ISA 419: Data-Driven Security

20: Authentication Using Behavioral Biometrics

Fadel M. Megahed, PhD

Endres Associate Professor
Farmer School of Business
Miami University

@FadelMegahed
fmegahed
fmegahed@miamioh.edu
Automated Scheduler for Office Hours

Spring 2024

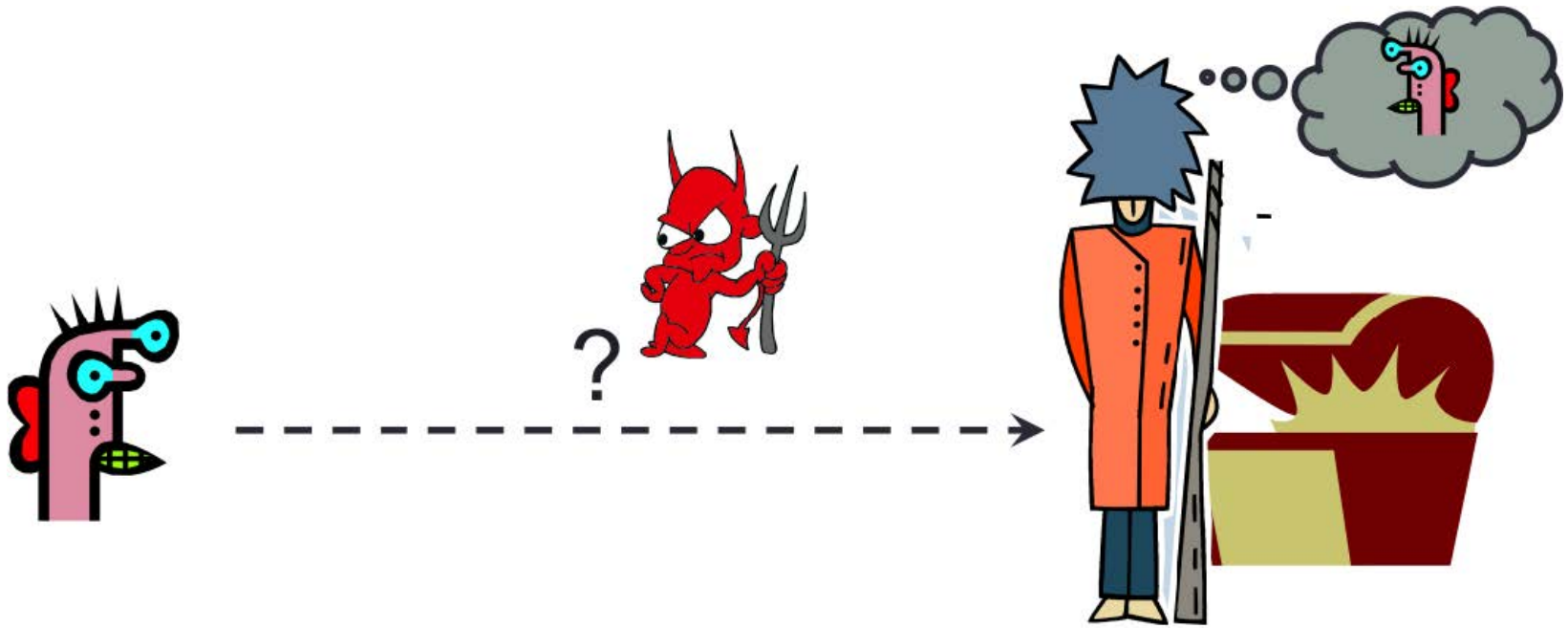
Learning Objectives for Today's Class

- Motivating the use of behavioral biometrics for authentication purposes
- Explain possible sources of data and application for behavioral biometrics
- Model behavioral biometrics data using machine learning techniques

Motivating the Use of Behavioral Biometrics for Authentication Purposes

Authentication

- **Authentication** is the process of verifying the identity of a user or system.



How do you prove to someone that you are who you claim to be?

Information Traditionally Used for Authentication

User authentication is the most common and best studied human security task.

Something you know Passwords, PINs, security questions

Something you have: Smart cards, tokens, mobile phones

Something you are: Biometrics (fingerprint, face, iris, etc.)

Lifecycle for Authentication Information

Issuance: User enrolls in the system and provides authentication information

Usage: User logs in and provides authentication information

Maintenance/Revocation: User's authentication information is revised or retired

Lifecycle: Something You Know

Issuance:

- User "has" to memorize a new password or (unfortunately) reuse an existing one.
- Should follow password guidelines (length, complexity, uniqueness)
- Needs to be sufficiently secure and memorable

Usage:

- Can I remember my password?
- If so, which one is it (out of the many I have or the updates I have made)?

Maintenance:

- Subject to loss/expiration
- Re-issuance may require secondary mechanism
- Rules on freshness/variation for re-issued data

Lifecycle: Something You Have

Issuance:

- User "has" to carry a new token or smart card.
- Should be kept secure and not shared with others.
- Needs to be sufficiently secure and portable.

Usage:

- Requirements on human memory: Where did I leave my card? Which card is it?
- Requirements on physical presence: Do I have it with me?

Maintenance:

- Subject to loss/theft
- Requires periodic replacements

Lifecycle: Something You Are

Issuance:

- “Reverse issuance” required to submit biometrics
- Accessibility: not all humans have readable fingerprints, irises, etc.

Usage:

- Minimal requirements placed upon human memory, e.g., “Which finger did I use?”, though may be specified at authentication
- Human-machine interface issues, e.g., what to do with a cut finger?

Maintenance:

- May possibly change with aging, injury, etc.
- In general, there are limited options for renewal due to finite set of biometrics.

Lifecycle: General Guidelines

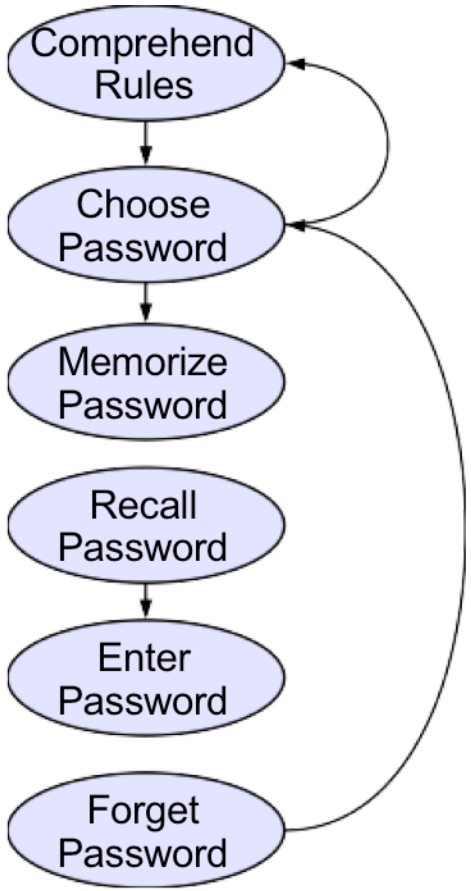
- **Issuance:**
  - Limit amount of physical interaction
  - Limit human processing and learning requirements
  - Limit number of seemingly artificial constraints
- **Usage:**
  - Limit human memory requirements
  - Limit requirements for perfect accuracy
- **Maintenance:**
  - See Issuance guidelines
  - Limit excessive update requirements

Password Lifecycle

Issuance

Use

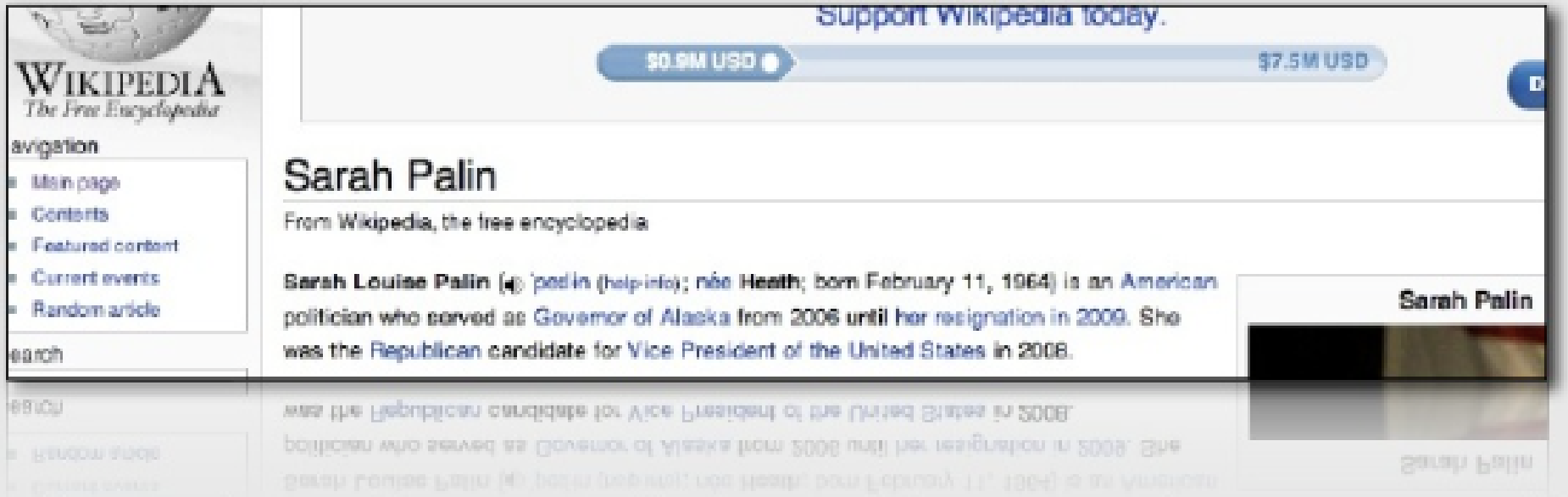
Maintenance



The "Traditional" Approach to Replacing Passwords ⊗

The "Traditional" Approach to Replacing Passwords ⊗

- Sarah Palin's Yahoo! Mail account was hacked in Sep 2008 via her secret question
- First secret question was... "what is your birthdate?"



- Second question was... "where did you meet your spouse?"

Behavioral Biometrics for Authentication

The Keystroke Dynamics - Benchmark Data Set

Based on Killourhy and Maxion (2009), the authors set up their experiment as follows:

- **Password:** .tie5Roanl
- **Feature sets:**
  - The **Enter** key was considered to be a part of their password
  - **Keydown-Keydown:** time between the key presses of consecutive keys is used as a feature;
  - **Keyup-Keydown:** time between the release of one key and the press of the next is used;
  - **Hold:** time between the press and release of each key is used
- Note that their features are linearly dependent; for example, $Keydown.Keydown = Hold + Keyup - Keydown.$
- Additionally, their features are all recorded in seconds.

Choice of Features

	Source Study	Detector	Feature Sets				Password	
			Enter Key	Keydown-Keydown	Keyup-Keydown	Hold	Length	Reps
1	Joyce & Gupta (1990) [10]	Manhattan (filtered)	✓	✓			N/A	8
2	Bleha et al. (1990) [2]	Euclidean (normed)		✓			11–17	30
		Mahalanobis (normed)		✓			11–17	30
3	Cho et al. (2000) [4]	Nearest Neighbor (Mahalanobis)	✓		✓	✓	7	75–325
		Neural Network (auto-assoc)	✓		✓	✓	7	75–325
4	Haider et al. (2000) [8]	Fuzzy Logic		✓			7	15
		Neural Network (standard)		✓			7	15
		Outlier Count (z-score)		✓			7	15
5	Yu & Cho (2003) [21]	SVM (one-class)	✓		✓	✓	6–10	75–325
6	Araujo et al. (2004) [1]	Manhattan (scaled)		✓	✓	✓	10+	10
7	Kang et al. (2007) [11]	k-Means			✓	✓	7–10	10

Subject × Sessions

We recruited 51 subjects from within the university. Subjects completed 8 data-collection sessions (of 50 passwords each), for a total of 400 password-typing samples. They waited at least one day between sessions, to capture some of the day-to-day variation of each subject's typing.

Our set of subjects consisted of 30 males and 21 females. We had 8 left-handed and 43 right-handed subjects. The median age group was 31–40, the youngest was 18–20 and the oldest was 61–70. The subjects' sessions took between 1.25 and 11 minutes, with the median session taking about 3 minutes.

Non-Graded Class Activity

10:00

- Use the following link to access the data set.
- Download the data set and explore the features.
- In plain English, define three different approaches for modeling this dataset and write them below.

- Edit me to answer the question above.
- Edit me to answer the question above.
- Edit me to answer the question above.

In-Class Demo

Recap

Summary of Main Points

By now, you should be able to do the following:

- Motivating the use of behavioral biometrics for authentication purposes
- Explain possible sources of data and application for behavioral biometrics
- Model behavioral biometrics data using machine learning techniques

Review and Clarification

- **Class Notes**: Take some time to revisit your class notes for key insights and concepts.
- **Zoom Recording**: The recording of today's class will be made available on Canvas approximately 3-4 hours after the end of class.
- **Questions**: Please don't hesitate to ask for clarification on any topics discussed in class. It's crucial not to let questions accumulate.