Long-term Active User Authentication Using Multi-modal Profiles

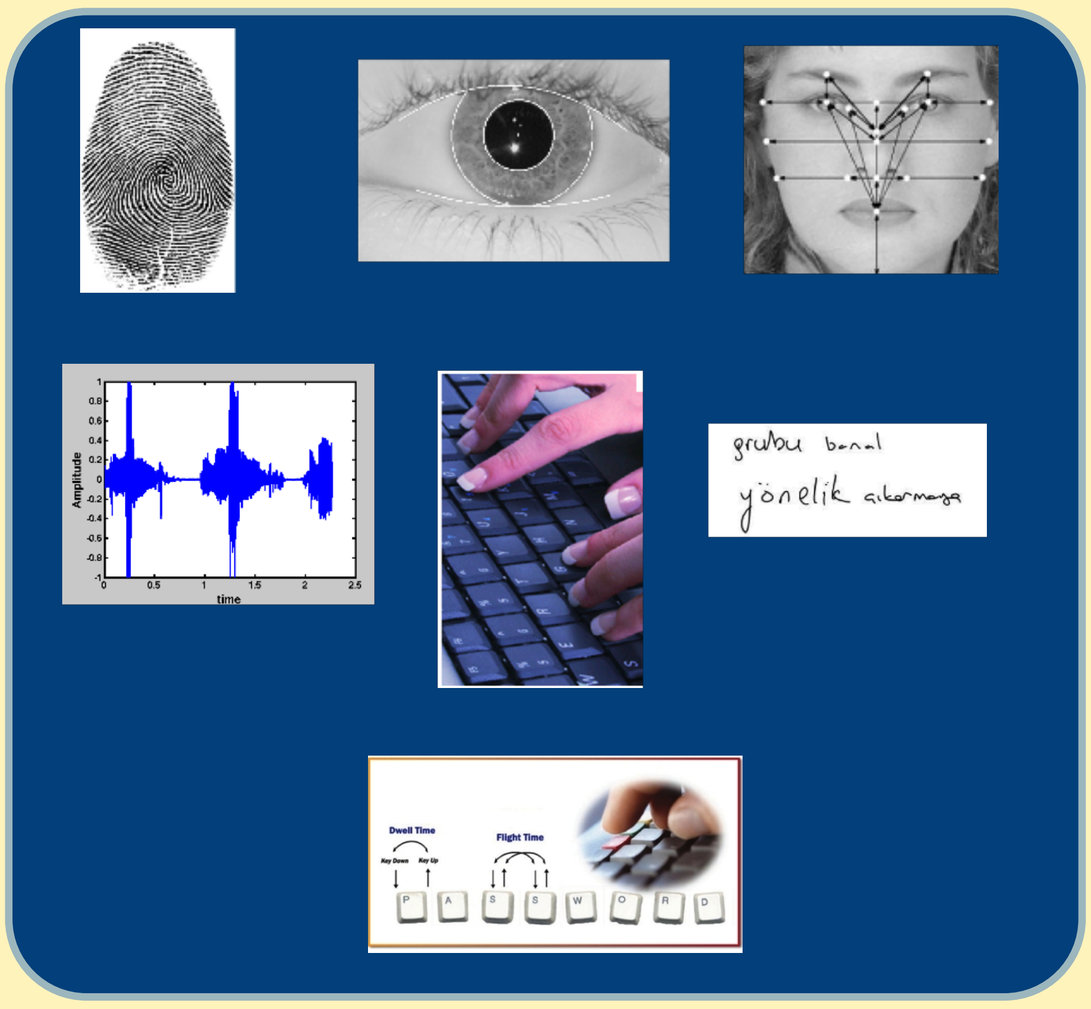
Stephanie Schuckers, Daqing Hou

Electrical and Computer Engineering Department, Clarkson University, Potsdam NY 13699

The goal of this project (with SUNY at Buffalo: Venugopal Govindaraju, Ifeoma Nwogu, Shambhu J Upadhyaya) is to develop a new framework for long-term, active user authentication, using multi-modal profiles that consist of physiological, behavioral, and cognitive biometric signatures.

The specific objectives of this project include the development of novel probabilistic biometric models as well as adaptive fusion algorithms that are capable of effectively adapting to changes in long term biometric signatures. Extensive usability tests are planned to ensure that the proposed framework can be effectively used in real-life computer and network systems.

In this collaborative project, Clarkson University has focused on keystrokes and mouse dynamics, and SUNY at Buffalo on the physiological and cognitive modalities.



Approach

- Establishment of large, shared datasets for keystrokes and mouse movements
  - Controlled, laboratory based data collection using browser based logger
  - Completely uncontrolled collection via installing logger on user's own computer
- Replication and improvement of state of the art algorithms using the established datasets
- Investigation of new algorithms

Clarkson Dataset 1: Controlled Keystrokes

Study	#subjects	Nature of text	Shared?
Joyce and Gupta [3]	27	short phrases (user names and passwords)	No
Killourhy and Maxion [4]	51	single password (All users type the same password ".tie5Roanl" 400 times over 8 sessions)	Yes
Loy, Lai, and Lim [8]	100	single password (recording both digraph latency and pressure. All users type the same password "try4-mbs" 10 times)	Yes
Lin [7]	125	passwords	No
Leggett *et al.* [5]	17	transcription of two proses of 1,400 and 300 chars each	No
Gunetti and Picardi [2]	40	free text (Italian, 40 subjects each contributing 15 samples; another 165 impostors each contributing one sample; sample length: 700-900 chars.)	Yes
Messerman *et al.* [9]	55	free text (multiple sessions spanning 12 months, a total of 3,000 to over 6,000 chars per subject)	Unkn
Clarkson University's Dataset (this paper)	39	mixed data of passwords, fixed text, and free text (two sessions spanning 11 months, on average 21,533 chars per subject)	Yes

Table 1: Characteristics of datasets from representative prior studies.

Statistics	#keys (Session 1)	#keys (Session 2)
average	11,066	10,467
stdev	2,358	1,823
min	2,948	4,731
max	14,184	13,069

Table 3: Statistics of raw keystrokes in Clarkson University's dataset. The minimum of 2,948 is considered an outlier due to a subject leaving without completing Session 1.

Clarkson Dataset 2: Uncontrolled Keystrokes

Study	#User	Time Span	Collection Setting	#Keystrokes	Data Availability
Dowland and Furnell [4]	35	3 months	Uncontrolled	3.4 M	NO
Gunetti and Picardi [5]	40	6 months	Browser	400 K	YES
Messerman et al. [7]	55	12 months	Predefined tasks	293 K	NO
Monaco et al. [8]	30	1-3 sessions	Fixed text	280 K	NO
Ahmed and Traore [1]	53	5 months	Uncontrolled	9.5 M	NO
Vural et al. [11]	39	2 sessions	Browser	840 K	YES
Ours	95	2 years	Uncontrolled	9.7 M	YES

Table 1. Comparison of datasets in literature

Impact of "size" on authentication

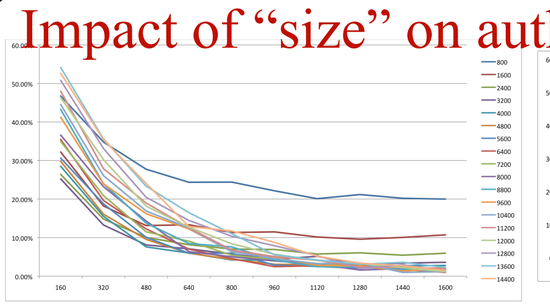


Figure 2. Effect of Test Sample Size on False Alarm Rate (FAR). X Axis Represents Test Sample Sizes (number of keystrokes). Y Axis Represents FAR Values. Each Line Represents the Results of Testing Reference Profiles of the Same Size, Ranging from 800 to 14,400 Keystrokes, against Test Samples of Different Sizes, Ranging from 160 to 1,600 Keystrokes.

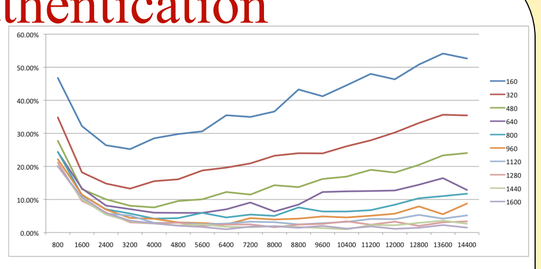


Figure 3. Effect of Reference Profile Size on False Alarm Rate (FAR). X Axis Represents Reference Profile Sizes. Y Axis Represents FAR Values. Each Line Represents the Results of Testing Test Samples of the Same Size, Ranging from 160 to 1,600 Keystrokes, against Profiles of Different Sizes, Ranging from 800 to 14,400 Keystrokes.

Impact of "Gibberish" text on authentication

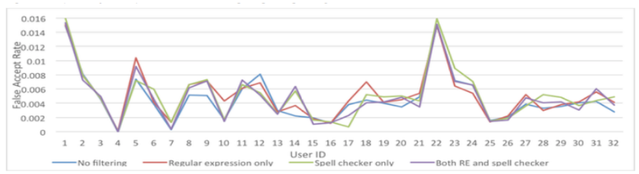


Fig. 4. False Accept Rate for different strategies of filtering gibberish text.

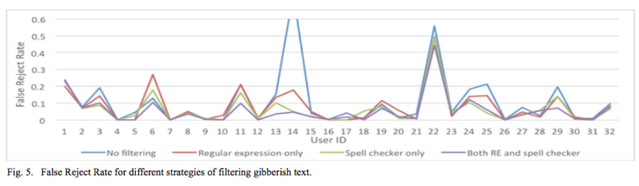


Fig. 5. False Reject Rate for different strategies of filtering gibberish text.

Interested in meeting the PIs? Attach post-it note below!

NSF Secure and Trustworthy Cyberspace Principal Investigators' Meeting
Jan. 9 -11 2017
Crystal Gateway Marriott Hotel, Arlington, Virginia

National Science Foundation
WHERE DISCOVERIES BEGIN

Clarkson UNIVERSITY defy convention

University at Buffalo
The State University of New York