Long-term Active User Authentication Using Multi-modal Profiles

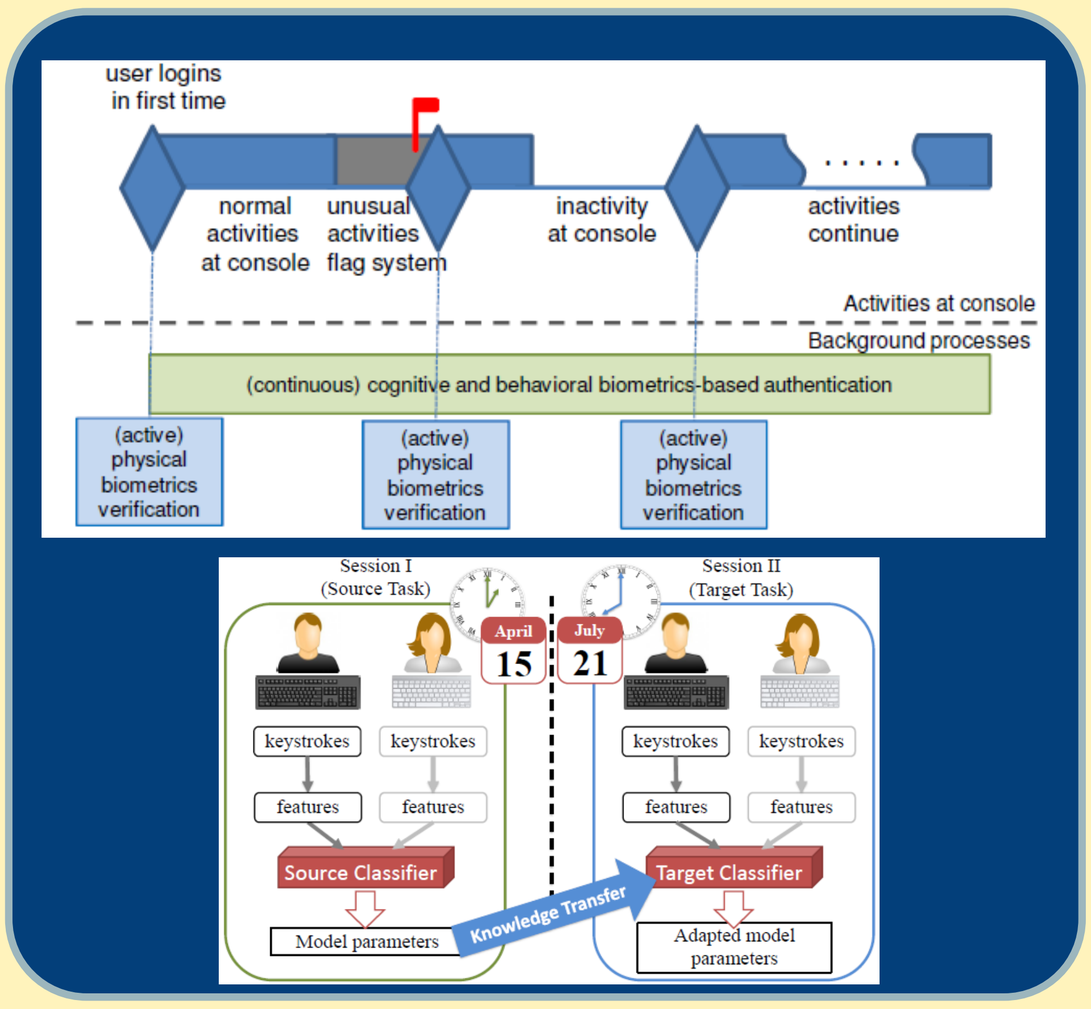
PIs: Venu Govindaraju, Ifeoma Nwogu, Shambhu Upadhyaya
Computer Science and Engineering, University at Buffalo

CNS-1314803

Project Objectives

Develop an adaptive process to advance the current state of identity authentication in computer systems, using different novel individual biometrics traits

An intelligent environment where a user identifies him/herself at a console and simply starts working, while **continuous authentication** occurs in the background, actively increasing the authentication processes when the **evidence is weak**, but invisible to the user, as he/she works without interruptions or performance compromises



a) Data collection
b) Analysis of secondary features of key stroke dynamics
c) Machine learning algorithms for biometric recognition
d) Development of language usage as a cognitive biometric

Approach

Shared Keystroke Dataset for Continuous Authentication

- One of the largest publicly accessible keystroke datasets for long text (> 300 users)
- Characterized to reflect
  - Temporal aspects of typing patterns
  - Effect of keyboard layout variability
- Textual data included
- Mouse movements and system events data

Related Work

	#Subject	#Sessions	Duration	Gap b/w Sessions	Clock Resolution	Keyboard variability	Gender (M:F)	Age
Clarkson	39	2	1 hour	Mostly 1 or 2 month	-	-	-	-
MSU P1	51	2	10 – 16 min	Same day	-	-	-	-
MSU P2	30	Around 5	60 sec	-	-	-	-	-
Ours	148	3	50 min	28 days	15 ms	Yes	113 : 35	20-30

Our Experiment

A large scale data collection campaign • 4 months from Sept. to Dec. 2015 • 157 volunteers recruited • 2 keystroke activities involved • Transcribed and free text • 3 sessions for each participant • 50 minutes for each session • Approximately 1 month between sessions • 4 different types of keyboards utilized

Data Anonymization and Quality Assurance

Privacy protection

- Randomly generated ID; dummy email accounts
- Built-in rule based sanitization functions
  - Remove username, email address, phone numbers

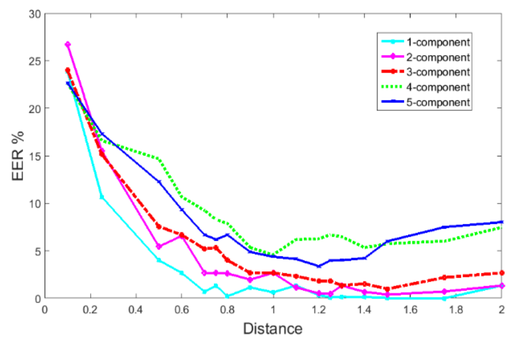
Incomplete data files removed

- 9 subjects removed
- 157 subjects −> 148 subjects in first round

Data Quality

Experiments on digrap using GMM

- Create a reference profile for each user using training feature set
- Test records compared against reference profile to measure similarity score
- EER comparison
- 0.01% (ours), 0.08% (Clarkson)



# Components	1	2	3	4	5
EER (%)	0.01	0.39	0.98	4.58	3.37

Dataset Statistics

Number of raw keystrokes

- 5,700 keystrokes per session; Minimum 10,000 overall
- 17,000 keystrokes each subject

Time intervals between sessions

- 28 days on average

Gender information

- Female 35, Male 113

Publications

Y. Sun and S. Upadhyaya, "Secure and Privacy Preserving Data Processing Support for Active Authentication", *Information Systems Frontiers*, Volume 17, Issue 5, pp. 1007-1015, October 2015.

H. Çeker and S. Upadhyaya, "Enhanced Recognition of Keystroke Dynamics using Gaussian Mixture Models", *MILCOM* 2015, pp. 1305– 1310, 2015.

H. Çeker and S. Upadhyaya, "User Authentication with Keystroke Dynamics in Long-Text", *IEEE BTAS*, Niagara Falls, NY, 2016.

H. Çeker and S. Upadhyaya, "Adaptive Techniques for Intra-User Variability in Keystroke Dynamics", *IEEE BTAS*, Niagara Falls, NY, 2016.

Y. Sun, H. Çeker and S. Upadhyaya, "An Anatomy of Secondary Features in Keystroke Dynamics - Achieving More with Less", *8th WIFS 2016*, Abu Dhabi, UAE, December 2016.

N. Pokhriyal, I. Nwogu and V. Govindaraju, "Cognitive-Biometric Recognition from Language Usage: A Feasibility Study", *IEEE Trans. Information Forensics and Security*, 12 (1), Jan. 2017.

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