Louisiana Tech University

Vir V. Phoha, Md E. Karim

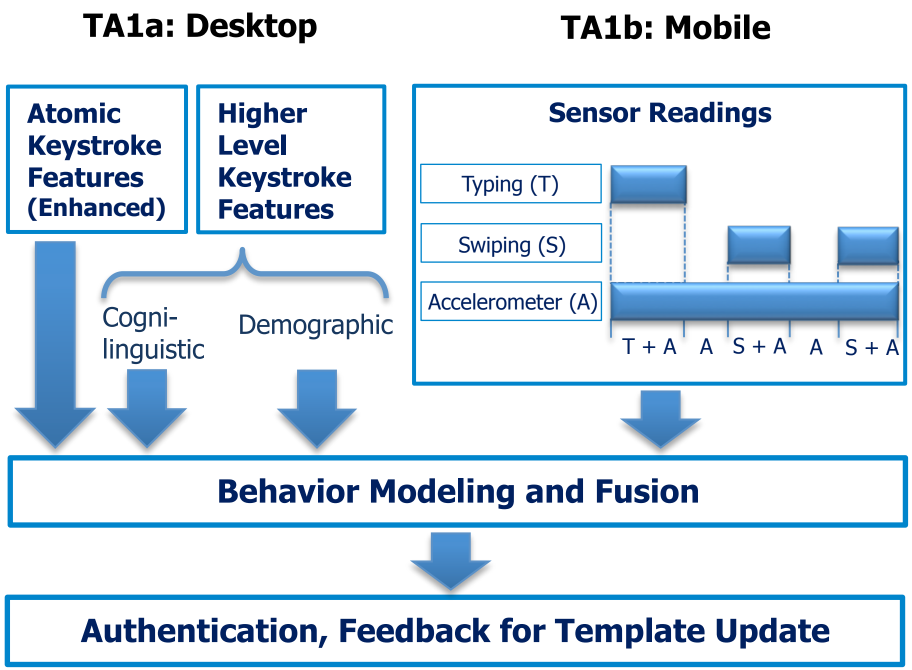
Active Authentication PI Meeting

Oct 30, 2014

This research was developed with funding from the Defense Advanced Research Projects Agency (DARPA). The views, opinions, and/or findings contained in this article/presentation are those of the author(s)/presenter(s) and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

Louisiana Tech University

BIOMETRIC: (Desktop) Atomic keystroke latencies enhanced with word context, Cogni-linguistic/ Demographic features; (Mobile) Typing/Swiping features, Body movements



Experiment:
(Desktop) Volunteer participants: 831

(Mobile) Volunteer participants: typing 74, swiping 47 and body movements 11

Experiment Results

- ✓ **(Desktop) Atomic Keystrokes:** Accuracy 96.641%, FAR .0295, FRR 0.0382

The above results were obtained using typing data collected from volunteers in a single application context. System accuracy may change with passively collected data from multi-application, multi-window environments.

- ✓ **(Mobile) Typing:** FAR: 11.80%, FRR: 12.60%
  Swiping: FAR: 10.30%, FRR: 10.30%
  Body Movements: FAR: 3.20% , FRR: 3.20%

TA1a: Desktop Design and Development of a Suite of Keyboard-based Biometrics for an Active Authentication System

TA1b: Mobile Context-aware Active Authentication using Touch Gestures, Typing Patterns, and Body Movements

TA1a: Desktop

Refinement & Integration
Md E Karim

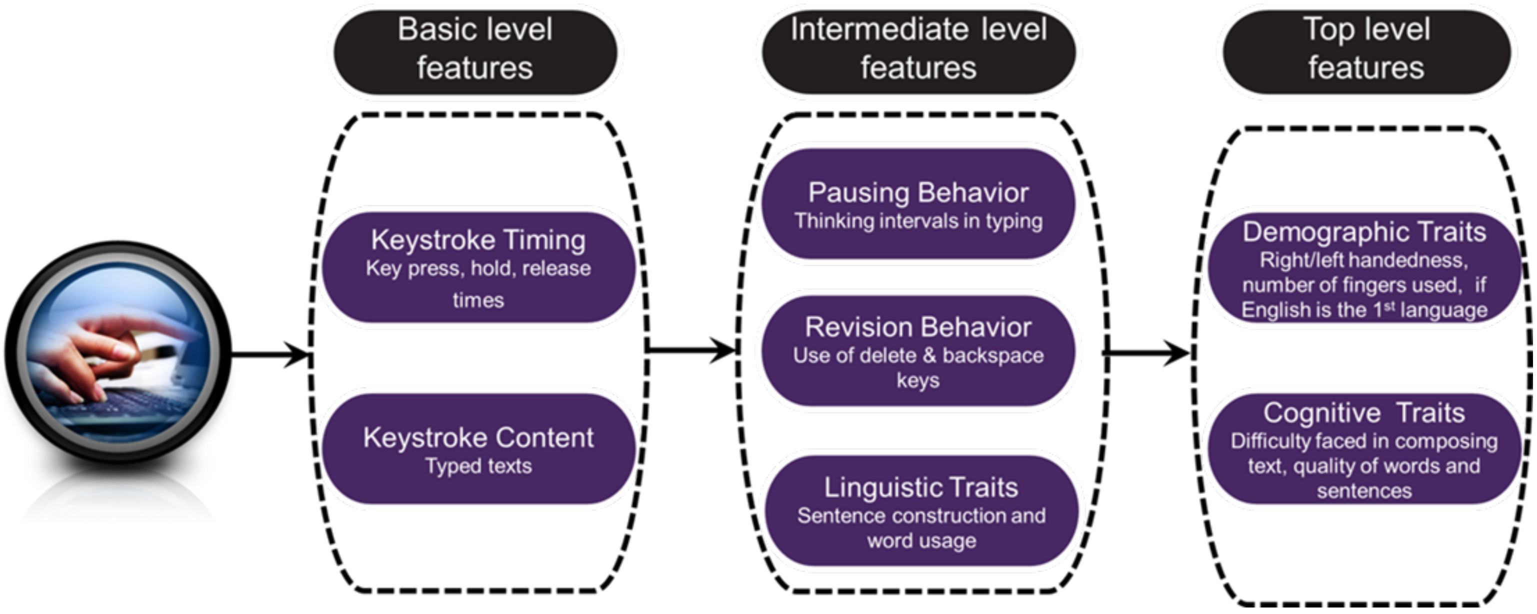
TA1b: Mobile

Final Findings
Vir V Phoha

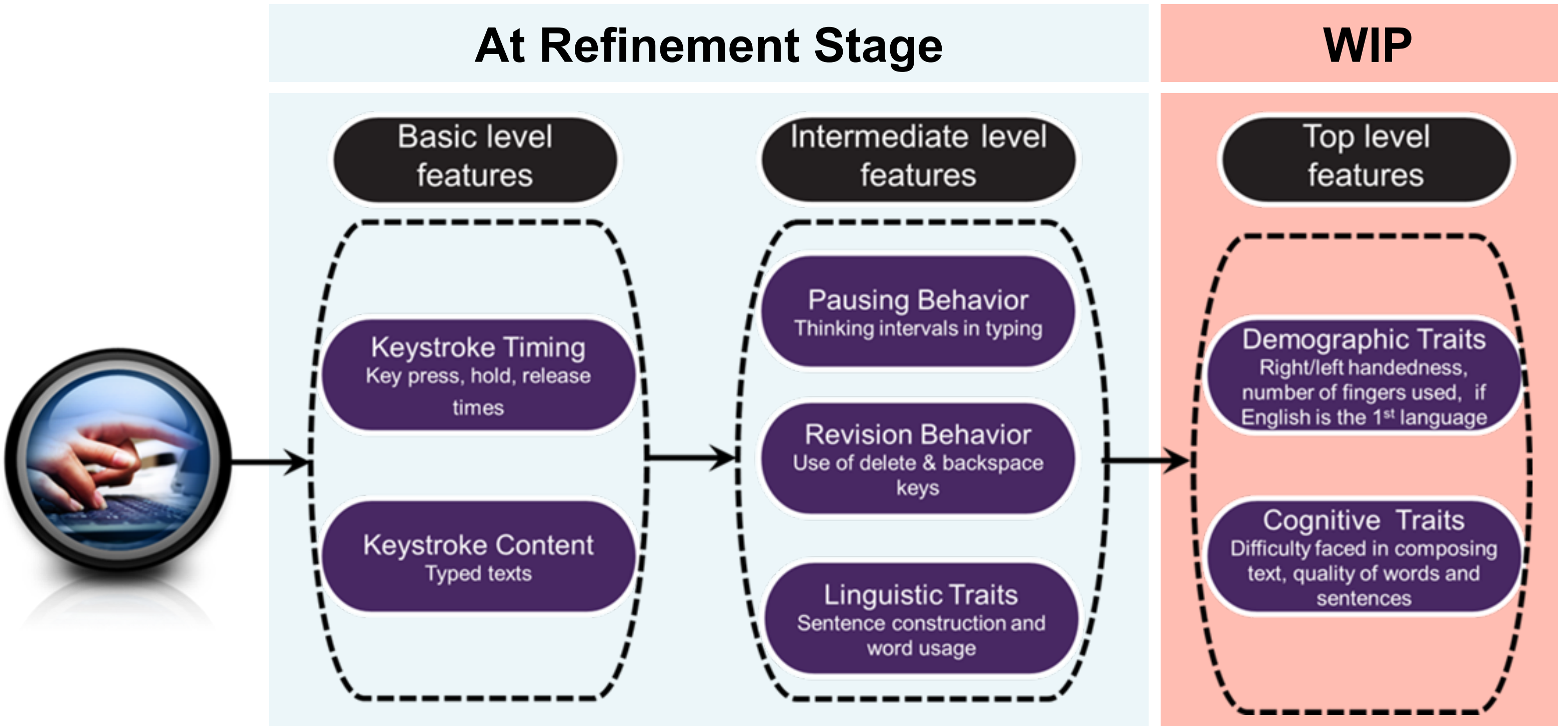
TA1a: Desktop

LOUISIANA TECH UNIVERSITY
cyber innovation center
Collaboration • Research • Technology
CUNY NYIT {AEGIS} RESEARCH LABS SU

Overall Scheme



Current Status



Challenges

- **One Size Does Not Fit All**

Initial investigation shows that the system accuracy highly depends on the individual-specific tuning of system parameters.

- **Multi-application Environments**

Work to date has focused on data collected in a single application (question and answer) environment which is not reflective of the real life scenarios. Data passively collected from West Point's freshmen class will become available shortly which will allow us to begin looking at system performance in multi-window, multi-application environments.

- **Resistance to Spoof Attacks**

If someone's typing data were captured by a key-logger it might be possible for an attacker to create a program that simulated the atomic keystroke behavior of that user. We are exploring the use of pause and revision behavior, and demographic and cognitive traits (extracted from the keystroke input) to detect many such attacks.

- **Time Drift**

We are also investigating the effect of temporal drift in users' typing behavior on authentication accuracy using a dataset collected over a period of 3+ years.

Current System Performance

Accuracy:

✓**Basic Level Features: 96.641%** for decisions made approximately every 10-15 seconds of continuous typing (initial decision takes 100 - 200 seconds)

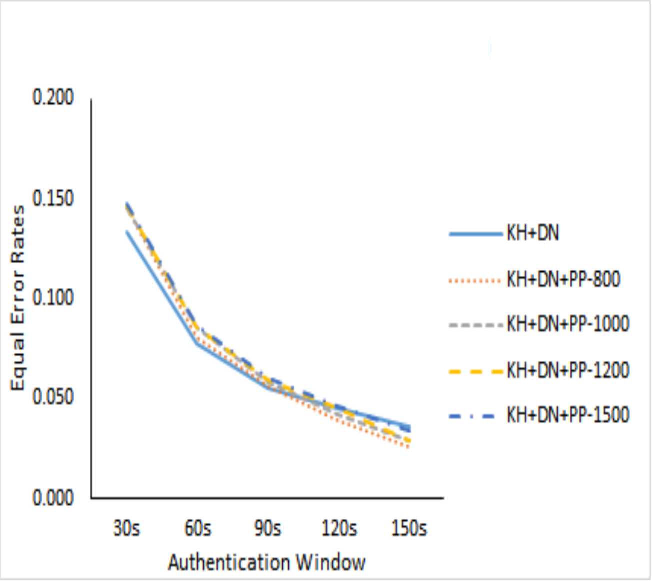
❖ Smaller scale tests with user specific thresholds show an average 2% improvement in accuracy.

✓**Basic + Intermediate Level Features: 97.36%** for decisions made every 150 – 200 seconds of continuous typing (100 volunteer participants)

✓**Top Level Features:** Work in progress.

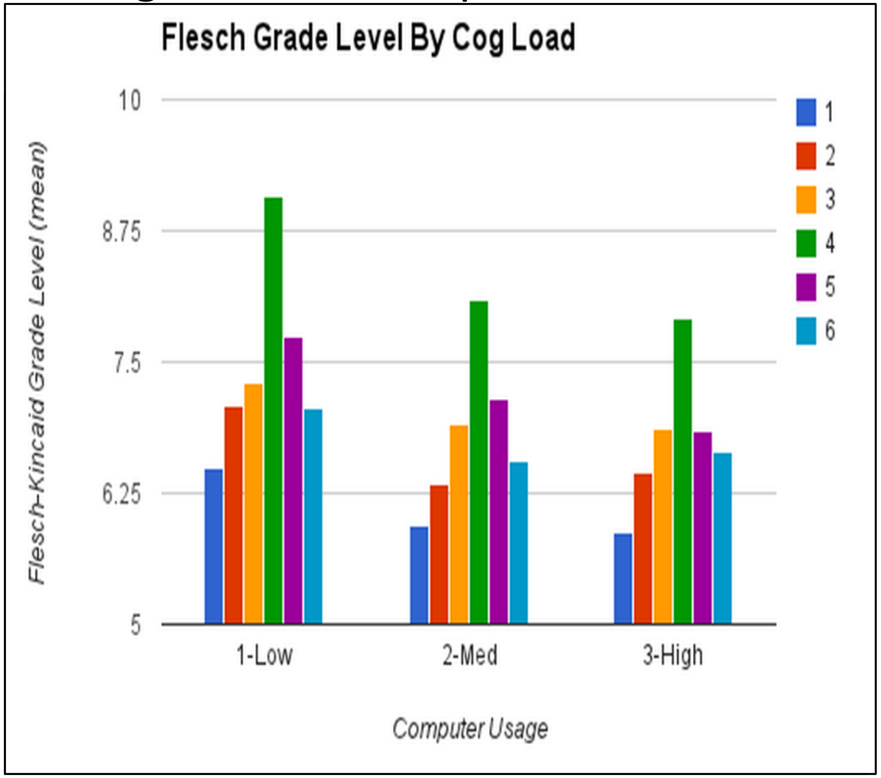
- The above results are obtained using typing data collected from volunteers in a single application context. System accuracy with passively collected data from multi-application, multi-window environments may be lower.
- Time before decision will vary with typing speed; above numbers are typical of an average typist.

Fusion EERs of PP-Burst with Keystroke Dynamics



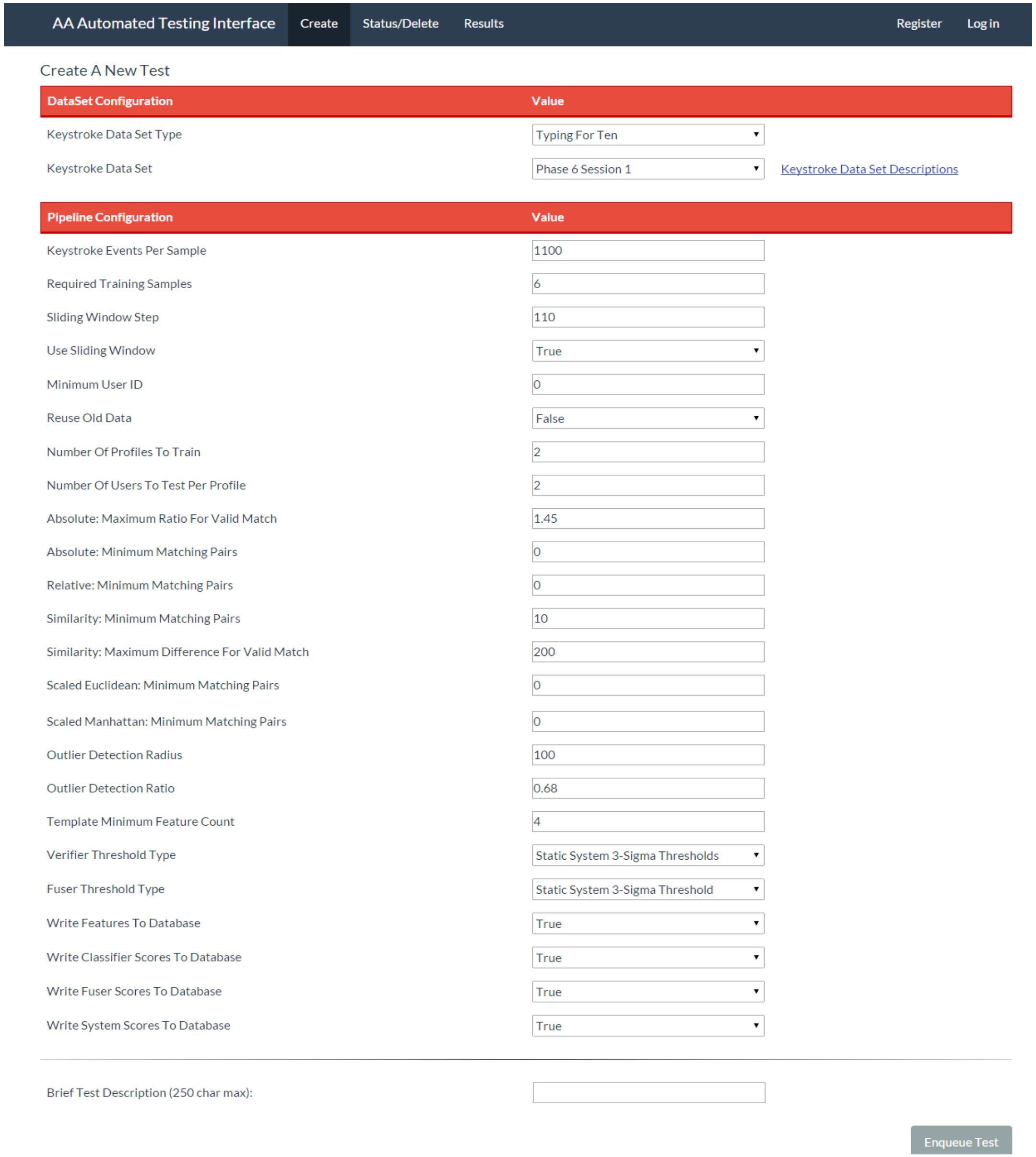
Relationship between Cognitive Load, Computer Usage and Lexical Complexity

- How does the typing proficiency and cognitive load interact and impact a typist's behavior?
  - Typing proficiency is based on subject's self reports of how many hours per day they use a computer. Bottom 25% - Low, Middle 50% - Middle, Top 25% - High.
  - Cognitive Load is based on the expected cognitive load required to respond to a prompt.
- Mean typing burst (words) Increases with computer usage.
- Mean Typing burst is impacted by cognitive load, though not linearly.
- The lexical complexity (Flesch-Kincaid grade level) of the generated text varies with cognitive load, but not linearly.
  - Under higher and lower cognitive loads, Lexical complexity is lower. Moderate cognitive load leads to the most complex responses.

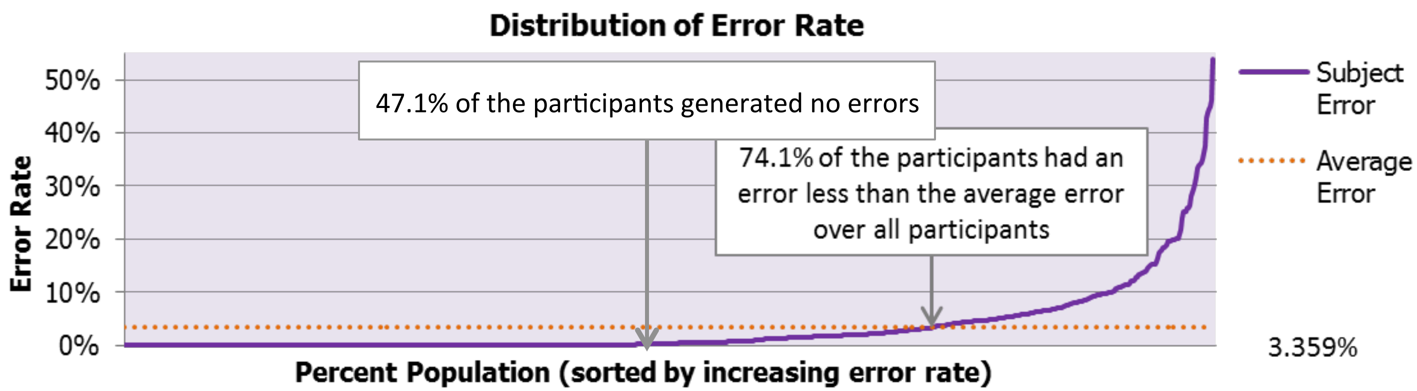


Web-Based Testing Environment

- Multi-tiered testing environment with a centralized database, web front end and different test running on individual server at the back end
- Allows the selection of tests, displays the system configuration matrix, user specific confusion matrices, and classifier confusion matrices.
- Additional features of the system include data reuse for data persisted from various points of the processing pipeline



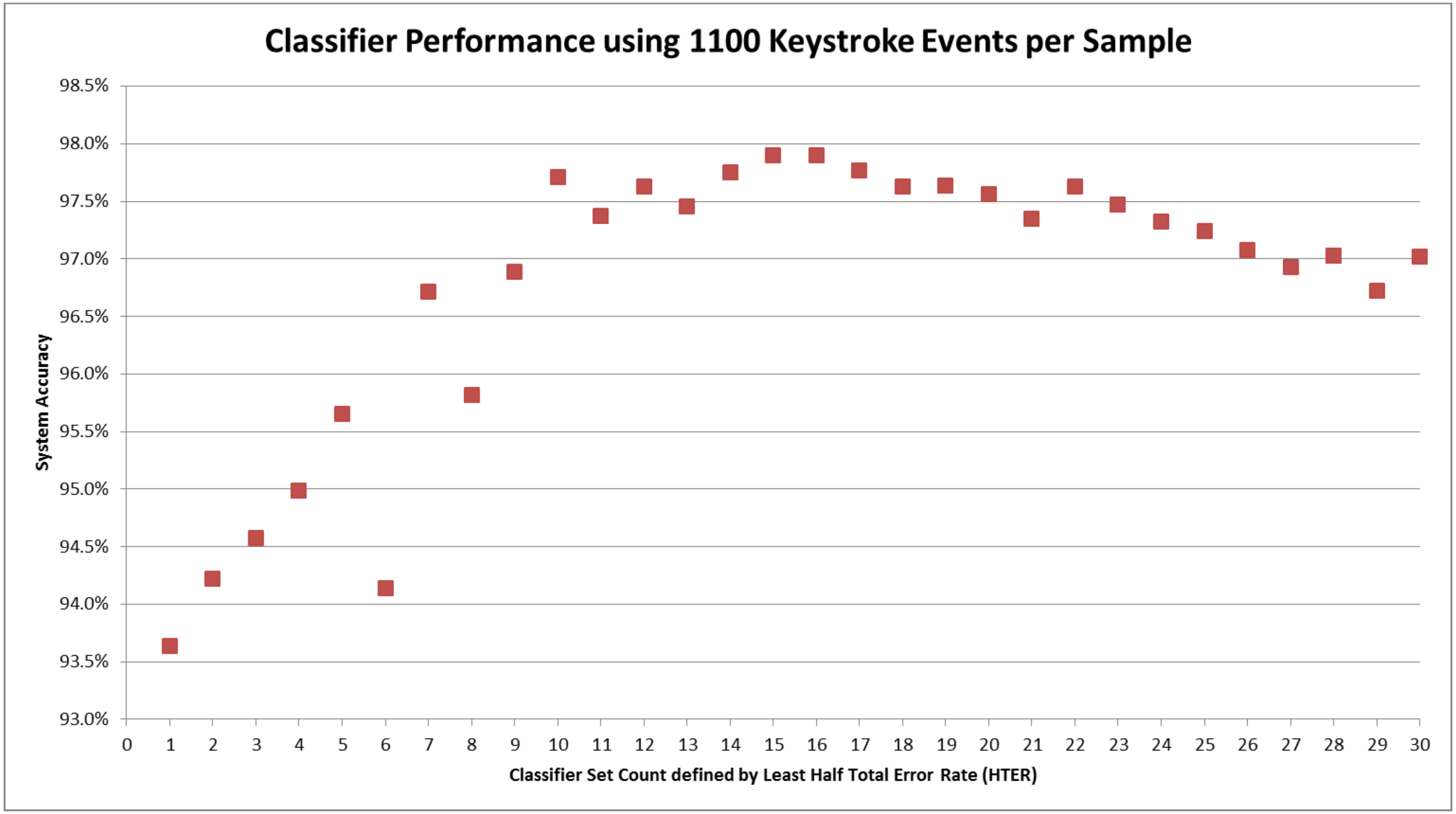
User-Specific VS Population/Global Threshold



	Absolute							
	IK		KH		KPL		KRL	
	User_Sp	Global	User_Sp	Global	User_Sp	Global	User_Sp	Global
FPR	0.156093	0.204012	0.16189	0.235035	0.098446	0.166235	0.076458	0.128454
FNR	0.162241	0.215586	0.17836	0.117013	0.101937	0.173721	0.087325	0.1496
Accuracy	0.843873	0.795925	0.838019	0.765624	0.901535	0.833724	0.923482	0.871428

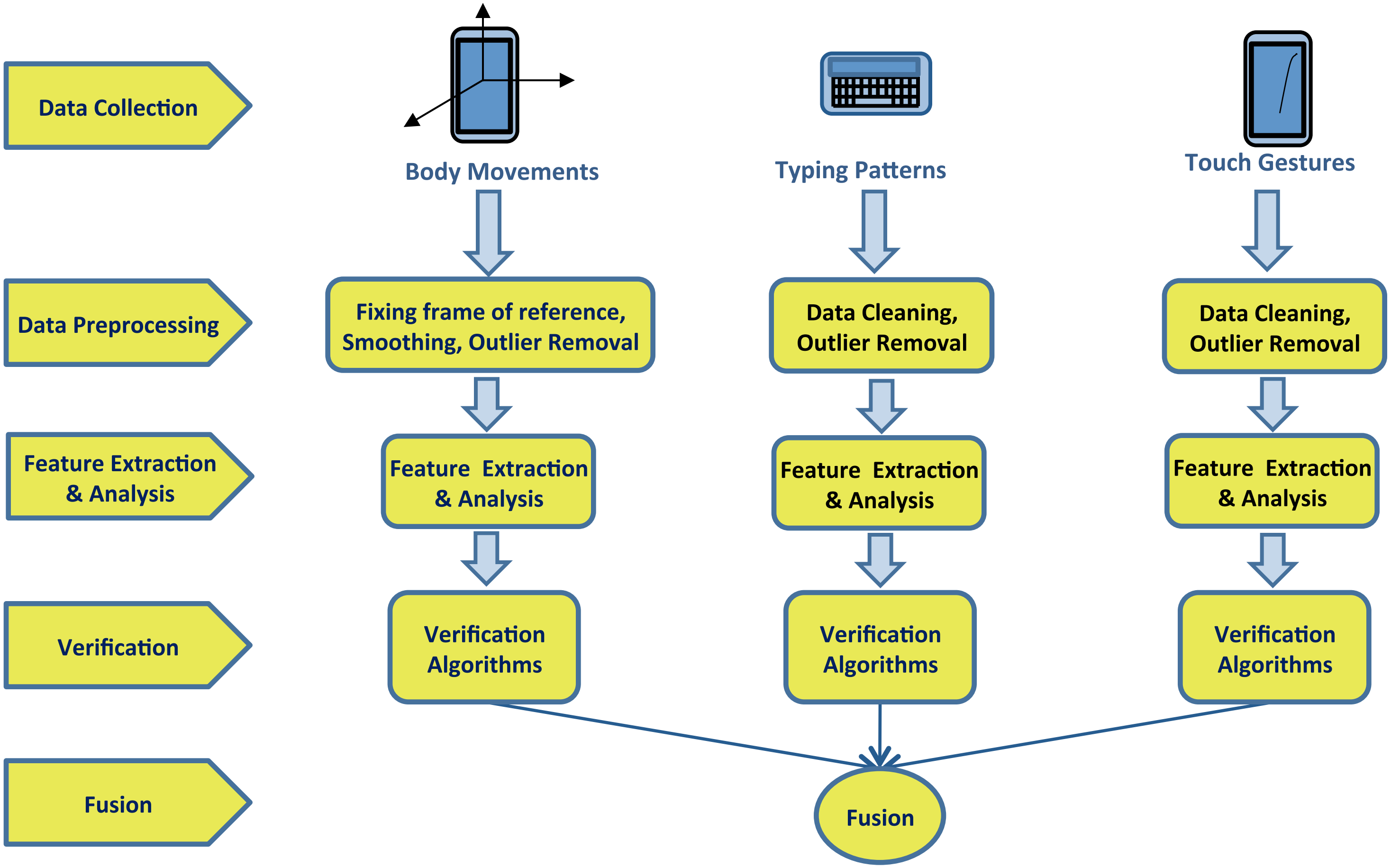
	Relative							
	IK		KH		KPL		KRL	
	User_Sp	Global	User_Sp	Global	User_Sp	Global	User_Sp	Global
FPR	0.051546	0.071599	0.080002	0.113914	0.095538	0.120356	0.063338	0.082524
FNR	0.051258	0.085121	0.087093	0.125014	0.094167	0.131509	0.061	0.091731
Accuracy	0.948457	0.928327	0.919959	0.886025	0.90447	0.879582	0.936675	0.917426

Classifier-Ensembles VS Performance



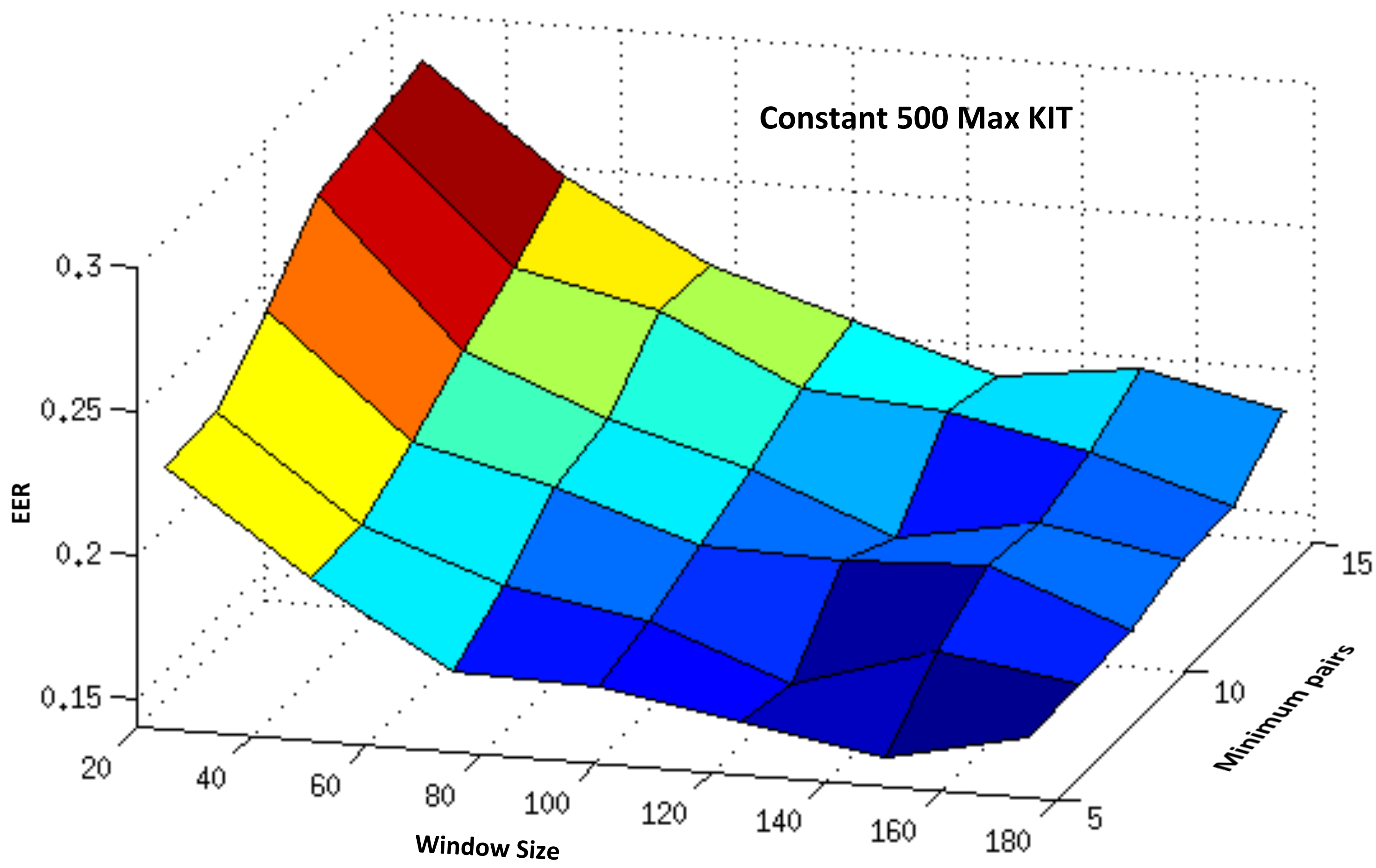
TA1b: Mobile

Tri-modal Fusion

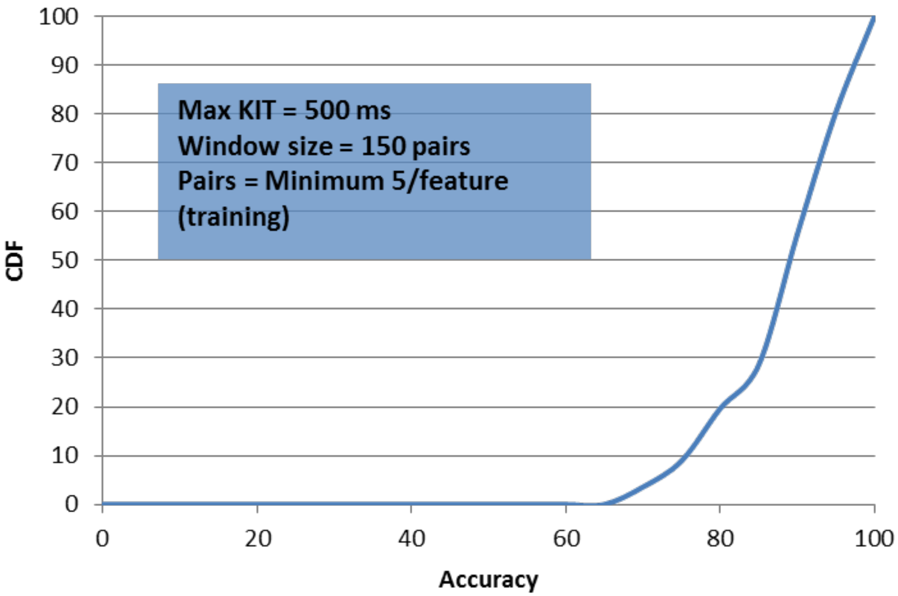


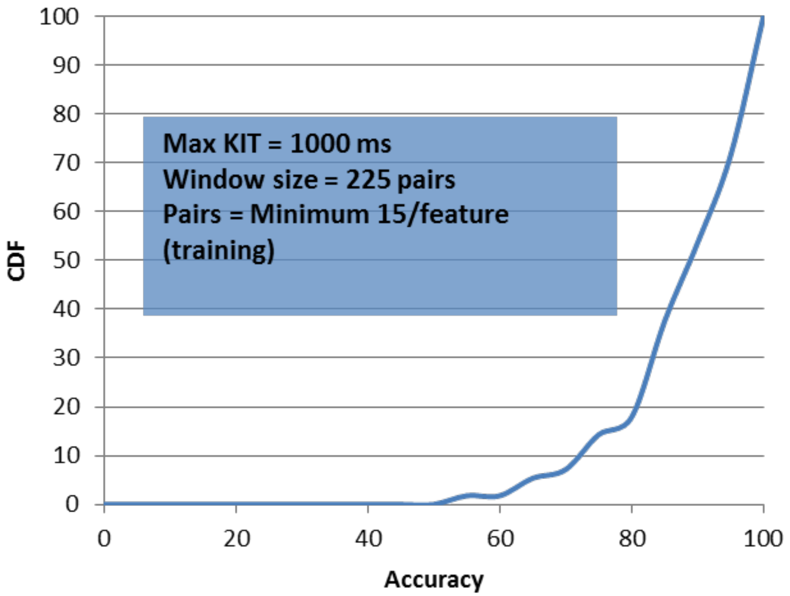
- "*Beware*, Your Hands Reveal Your Secrets" Accepted in the ACM Conference on Computer and Communications Security (ACM CCS), 2014, Scottsdale, Arizona, USA.

Typing: EER Sensitivity Analysis for Different Parameter Settings



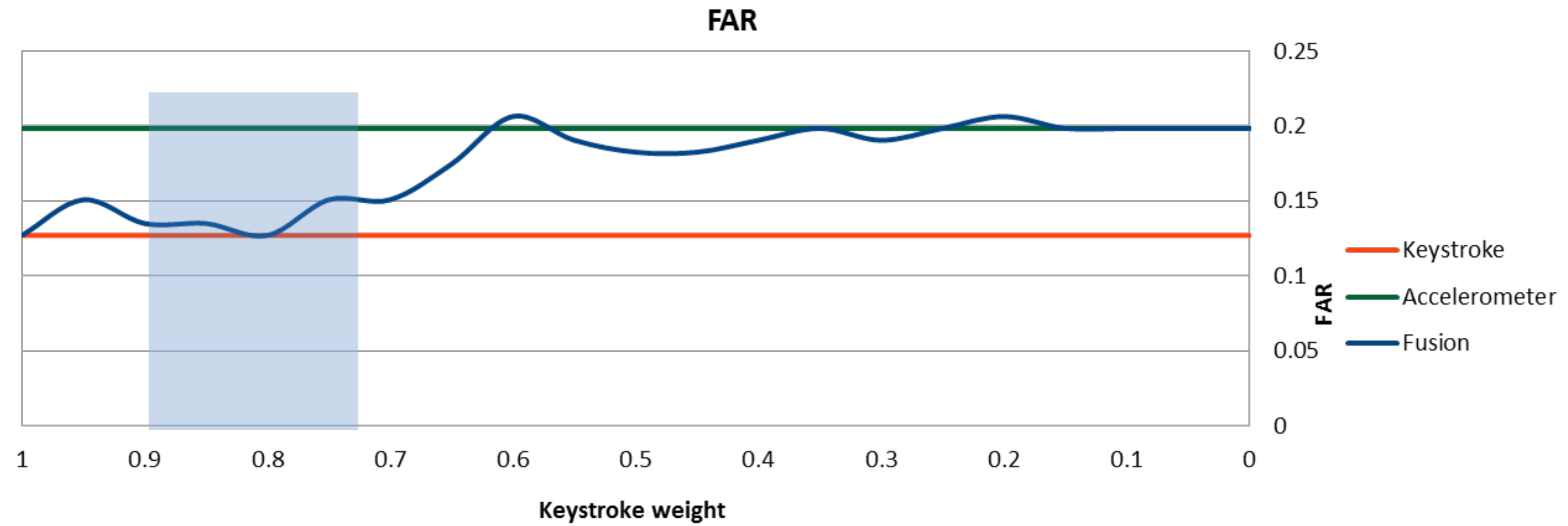
Typing: Classification Accuracy for Selected Parameter Settings

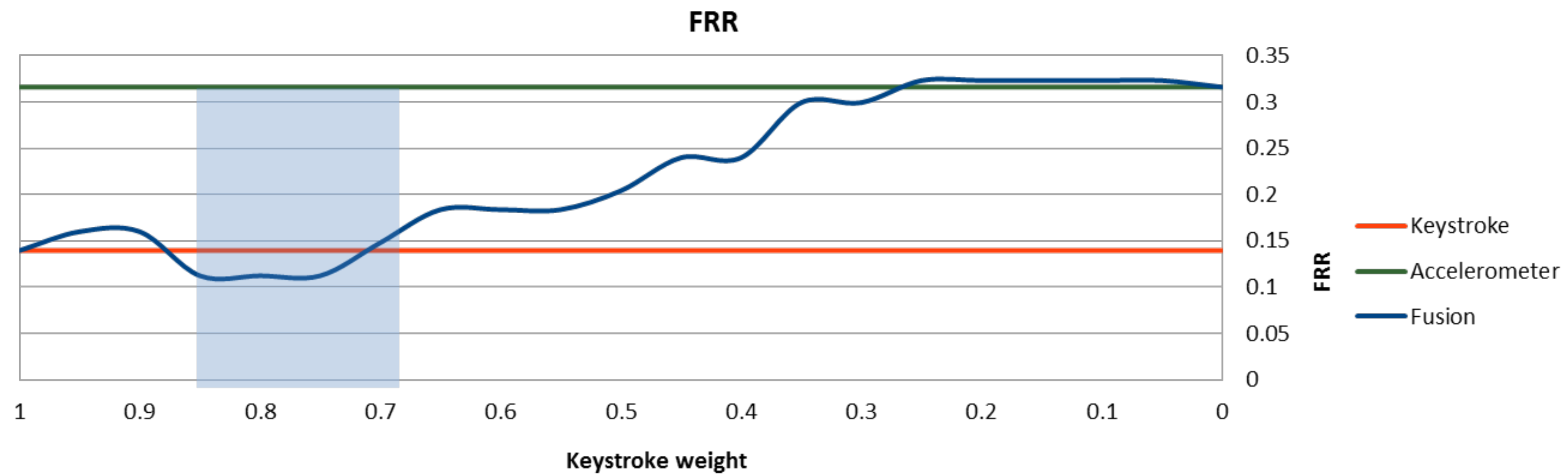




Max KIT	Window Size	Min Pairs	FRR	FAR	% Accuracy
1200	225	5	0.142	0.136	86.5
1100	225	5	0.135	0.129	87.0
1300	225	9	0.136	0.131	86.9
800	200	7	0.130	0.133	86.9
700	200	9	0.112	0.135	87.2
1000	200	5	0.154	0.135	86.1
1000	225	11	0.133	0.138	86.7
400	100	5	0.133	0.153	85.5
1300	225	11	0.146	0.141	86.0
900	75	5	0.141	0.164	84.8

Modality Fusion: Typing + Accelerometer

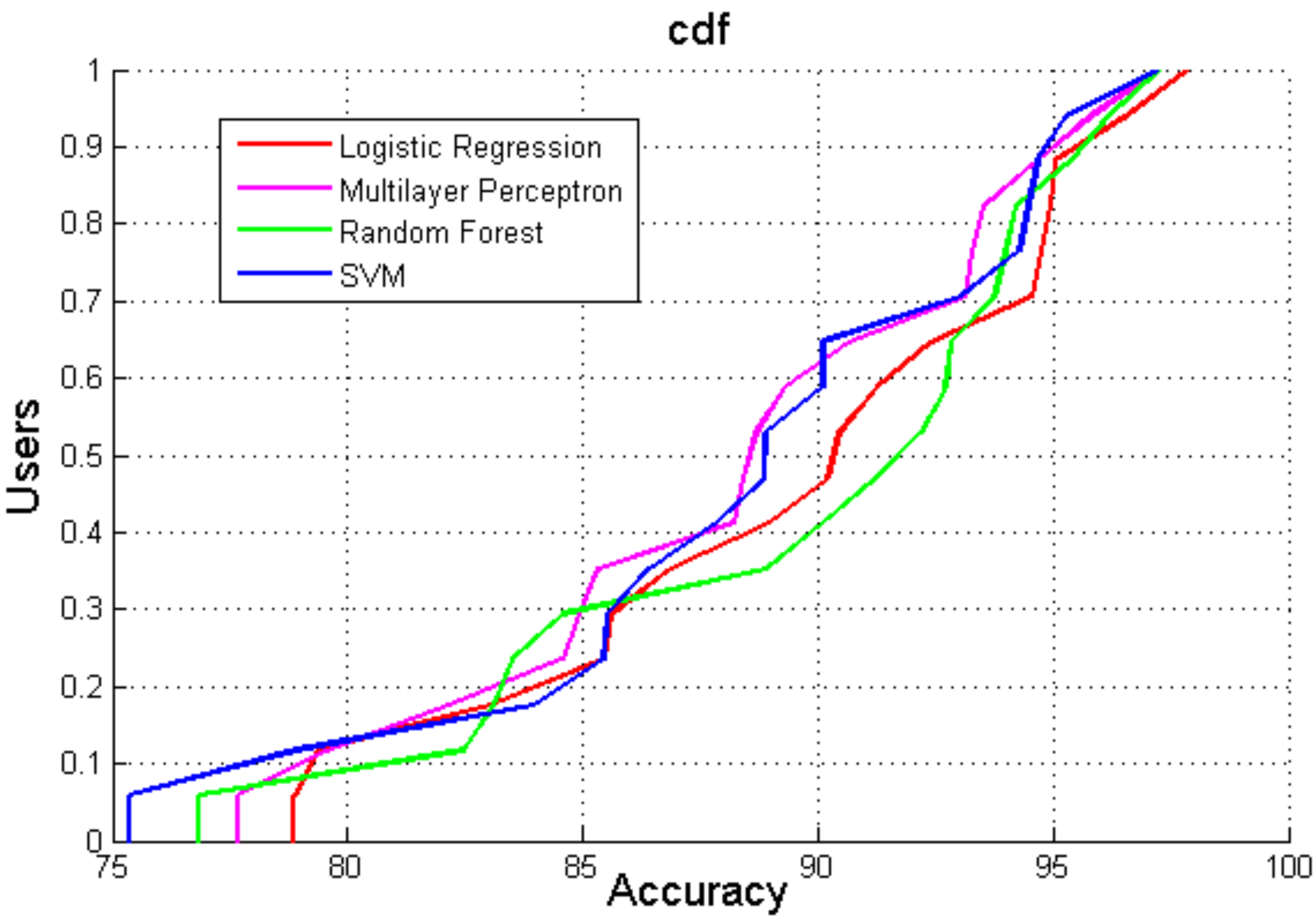




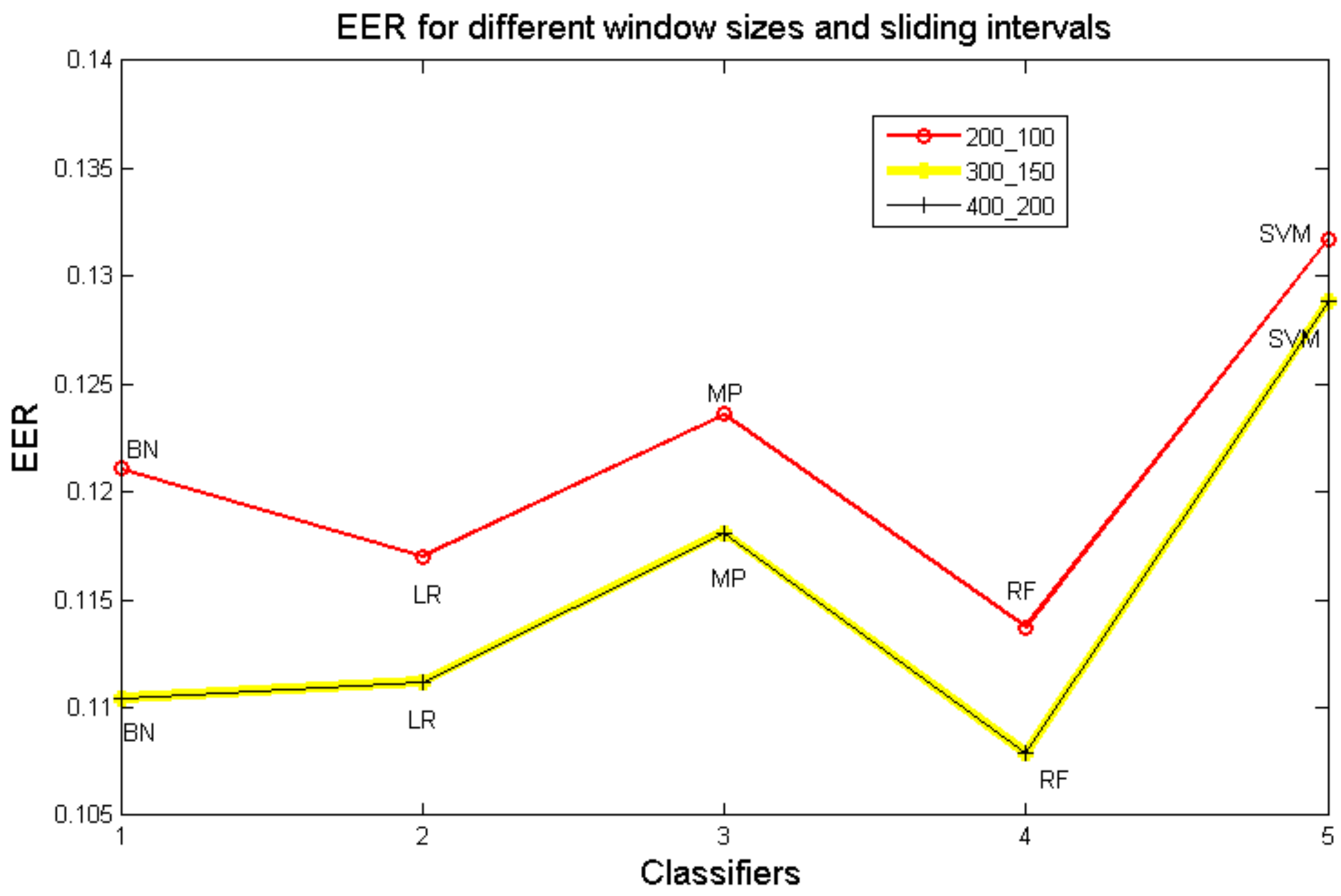
Gait: Different Verifiers Performance

Verifier Name	FAR (%)	FRR (%)	Accuracy (%)
Bayes Net	**10.15**	**15.72**	**87.83**
Logistic Regression	**13.86**	**08.26**	**89.24**
Multilayer Perceptrons	**16.23**	**08.65**	**88.19**
Random Forest	**12.16**	**10.63**	**89.33**
SVM	**13.86**	**11.10**	**88.37**

Gait: Distribution of accuracies across the population



Gait: EER for different window sizes and sliding intervals



Accelerometer Data Sampling Rate: 40-45 samples per second

Thank You !