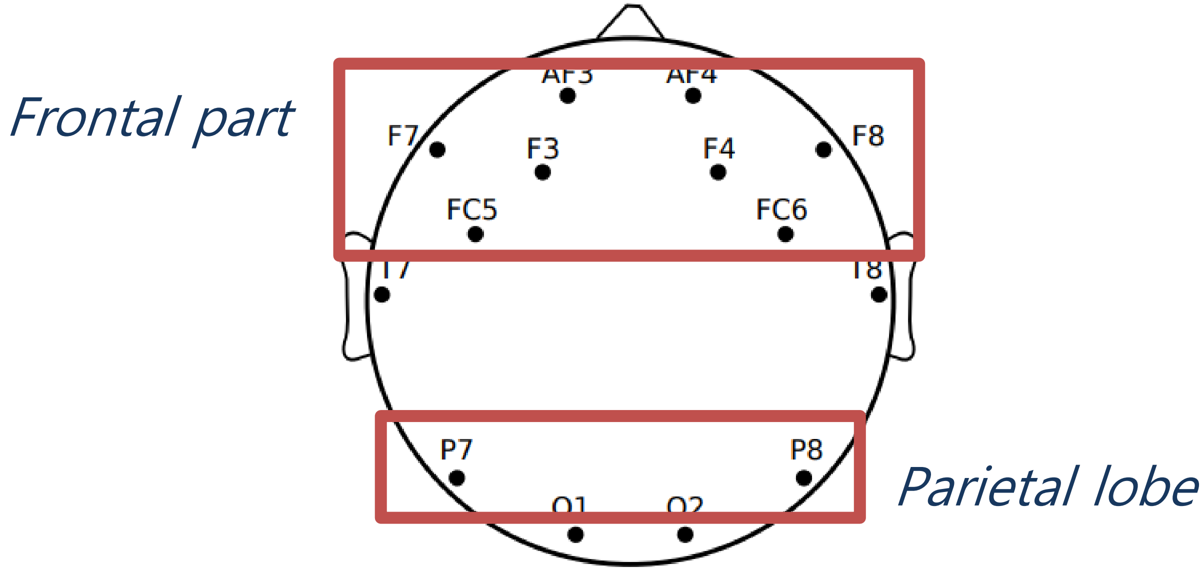
Data Collection

- EEG signals recorded by **14 electrodes**
- At a **128Hz sampling rate**, create tuples **(EEG signal, stimuli)**



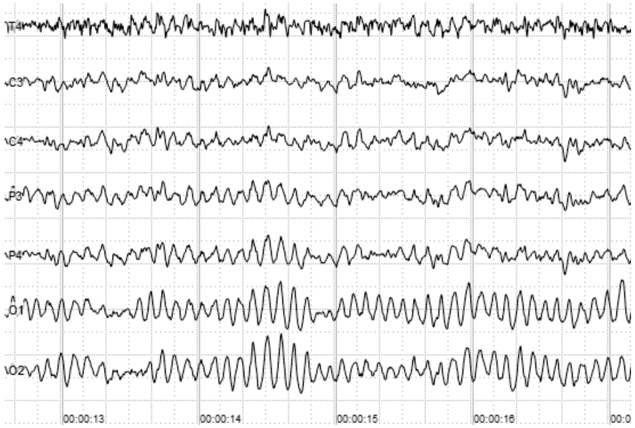
Data Collection - Challenge

- **Reliability** of P300 detection & **discrimination** of other EEG signals
- **Passive user**
- Target device **not made for detecting P300**



Binary Classification

- Set of (EEG data = **epochs**, **stimulus**)



- Two phases: **training phase, classification phase**

Training Phase

Idea

- Train the classifier to map an **epoch** to the correct **stimuli**

Input

- A **set of epochs** $x \in X^{\mathrm{tr}}$
- A **vector of label** $y \in Y$

Output

- A function $\boldsymbol{g}$ that maps epochs to target stimuli labels:

$$\boldsymbol{g(x) = y}$$

Classification Phase

Idea

- Use the model to **obtain stimulus from epoch** of the test set

Input

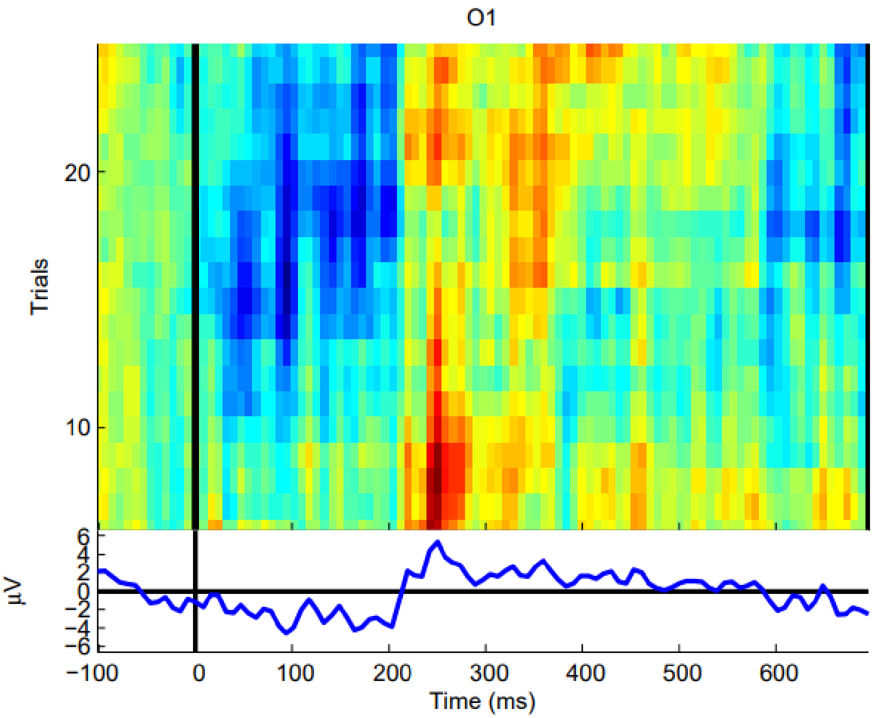
- A **set of new epochs** $x^{test} \in X^{test}$

Output

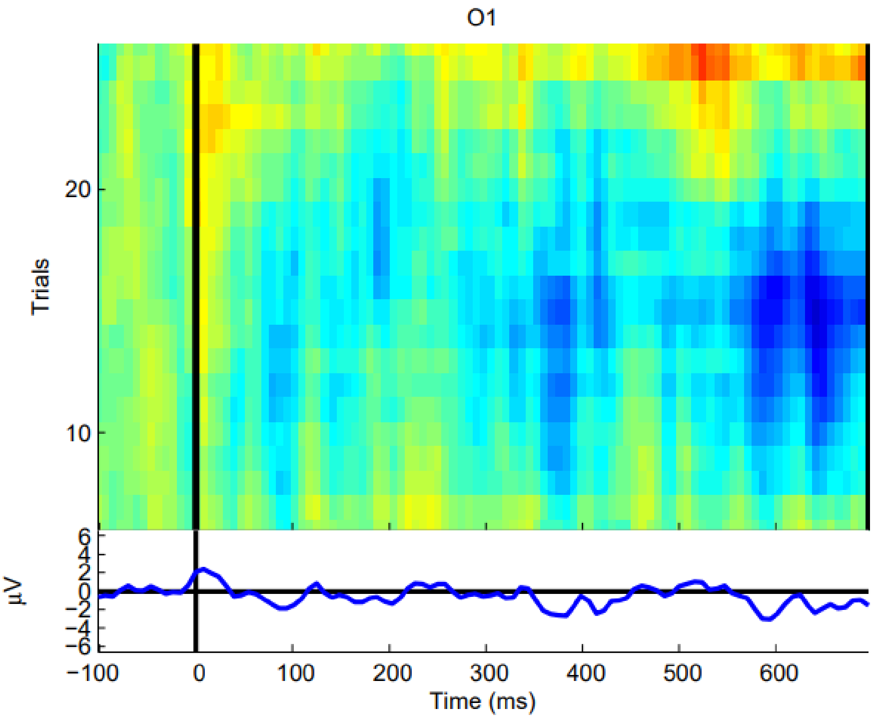
- A **set of estimation** $\{\ \hat{y} = g(x^{test})\ \}$

Classification Phase

- For a stimulus k, $\mathbf{N_k^{(+)}} = \sum_{\mathbf{i \in E_k}} \hat{\mathbf{y}}_\mathbf{i}$
- Highest Nk is used to **estimate the target stimuli**



(a) target stimulus



(b) non-target stimulus

Classifier Functions

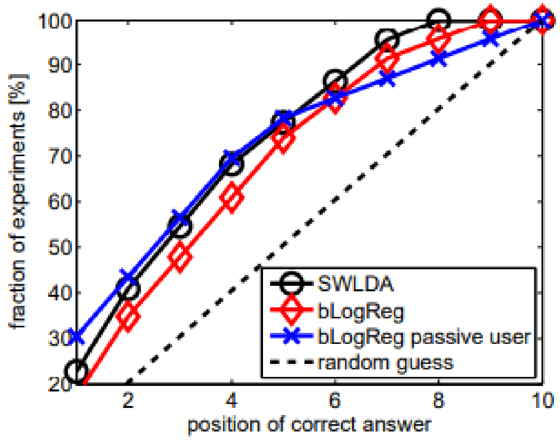
1. Boosted logistic regression (**bLogReg**)
   - Model trained on the training data
   - Minimize the negative Bernoulli log-likelihood
2. Stepwise Linear Discriminant Analysis (**SWLDA**)
   - Extension of Fisher's linear discriminant analysis (LDA)
   - **Robust to noise**

Results – Classifiers Calibration

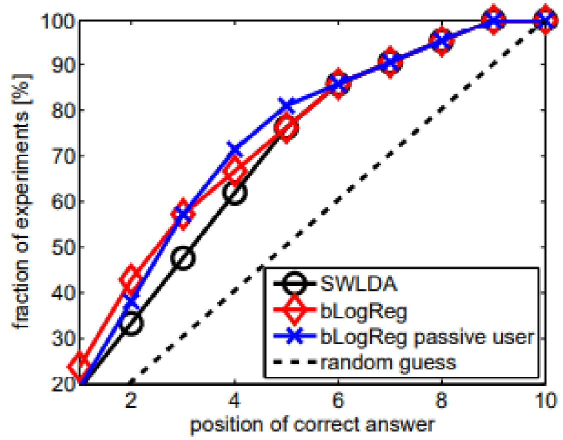
1. User-supported calibration
   - User **supports** the training phase
   - User **does not support** stimuli detection

2. On-the-fly calibration
   - User **does not support** the training phase (nor disturbs it)
   - User **does not support** stimuli detection

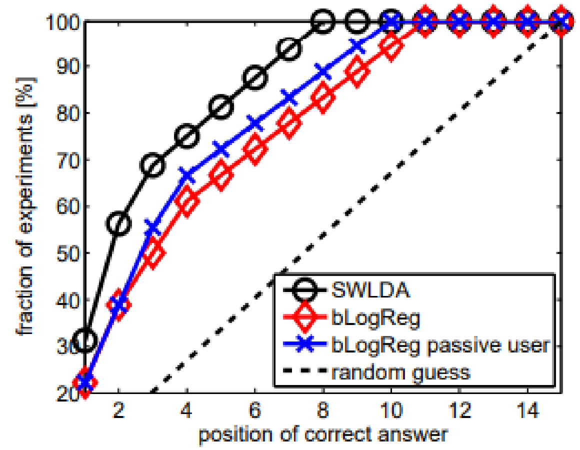
Results



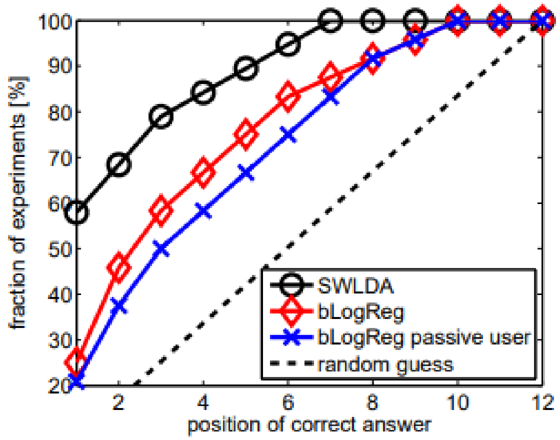
(a) 1st digit PIN



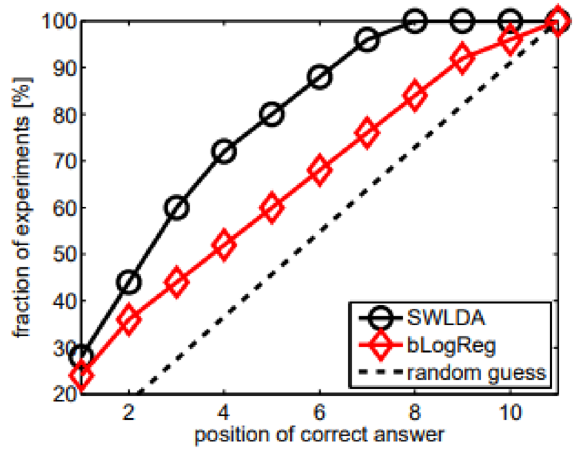
(b) Debit card



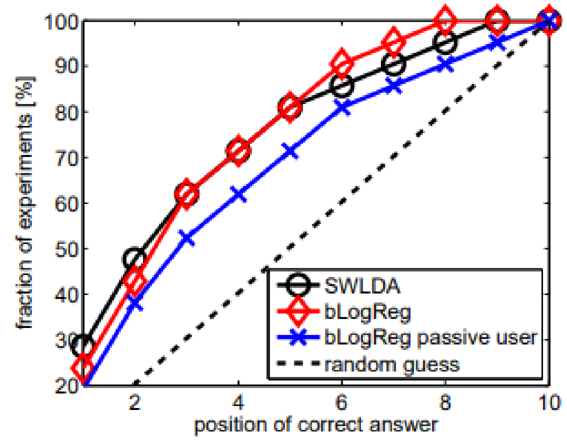
(c) Location



(d) Month of birth



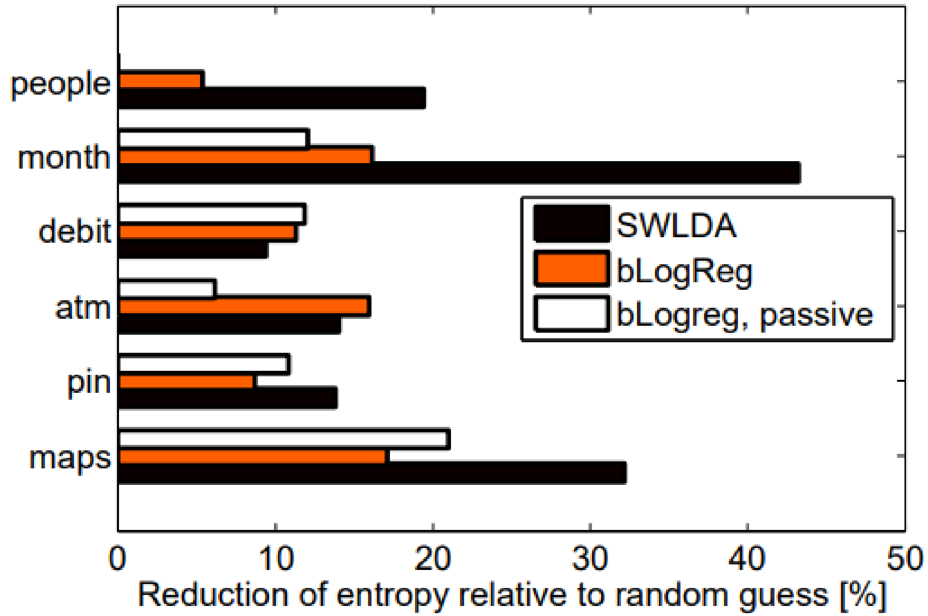
(e) People



(f) ATM machine

Results

- Entropy loss: **15 – 40%**
- Better accuracy with **user cooperation**
- Improve accuracy with **prior knowledge**



Summary & Contribution

Problem

- Rise of consumer-grade Brain Computer Interfaces (BCIs)
- No literature on security implications of using BCI devices

Contribution

- Used cheap EEG-based BCI devices to conduct simple and effective attacks

Result

- Entropy of private information decreased by 15-40 % compared to random-guessing attacks *(i.e., information are easier to get)*

Meaning

- First study of security risks related to consumer-grade BCIs
- Demonstrated that BCIs could be turned against users to reveal their private and secret information

Defenses

- Users can **focus on non-target stimuli** to hinder probing
  - Unrealistic
- Create restricted APIs
  - Stops exposure of raw data to third-party developers
  - But reduces their potential
- Add noise to the EEG raw data
  - Interferes with legitimate application

Questions

- How feasible would it be to implement more secure protocols to prevent these information leakage?
- How impactful can this attack be in real-world? Can it change the way BCIs or treated or commercialized?

Best Questions

- **[Hansung Bae]** I think not only brain waves but also eyes can show the unconscious movements. I'm curious if there has been any research on side-channel attacks using eye-tracking technology to exploit the movements of the eyes.

Al-Haiqi, M. Ismail and R. Nordin

"The eye as a new side channel threat on smartphones"

2013 IEEE

Best Questions

- **[SeongRyong Oh]** Can cryptography be applied to sending EEG signals?

Best Questions

- **[Hyeongju Lee]** Is it resilient when facing patterns resembling attacks in various noise environments?

Thank you for your attention!

KAIST