Computer Security
Dept. of Computer Engineering,
Chulalongkorn University.

Activity VII: Soft Biometric (Keystroke)

Instructors : Krerk Piromsopa, Ph.D

- Walking pattern
- Typing pattern

- Record typing pattern
- Recognize who is the typer
- n > 2

Overview

Soft Biometrics traits are physical, behavioural or adhered human characteristics, classifiable in pre–defined human compliant categories. These categories are, unlike in the classical biometric case, established and time–proven by humans with the aim of differentiating individuals. In other words the soft biometric traits instances are created in a natural way, used by humans to distinguish their peers.

(from https://en.wikipedia.org/wiki/Soft_biometrics)

In this activity, we will create a soft biometric based on keystroke patterns. There are several methods for identifying a person using keystroke patterns. For simplicity, this activity will focus on distinguishing the patterns between two persons.

There are at least three properties of keystroke that can be used: time for pressing and releasing a key, time between two (or more) consecutive keys, (average) time for typing a word (or a phrase or a sentence).

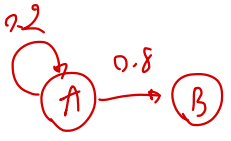
For the time between two consecutive keys, we can construct a digraph. A digraph table is a signature for identifying a person. The idea is to find an average time between a key pair for each person. To identify a person, we can ask a person to type a (long) paragraph and calculate time between each key pair. Then, we can apply the K-nearest neighbor to find a digraph profile that is best matched. A similar idea is used to create a trigraph (time between three consecutive keys).

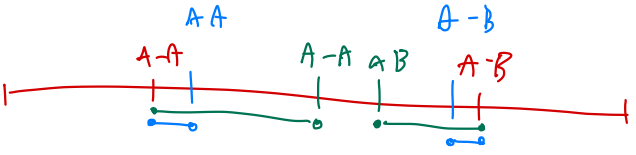
Digraph of Person A

Key Pair	Average Time
A A	0.2 ms
A B	0.8 ms
...	
...	

T → h → e
x sec y sec

Problem
- Use family member is fine





Z Z	0.21 ms

Exercise

Given a team of two persons. Use the knowledge of digraph and trigraph (explained by your professor) to create a template for you and your team mate. Please measure (at least) time between 2-consecutive keys and total time. Measure whether your software can correctly identify the person.

Please also answer the following questions.

1. How many words do we need to correctly identify the person?
2. Do you think this method is scalable? (to thousand persons) for either recognition system or identification system. Please provide your analysis.
3. Will you use this kind of authentication in your system? Please also provide a reason.

Hint:

Use python time.clock() to collect the time.

python keystroke ... smth