CODASPY - IWSPA 2017

Continuous Authentication Using Behavioral Biometrics

Shambhu Upadhyaya, SUNY at Buffalo

March 24, 2017

University at Buffalo The State University of New York

Acknowledgments

NSF Grant No. CNS-1314803 (SaTC Medium)

Clarkson University (Stephanie Schuckers and Daqing Hou)

Yan Sun, PhD student at UB

Hayreddin Ceker, PhD student at UB

Motivation

Desktop system authentication

- Password based
  - Creation, memorization and management daunting task
  - Typical systems do not guarantee the legitimacy of the person at the console
  - Leads to masquerade/impersonation attacks

Mobile device authentication

- Pin or patterns
  - Easily revealed by shoulder surfing

What is the solution?

- Just get rid of it!
- A single ignorant person can risk the entire system!

DARPA's Initiative in 2012

It all started at DARPA

Click here to receive GCN magazine for FREE!
inShare

DARPA: Dump passwords for always-on biometrics

- By Kathleen Hickey
- Mar 21, 2012

The Defense Advanced Research Projects Agency wants to eliminate passwords and use an individual's typing style and other behavioral traits for user authentication.

Click here to receive GCN magazine for FREE!
inShare

Why so many bad passwords? Because the rules allow them.

- By Kevin McCaney
- Mar 12, 2012

DARPA's Active Authentication Program

Active Authentication BAA-12-06

- March 6, 2012 for Phase I
  - Develop novel ways to authenticate using unique aspects of individual (Biometrics)
  - Use observables on how we interact with the world (Behavioral Biometrics)
  - Use of software-based biometrics
  - As a first step, do not use any additional hardware

Concept of "cognitive fingerprint"

- Pattern based on how our mind processes information
  - Use multiple modalities
  - Accuracy, robustness and transparency

DARPA's Advanced Program

Thrust I

- Goal is to deploy the new authentication platform on a DoD desktop or laptop

Thrust 2

- Securing mobile devices

Defense Advanced Research Projects AgencyNews And Events

Where DARPA is Going, You Don't Need Passwords

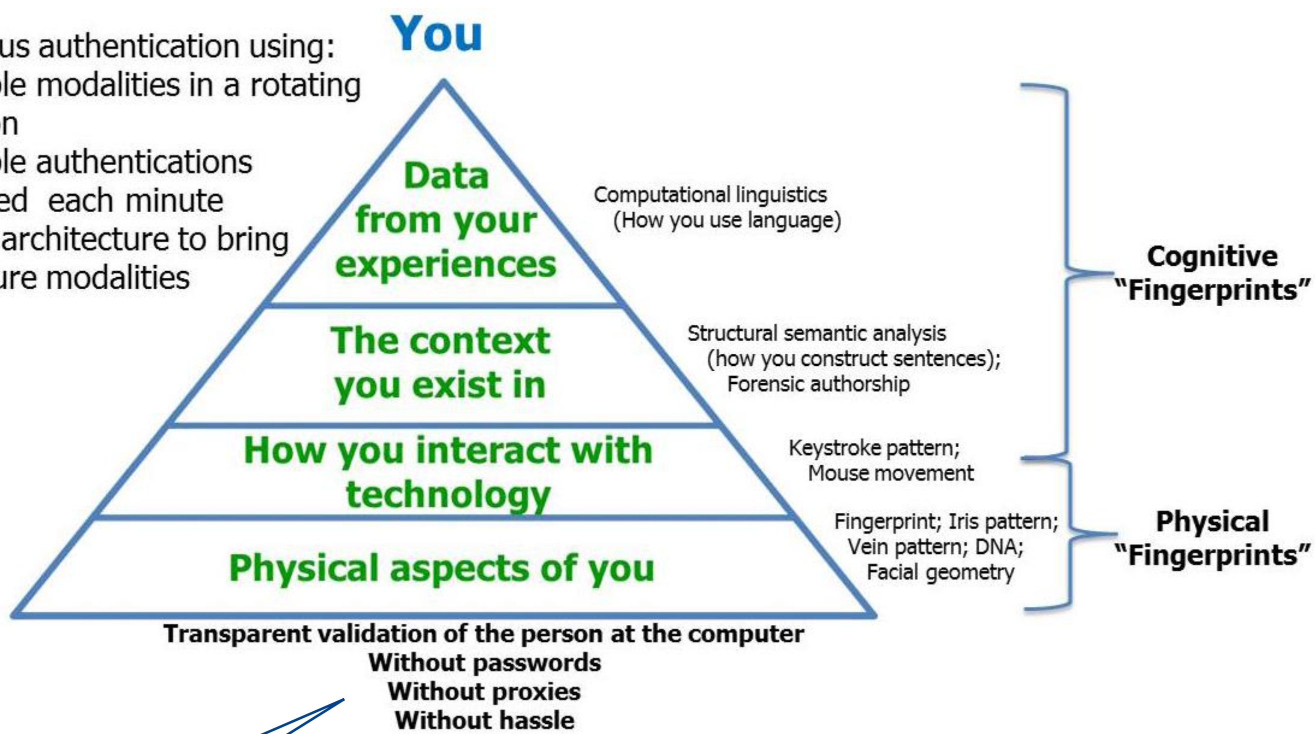
Active Authentication program investigates behavioral biometrics for mobile devices

outreach@darpa.mil
2/12/2013

DARPA's Vision of Continuous Authentication

Continuous authentication using:

- Multiple modalities in a rotating fashion
- Multiple authentications initiated each minute
- Open architecture to bring in future modalities



Courtesy: DARPA

Start here

Biometrics

Metrics related to human characteristics

- Physical
  - Fingerprint
  - Face
  - Iris, etc.
- Behavioral
  - Keystroke
  - Gestures
  - Voice
  - How user searches for information
  - How user reads material, etc.

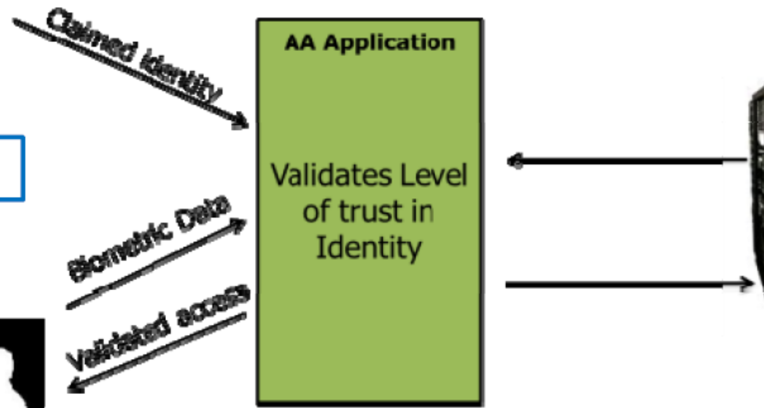
DARPA's Funded Programs

Iowa State
Stylometry focused on thought processing time

Drexel
Stylometry augmented by author classification and verification

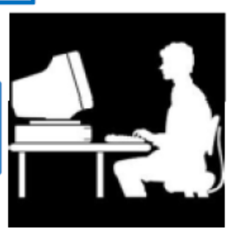
NY IT
Stylometry how a user types – ignoring the words

NPS
Behavioral manifestations of human thought processes



BehavioSec
Keystroke, mouse, in context

NRL
Identification of users through Web browsing behavior



UMD
Information processing from computer screens

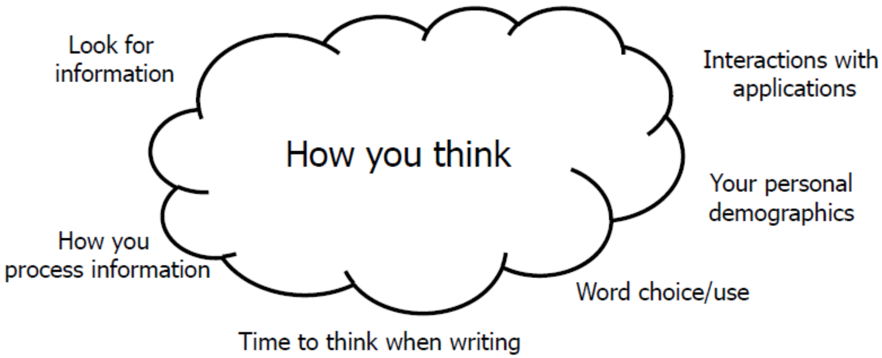
SWRI
Use covert games disguised as computer anomalies

Coveros
User Behavior Patterns as seen from the Operating System

Allure Security
User search behavior characteristics

Courtesy: DARPA

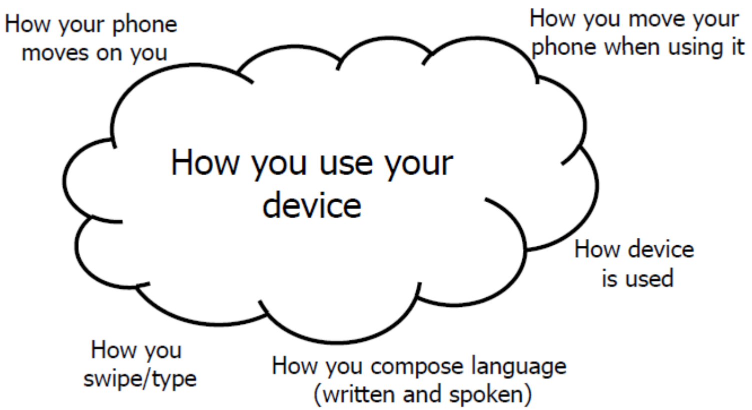
DARPA's Phase I Results

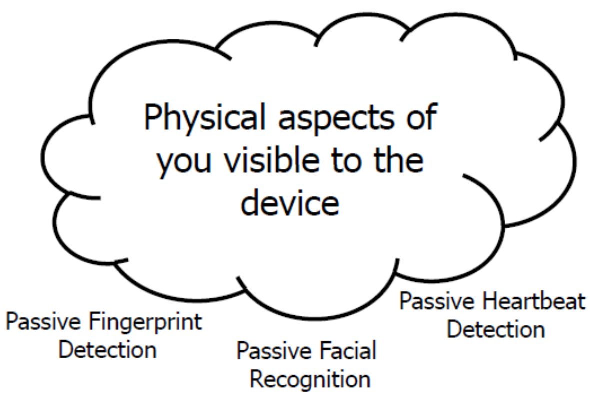


Phase 1 Results

	TP	FAR	Time
Allure Security	95.0%	1.0%	5m
Iowa State U (KRR)	92.7%	5.5%	29sec
NYIT	92.0%	4.0%	1min
Drexel U	92.0%	5.0%	50sec
University of MD	82.8%	20.0%	83sec
NRL	82.0%	6.0%	4hrs
Coveros	80.0%	10.0%	30sec
SWRI	75.0%	25.0%	8min
Alenka Brown	10.1%	10.0%	1min

TP = True Positive Rate
FAR = False Accept Rate
Time = time before decision





Courtesy: DARPA

Focus of the Talk

Keystroke dynamics as behavioral biometrics

- Short text
- Long text

Short text keystroke dynamics

- Generally useful for one-time authentication

Long text keystroke dynamics

- Necessary for continuous authentication

Rest of the talk will focus on this

Outline of the Talk

Introduction

- General approach to continuous authentication

Keystroke dynamics and mouse movements

- Feature selection
- Methodology - Gaussian model, SVM, transfer learning
- Datasets and anonymization

Results

- GMM, SVM, transfer learning

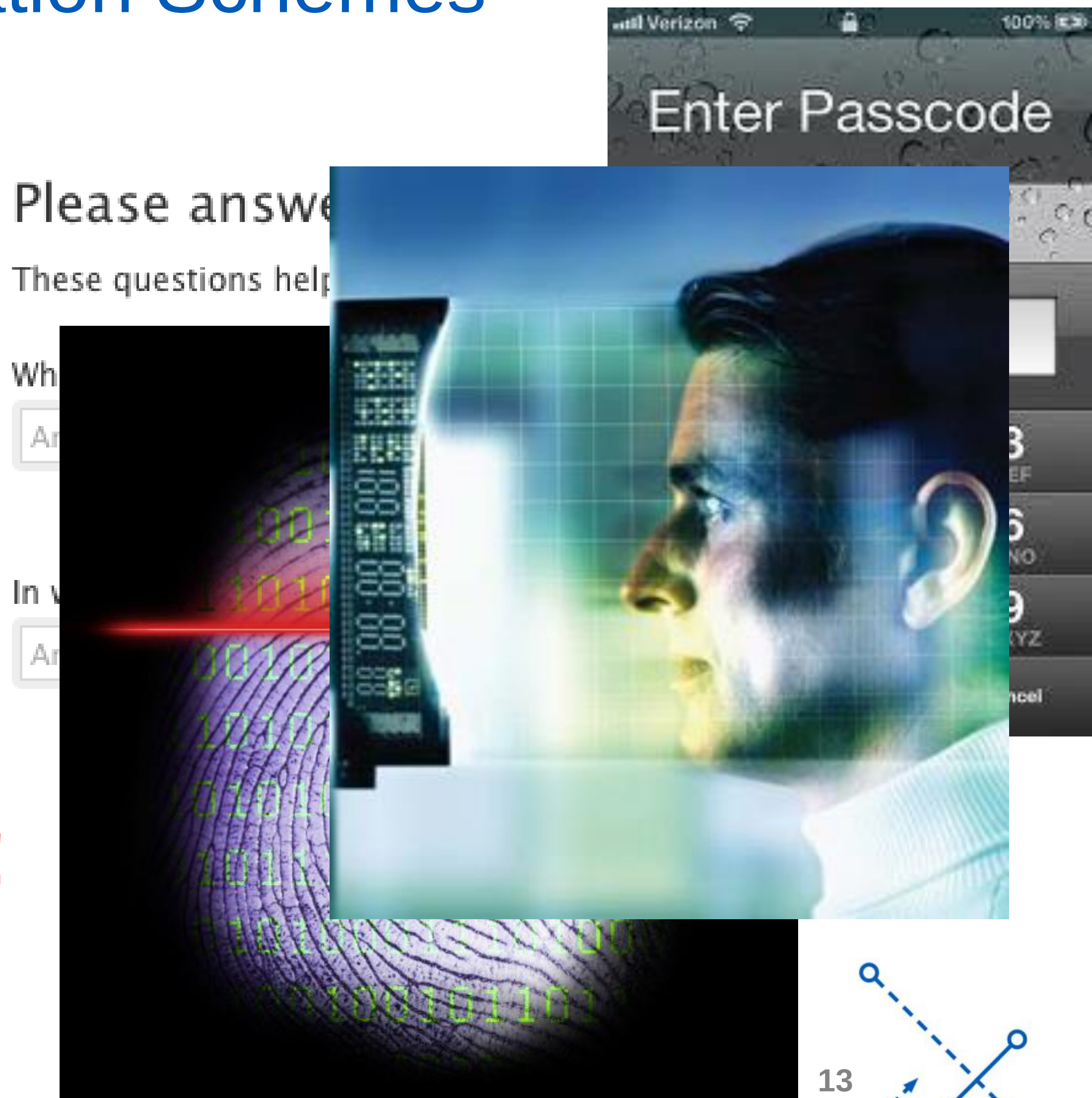
Research directions

- Secondary features
- Deep learning
- Adversarial learning
- Extension to network of smart devices

Current Authentication Schemes

The standard methods

- PIN/Password
- Security Questions
- Fingerprint
- Retina Scanner



They are all obtrusive!

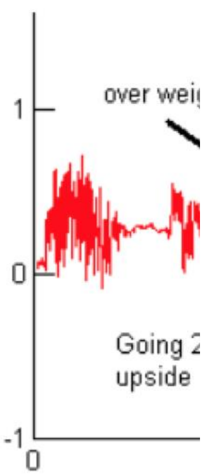
Popular Behavioral Biometrics

Humans recognize people by who they are and how they behave

- Rather than by the secrets that they know

Cues for recognition

- Typing patterns
- Gait
- Word/phrase choices



Displayed image	Worker A	Worker B
	A couple holding hands on a beach during sunset. The sun is creating an orange glow which reflects into the water.	The image shows a sunset. The image shows a beach. The image shows two people holding hands.

A General Approach to Active Authentication

Some call it “Continuous Authentication”, “Implicit Authentication”, “Transparent Authentication”

Users identify themselves at a console and simply start working

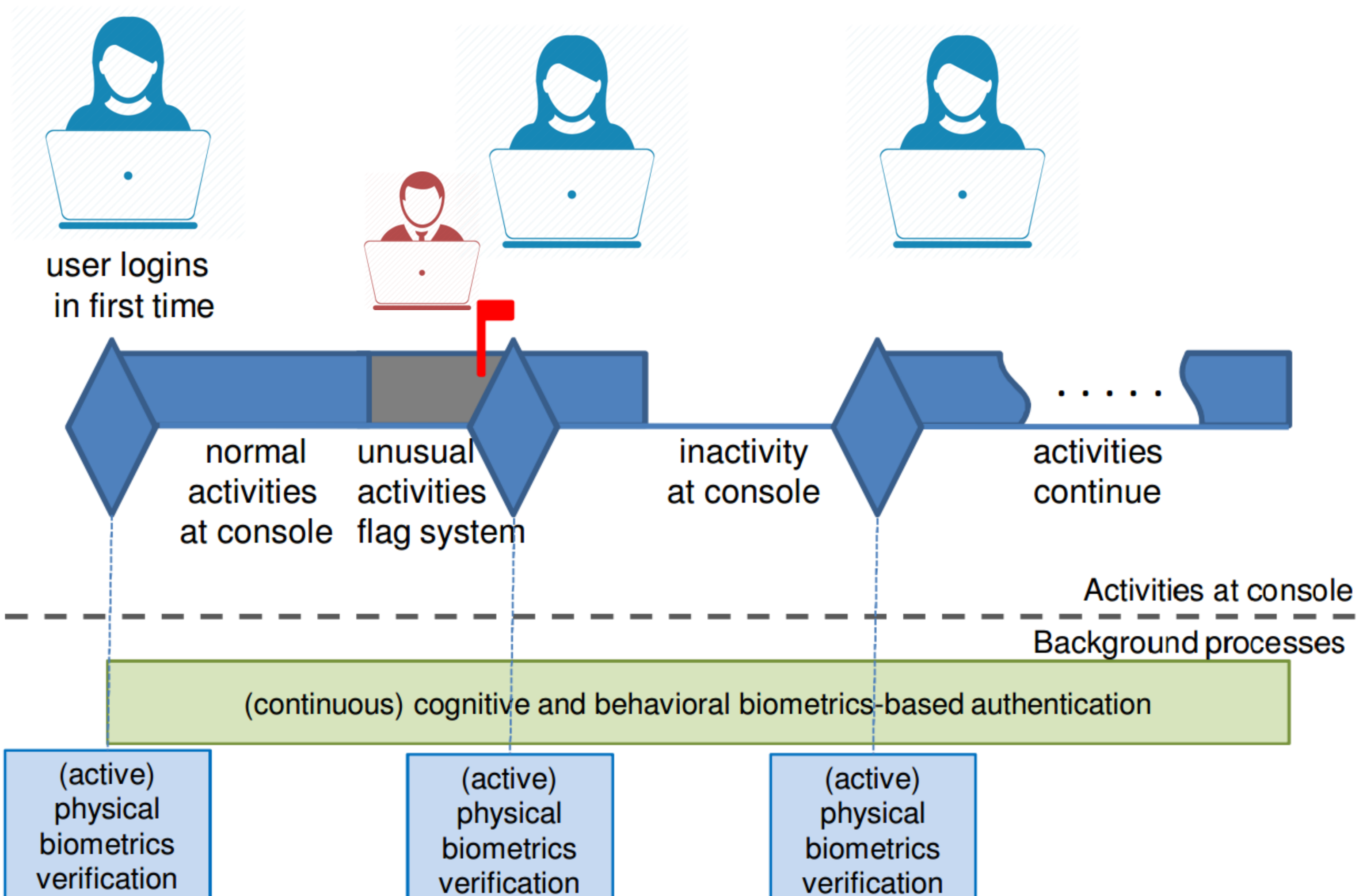
Authentication process occurring in the background

- Invisible and free of interruptions and no loss of performance

Device recognizes the user

- Adapts to changes

A Typical Desktop Scenario



The Big Picture

Transfer learning

Our objective is to design a standalone active authentication mechanism that can adapt to changing environmental conditions by using behavioral biometrics with respect to specific system requirements and certain standards. For instance, the European Stan-

Long-text data

GMM, SVM, Fusion

Keystroke dynamics

Outline of the Talk

Introduction

- General approach to continuous authentication

Keystroke dynamics and mouse movements

- Feature selection
- Methodology - Gaussian model, SVM, transfer learning
- Datasets and anonymization

Results

- GMM, SVM, transfer learning

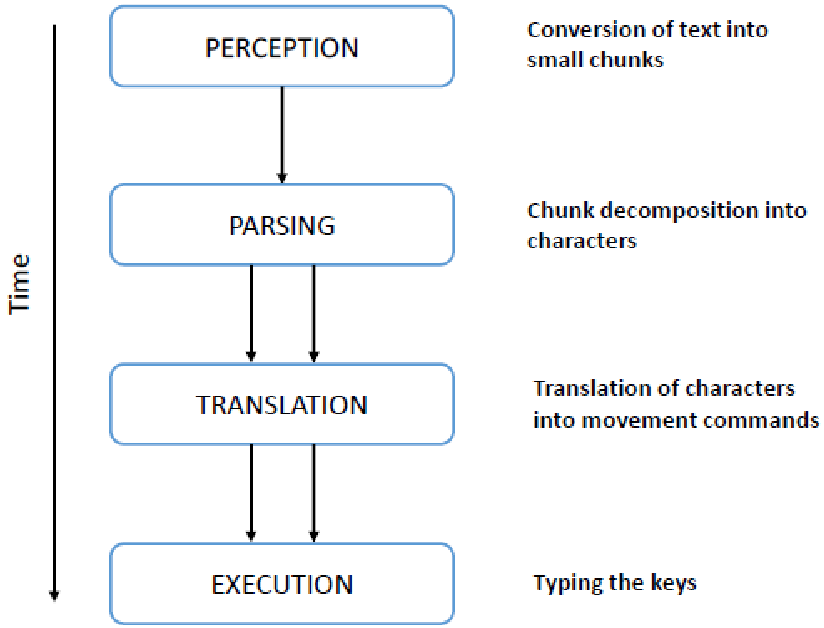
Research directions

- Secondary features
- Deep learning
- Adversarial learning
- Extension to network of smart devices

Why Keystrokes?

Keystroke dynamics in psychology

- Human computer interaction play a key role
- Salthouse [1] proposed a model
  for the steps taking place during typing
- It is a 4-stage process



Keystroke dynamics as a behavioral biometric

- Manner and rhythm of typing – Idiosyncratic
- It can be used as a means for authentication
- Low implementation and deployment cost; non-invasive, transparent
- Many methods and classifiers have been proposed

[1] Timothy A Salthouse. Perceptual, cognitive, and motoric aspects of transcription typing. Psychological bulletin, 99(3):303, 1986.

Rhythm in Keystrokes

John

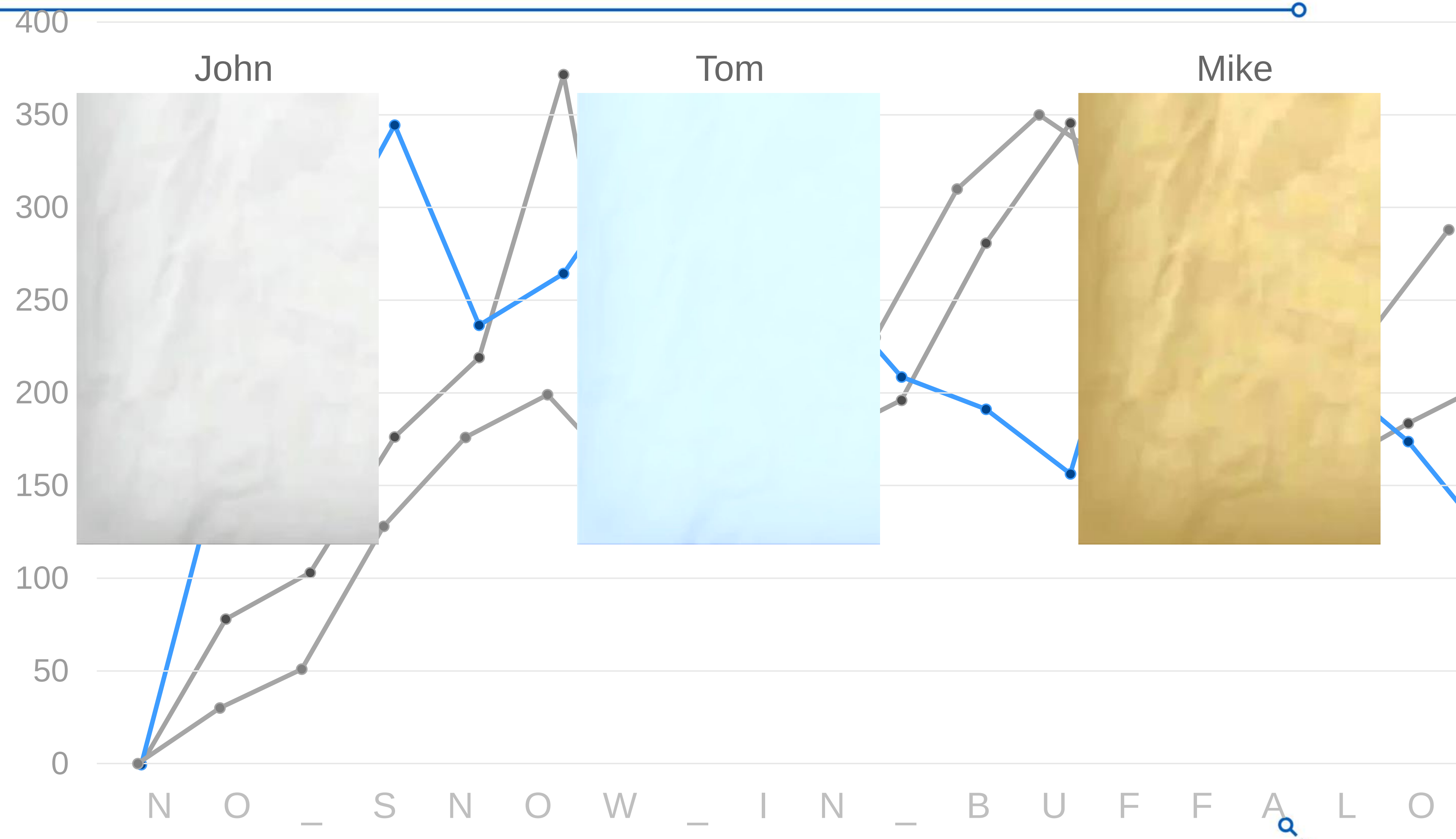
Fast typist

Tom

Medium typist

Mike

Slow typist


400
350
300
250
200
150
100
50
0
John
Tom
Mike
N O _ S N O W _ I N _ B U F F A L O

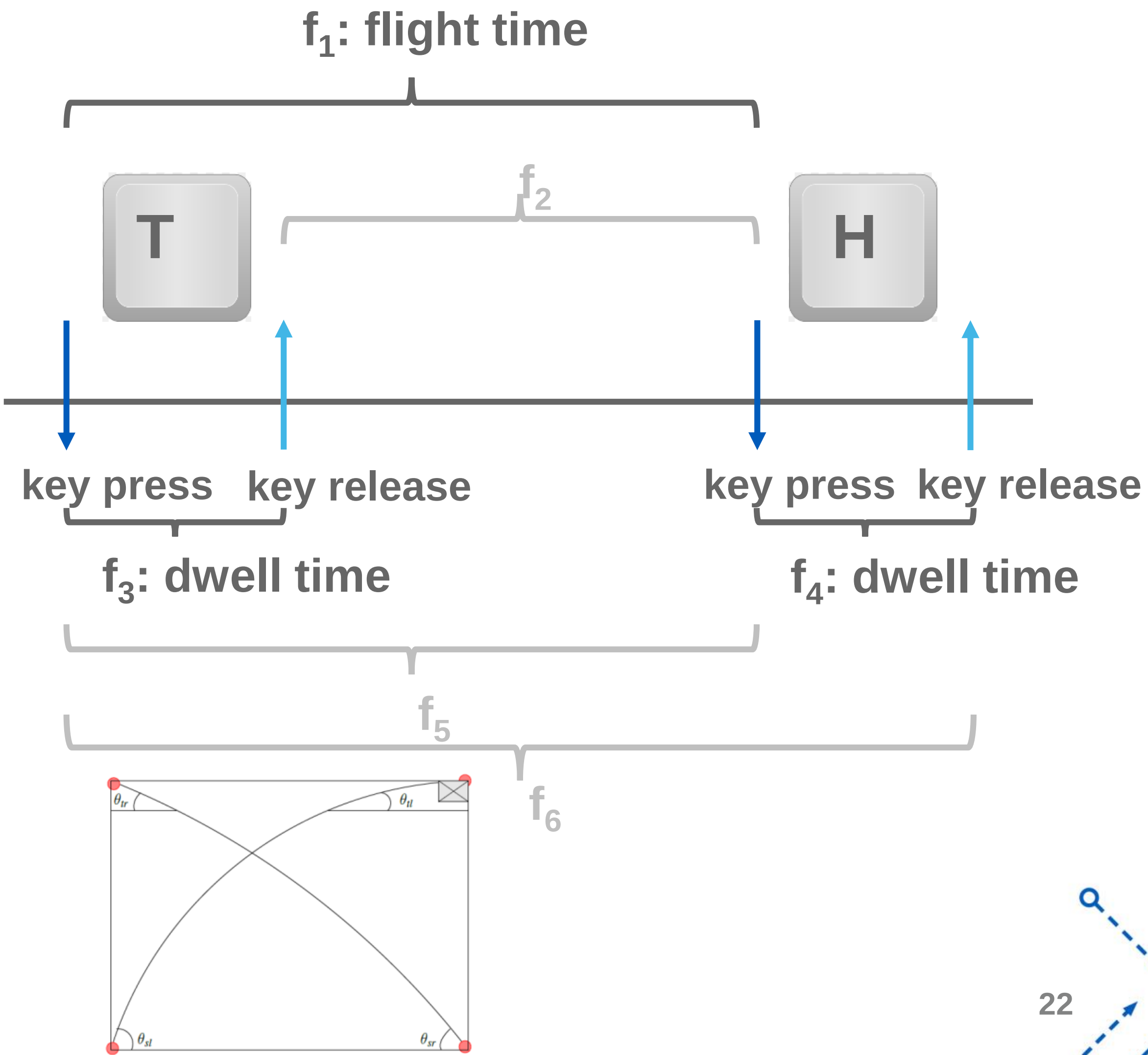
Feature Selection

Keystroke features

- Digraphs

Mouse movements

- Clicks
- Distance
- Speed
- Angle



Methodology

Classification

- Keystroke dynamics recognition is a pattern recognition problem
- Three categories of algorithms [1]
  - Statistical (61%) – probabilistic, cluster analysis
  - Machine learning (37%) – Neural network, decision tree, SVM
  - Others (2%)

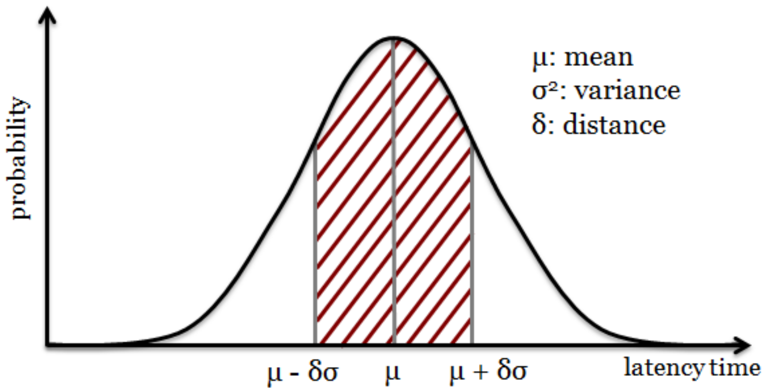
[1] Teh, P. S., Teoh, A. B. J., & Yue, S. (2013). A survey of keystroke dynamics biometrics. *The Scientific World Journal*, *2013*.

(1) Gaussian Model

Classification and authentication

- Every digraph latency exhibits a Gaussian distribution
- 26 X 26 digraphs – Flight time
  - E.g. TH, AB …
- Create profile for each user
- Measure similarity score

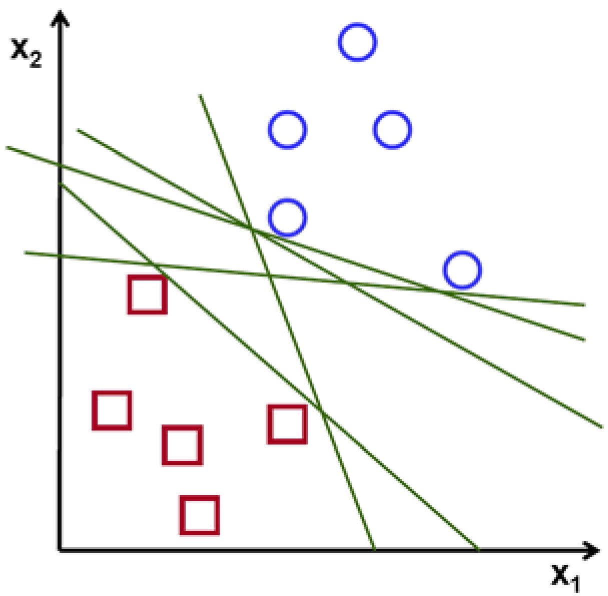
Zone of acceptance

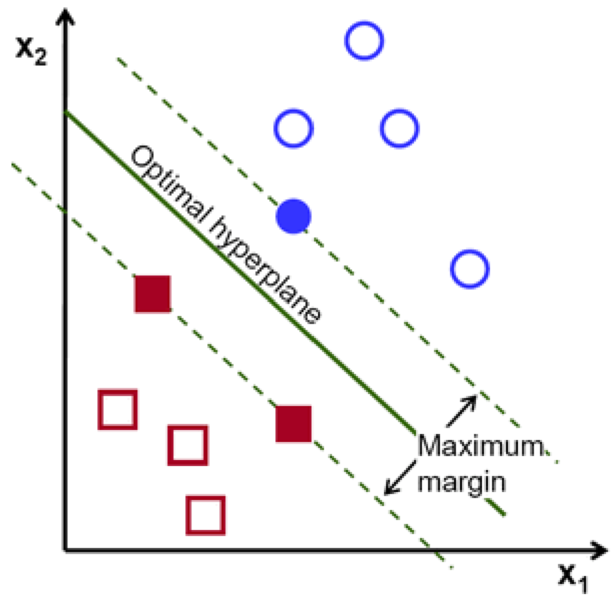


(2) Support Vector Machine (SVM)

A highly utilized classifier [1]

- Generates a region that separates majority of feature data related to a particular class
- By mapping the input vector into a high-dimensional feature space via the kernel function - linear, polynomial, sigmoid, or radial basis function
- Low energy consumption and high performance





[1] R. Caruana and A. Niculescu-Mizil. An empirical comparison of supervised learning algorithms. *Proceedings of the 23th International Conference on Machine Learning*, pages 161–168, 2006

(2) One Class Support Vector Machines (SVM)

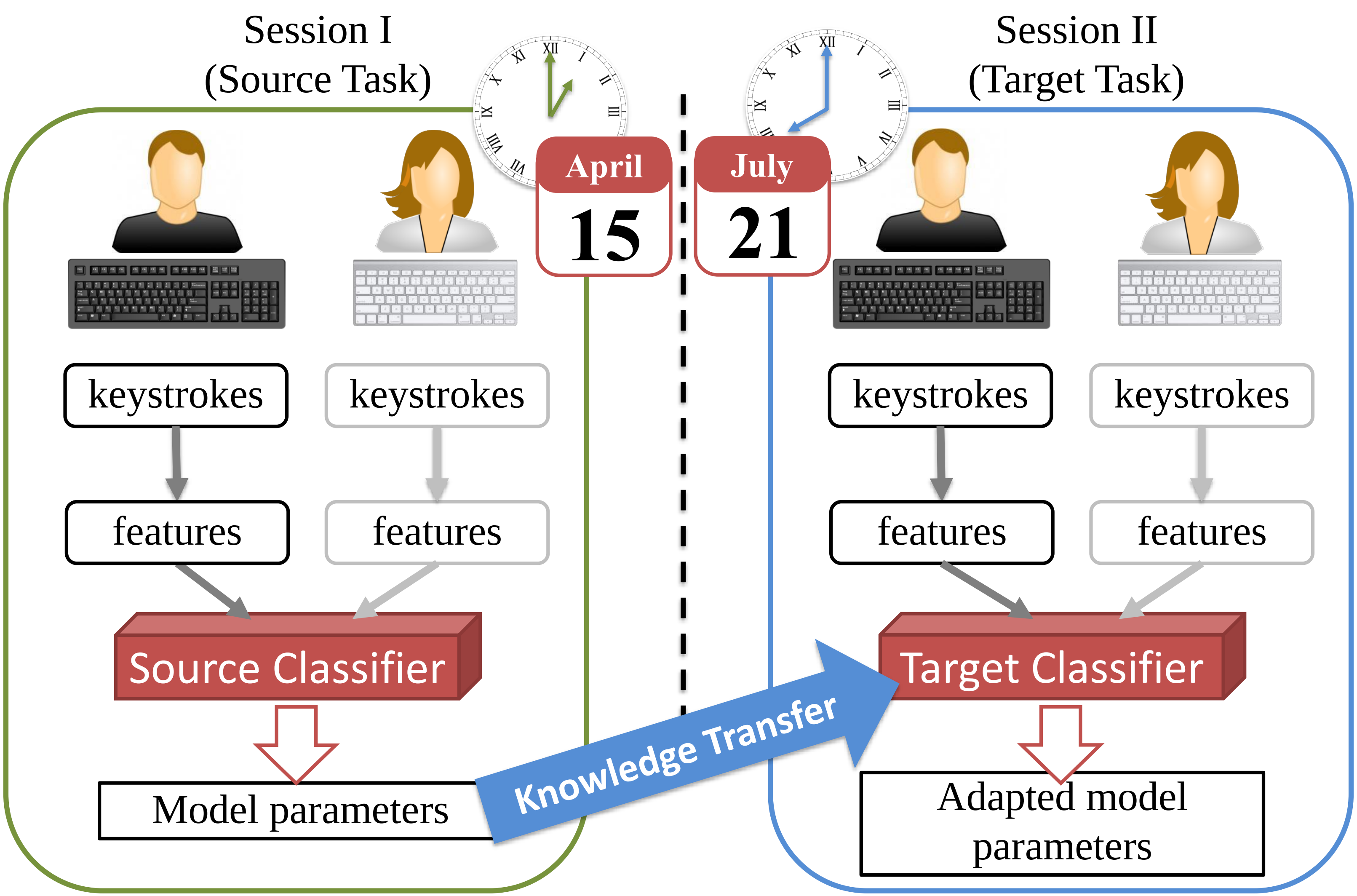
What fits our authentication goal?

- Legitimate data assumed as positive class (+1)
- Anything else as negative class (-1)
- Gaussian Radial Base Kernel Function (RBF)
- Optimal kernel scale

$$K(x, x\prime) = \exp\left(-\frac{||x - x\prime||^2}{2\sigma^2}\right)$$

Where $\sigma \in R$ is a kernel parameter and $||x - x'||$ is the dissimilarity measure

(3) Transfer Learning



Shared Keystroke Dataset

Why?

- Generalization of results
- Benchmarking various algorithms

What is missing?

- Very few high quality datasets in the public domain
- Those available are mostly on short texts
- Some are not accessible

Long text datasets are fundamental for continuous authentication

Related Work

Characteristics of current publicly available datasets – long text

	#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	289	3	50 min	28 days	15 ms	Yes	204 : 85	20-30

"-" symbolizes a feature not present in the original paper

Desirable Characteristics

- Large subject number
- Characterized to reflect
  - Temporal aspects of typing patterns
  - Effect of keyboard layout variability
- Textual data included
- Mouse movements and system events data

Institutional Review Board (IRB) permission

Overview of Experiments

- A large scale data collection campaign
  - 4 months in two campaigns from Sept. to Dec. 2015 and Sept. to Dec. 2016
- 157 + 143 volunteers recruited
- 2 keystroke activities involved
  - Transcribed and free text
- 3 sessions for each participant
- Approximately 1 month between sessions
- 50 minutes for each session
- 4 different types of keyboards utilized

Dataset Design - 1

Keyboard variability

- Baseline subset
- Keyboard variation subset

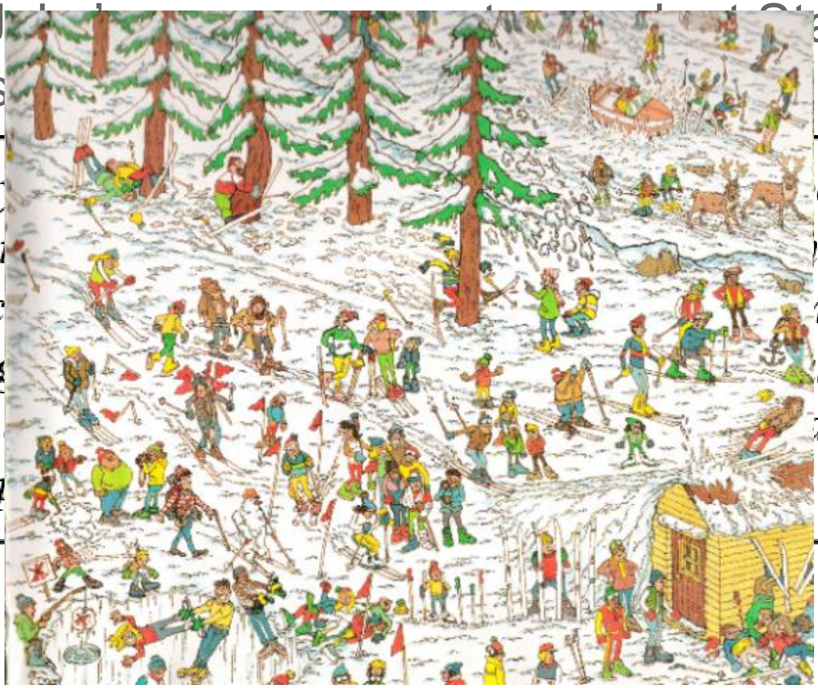
	#subjects	#sessions	#keyboard types
Baseline	157	3	1
Variation	132	3	3

(a) (b) (c) (d)

Dataset Design - 2

Typing activities

- Transcription of fixed text
  - Steve J
    Univers anford
- Describe ol and why you l college or high sc ore interestin t do you think rning the conce eb



Data Acquisition Tool

Active system logger

- Collect system events such as keyboard activity, mouse movement coordinates and mouse clicks
- Desktop based vs. web-based
- Clock resolution: ~15 ms

CPU Ticks	Event	Value
634564465190625000	Mouse Coordinates	464, 348
634564465190625000	Left Click	
634564462834375000	New Process	chrome.exe
634564462895937500	Key Down	G
634564462897187500	Key Up	G

Data Anonymization and Quality Assurance

Privacy protection

- Rule-Based sanitization tool

Secure Transportation

- Data transmission tool

Quality Assurance

- Incomplete data files removed
  - Several subjects removed
  - 300 subjects −> 289 subjects

Evaluation

Statistics

- Parameterize various experiments

Experiments

- Show overall quality

Statistics - 1

Number of raw keystrokes

- 5,800 keystrokes each subject per session
- 17,600 keystrokes each subject
- Minimum 10,000 keystrokes per subject

	# keys Session 1			# keys Session 2			# keys Session 3		
	Task 1	**Task 2**	**Sum**	**Task 1**	**Task 2**	**Sum**	**Task 1**	**Task 2**	**Sum**
Average	3729	2082	5811	3666	2101	5767	3912	2117	6028
Stdev	467	650	891	393	750	913	401	634	768
Min	2334	393	3426	2012	175	3413	1338	169	3560
Max	5332	5235	9506	6611	8751	12414	6027	5116	8425

Statistics - 2

Time intervals between sessions

- 28 days in average

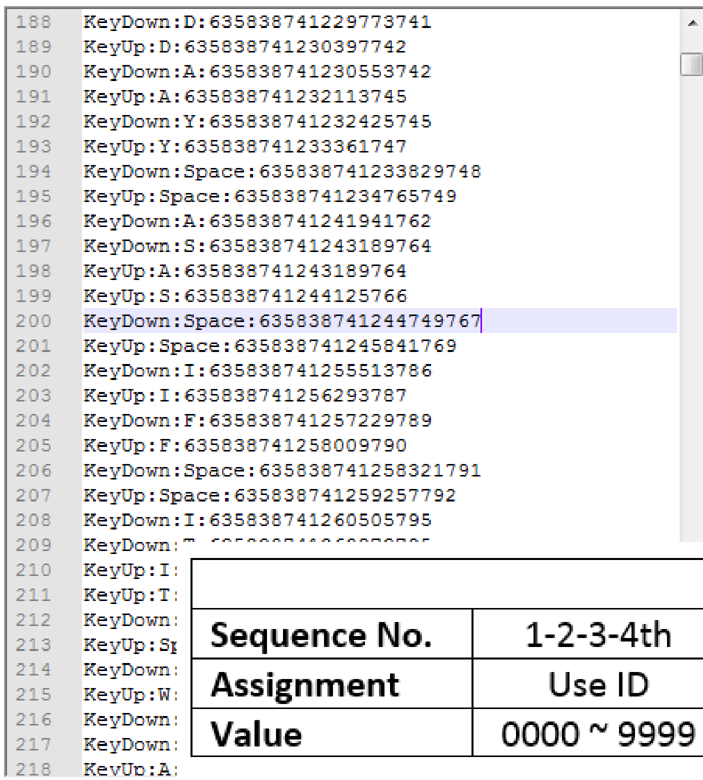
	S1 to S2 (days)	S2 to S3 (days)
Average	28.83	27.35
Stdev	5.99	5.11
Max	47	42
Min	18	14

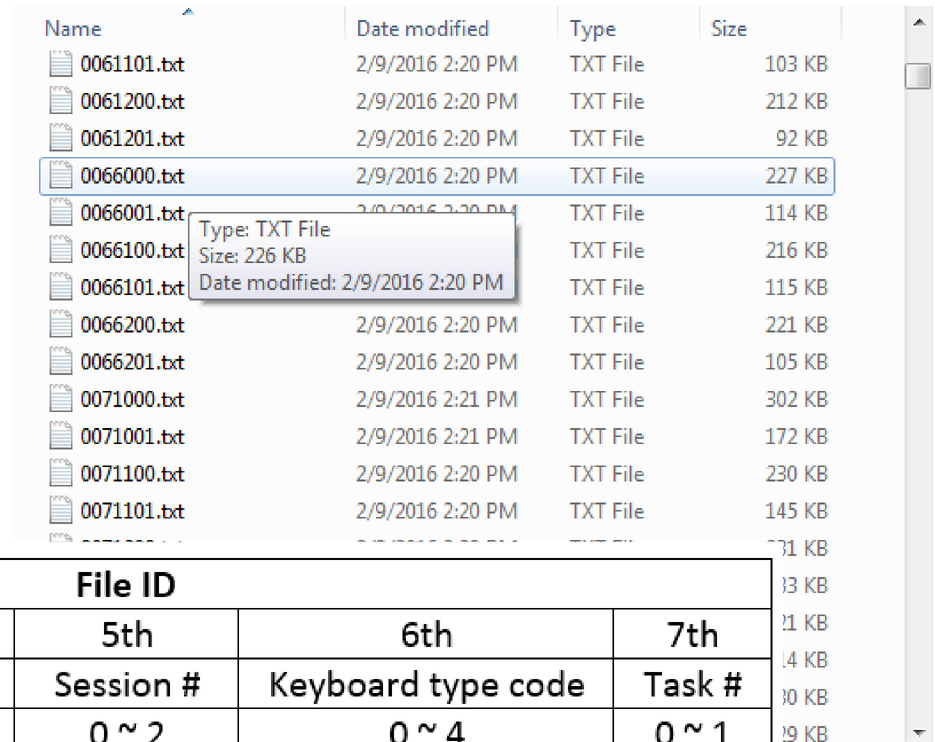
Gender information

- Female 85
- Male 204

	# Male	# Female
Baseline subset	115	43
Keyboard variation subset	89	42
Sum	204	85

Dataset





File ID				
Sequence No.	1-2-3-4th	5th	6th	7th
Assignment	Use ID	Session #	Keyboard type code	Task #
Value	0000 ~ 9999	0 ~ 2	0 ~ 4	0 ~ 1

http://cubs.buffalo.edu/research/datasets

Outline of the Talk

Introduction
- General approach to continuous authentication

Keystroke dynamics and mouse movements
- Feature selection
- Methodology - Gaussian model, SVM, transfer learning
- Datasets and anonymization

Results
- GMM, SVM, transfer learning

Research directions
- Secondary features
- Deep learning
- Adversarial learning
- Extension to network of smart devices

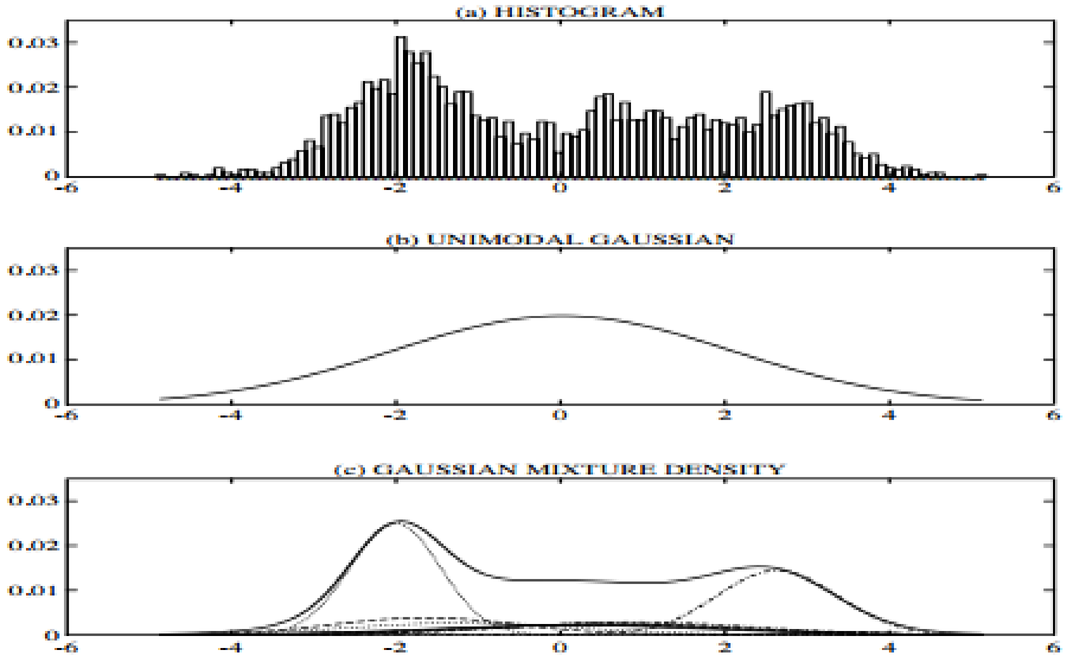
Metrics

Evaluation criteria

- False Reject Rate (FRR) – falsely denied genuine users – (Type 1 error)
- False Accept Rate (FAR) – falsely accepted unauthorized users – (Type 2 error)
- Equal Error Rate (EER) – overall accuracy – (Cross-over error rate)
- Receiver Operating Characteristic (ROC) – true positive rate vs. false positive rate
- Area under the curve (AUC) – scalar representation of ROC

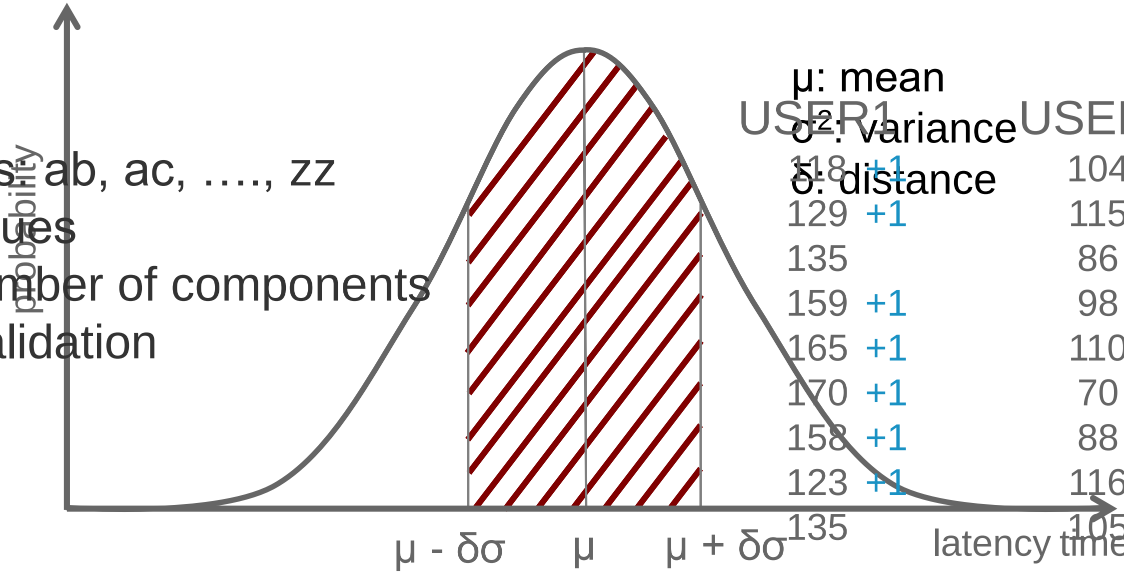
(1) GMM as the Classification Algorithm

- GMM can represent complex and hard-to-map data to an understandable and distinguishable format
- Perturbations can be acquired
- Easy to implement



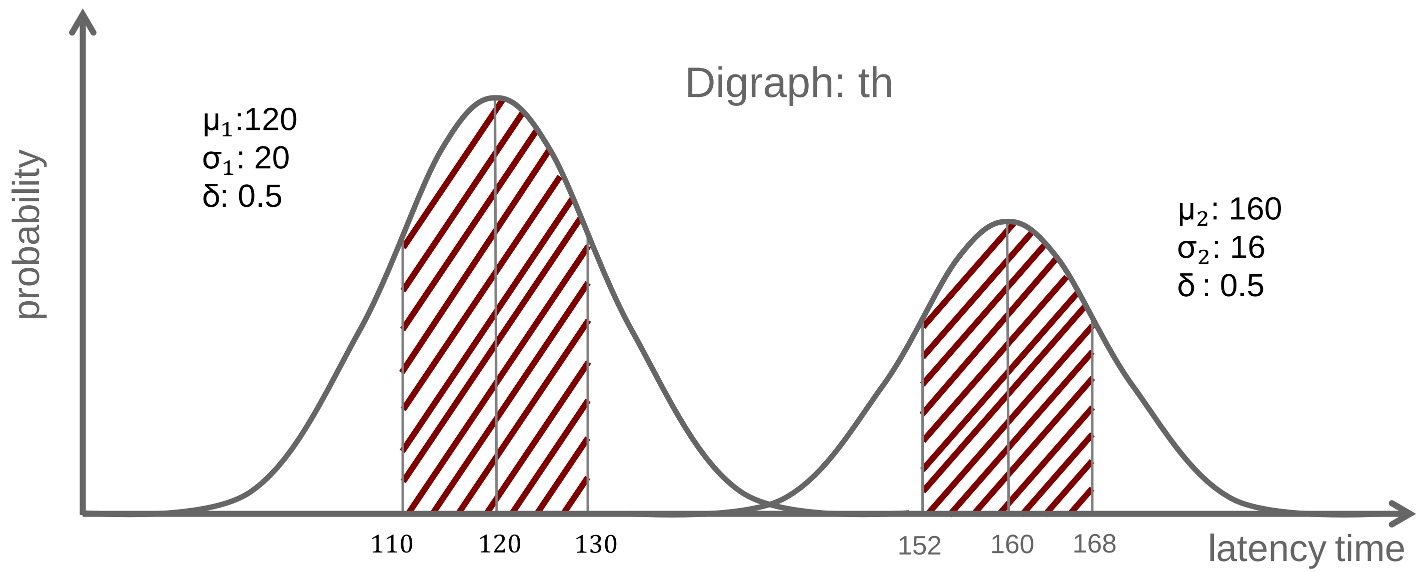
Traditional Approaches

- Considering all digraphs: ab, ac, …., zz
- Various distance (δ) values
- GMMs with different number of components
- Leave-one-out cross-validation

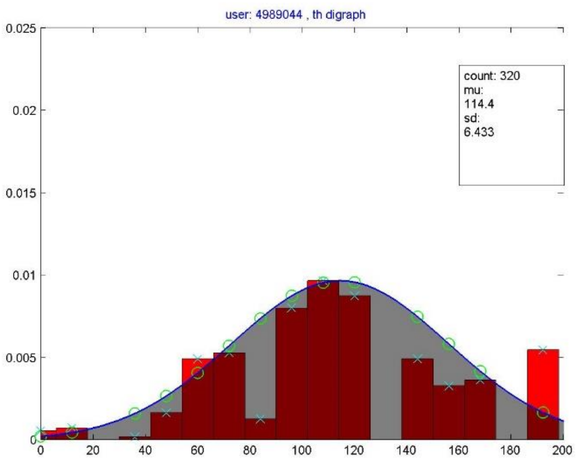


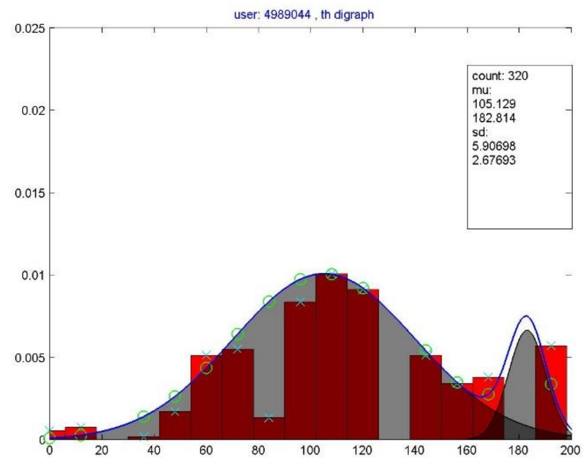
USER1		USER2	
118	+1	104	
129	+1	115	+1
135		86	
159	+1	98	
165	+1	110	+1
170	+1	70	
158	+1	88	
123	+1	116	+1
135		105	

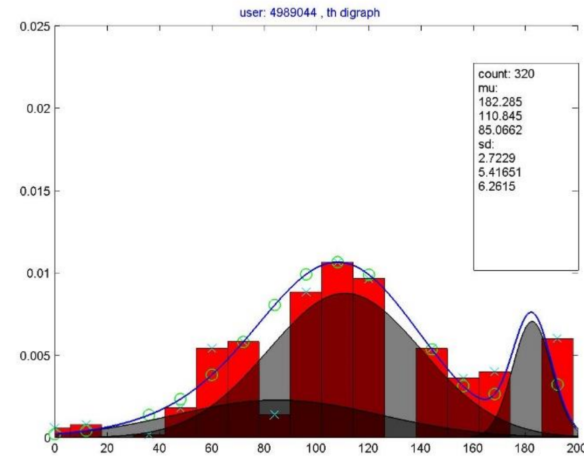
Gaussian Mixture Model

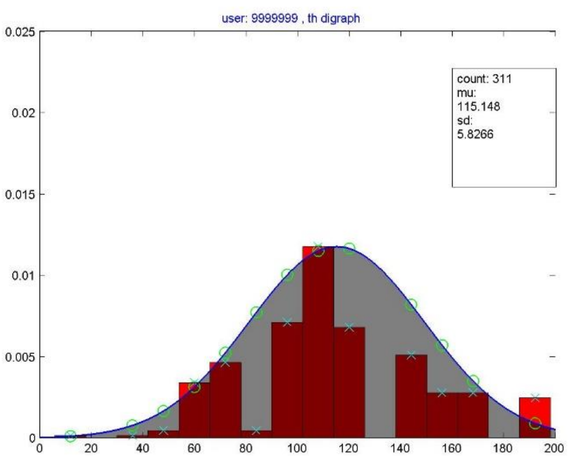


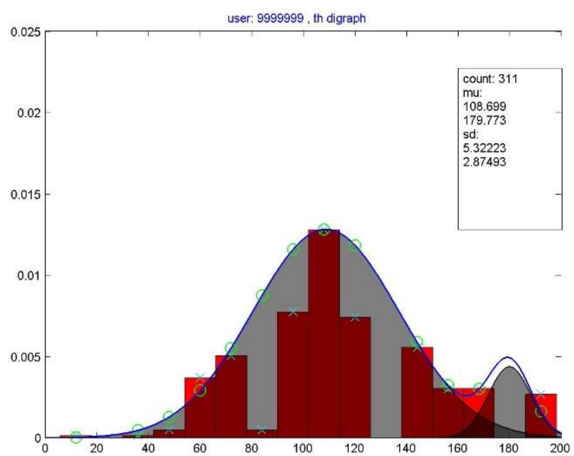
Keystroke Dynamics with GMM

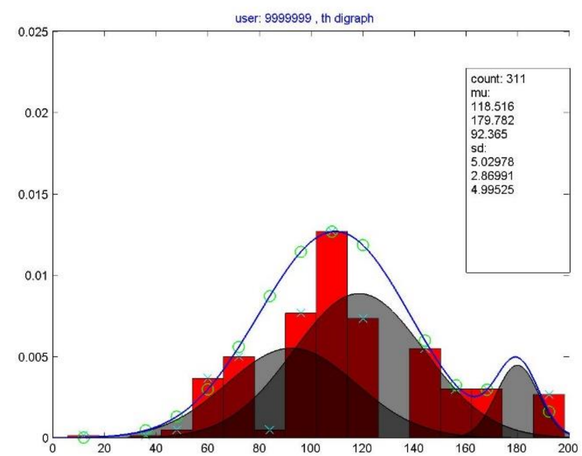












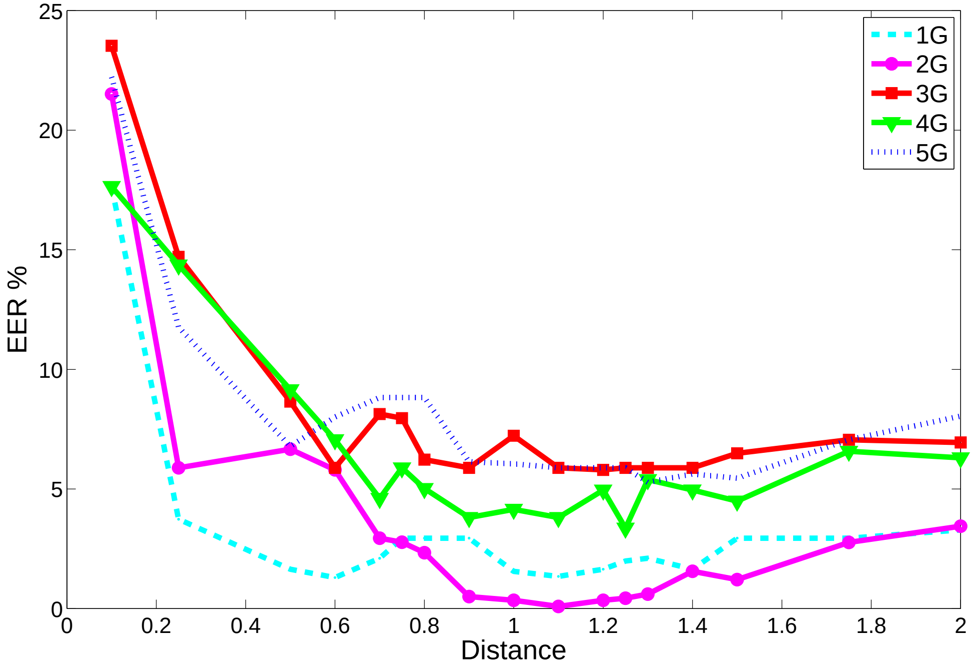
Hard to separate with 1G

Somewhat better

More separable

Results

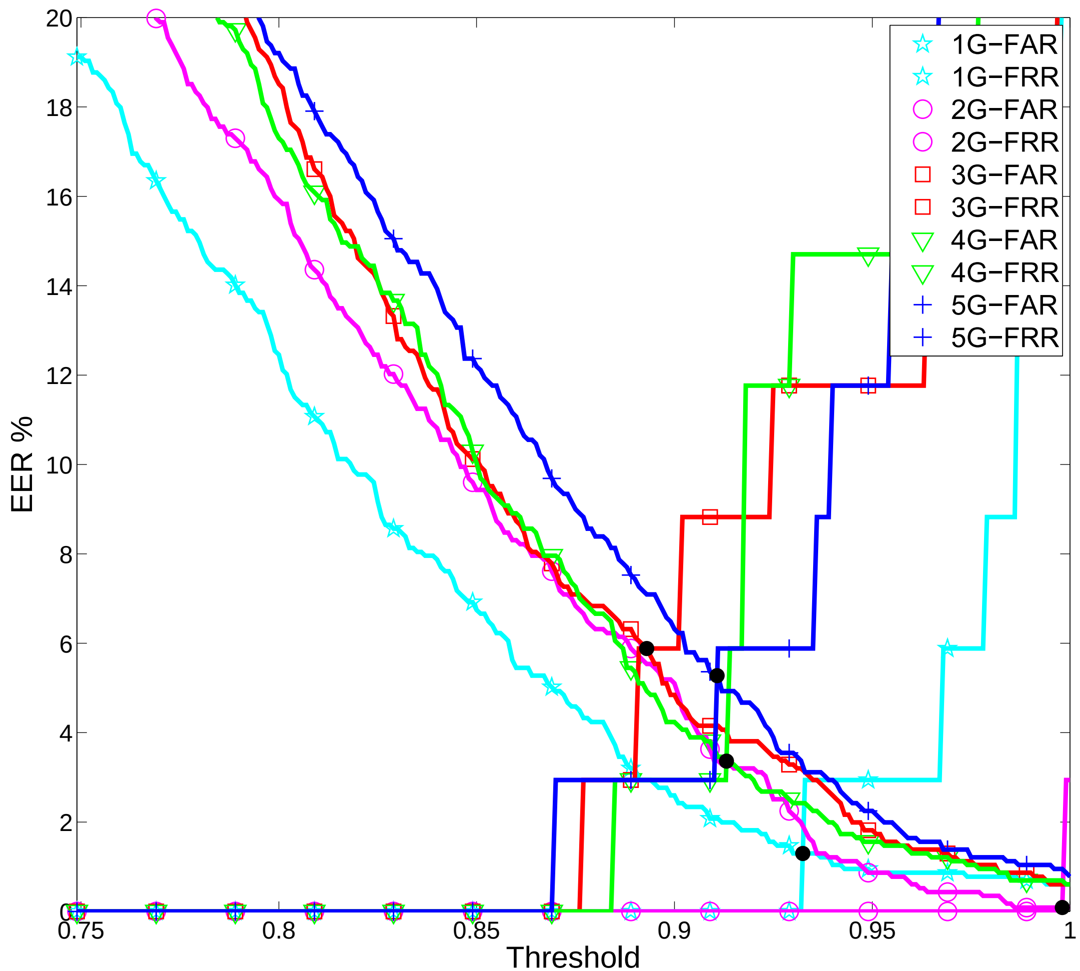
- Clarkson dataset (39 users) is used



- Word-initiation effect
- Curse of dimensionality
- Presence of singularities

False Accept/Reject Rate

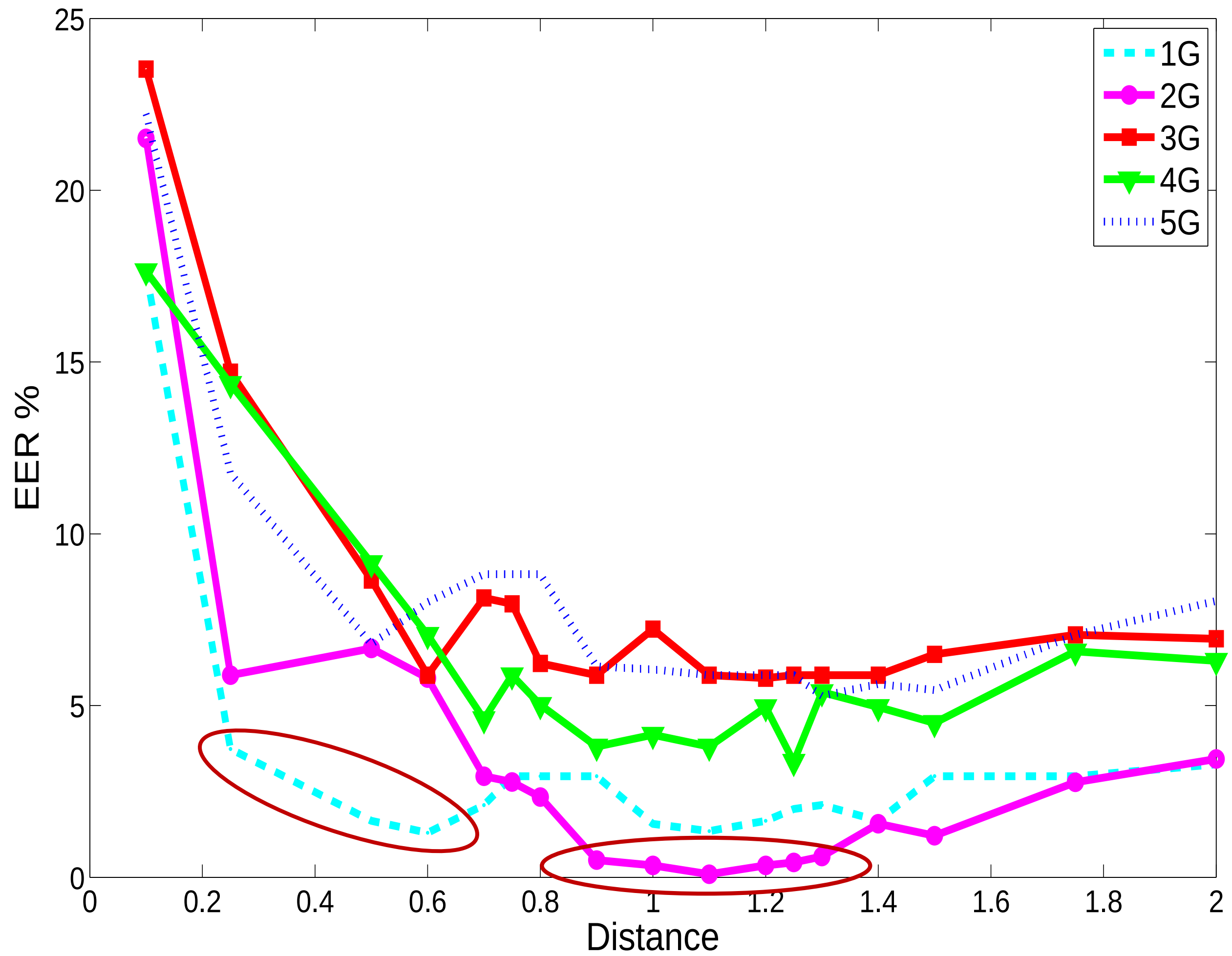
1G	1.3%
2G	0.08%
3G	5.88%
4G	3.36%
5G	5.28%



University at Buffalo The State University of New York

Is GMM Enough?

- No winner model
- Consolidating the strengths
- Anomalous characteristics are suppressed

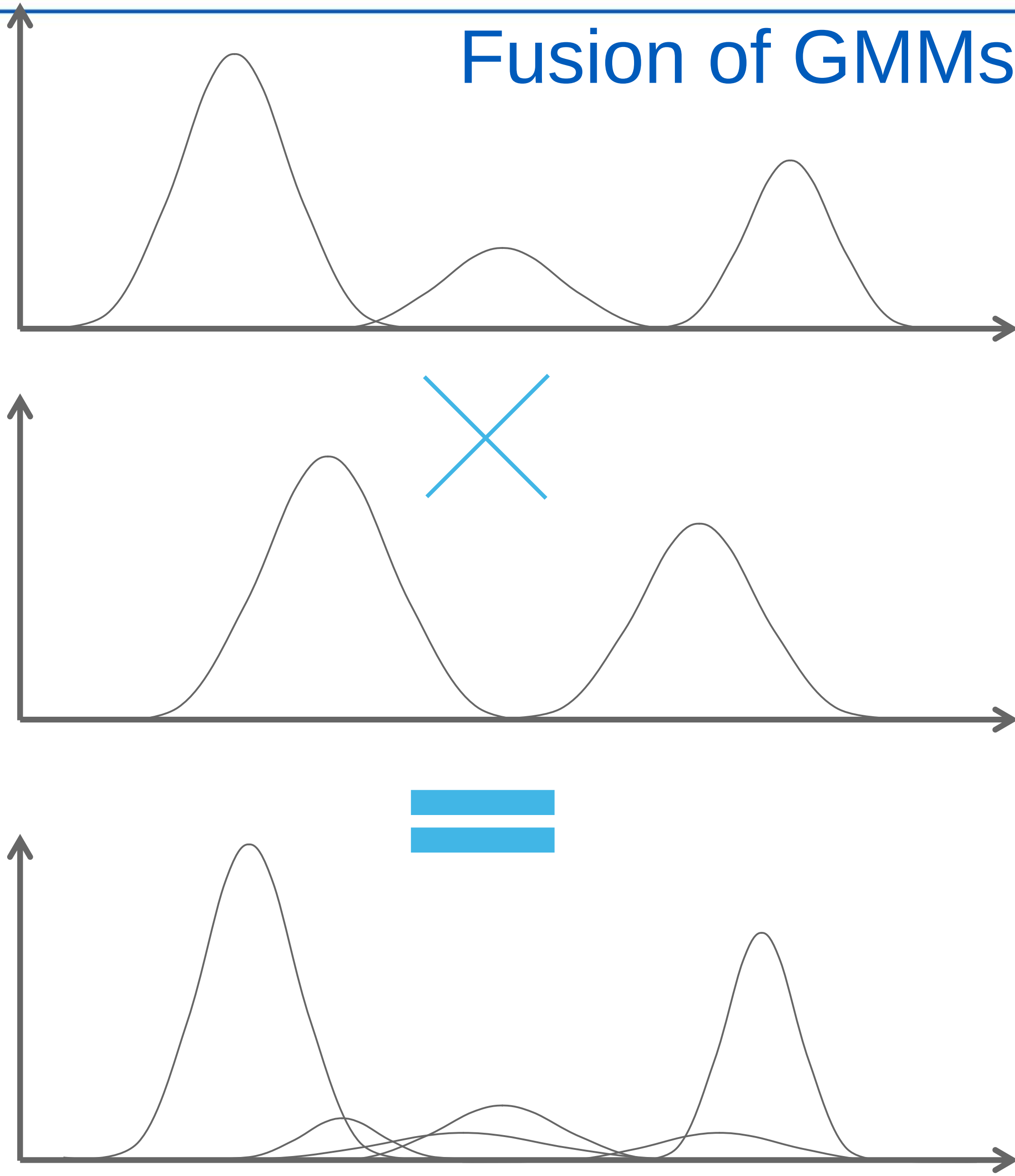


We take one step further!

Information Fusion

- Multiple sources, modalities or decisions
- Parameters from different classifiers are consolidated
- More refined set of criteria

Fusion of GMMs



Naïve Bayes

$$A_{ij}^{-1} = B_i^{-1} + C_j^{-1}$$

$$a_{ij} = A_{ij}\left(B_i^{-1} b_i + C_j^{-1} c_j\right)$$

$$r_{ij} = \frac{p_i q_j}{\sum_{k=1}^{M_b} \sum_{l=1}^{M_c} p_k q_l}$$

Covariance Intersection

$$A_{ij}^{-1} = \omega_{ij} B_i^{-1} + (1 - \omega_{ij}) C_j^{-1}$$

$$a_{ij} = A_{ij}\left(\omega_{ij} B_i^{-1} b_i + (1 - \omega_{ij}) C_j^{-1} c_j\right)$$

$$r_{ij} = \frac{\omega_{ij} p_i + (1 - \omega_{ij}) q_j}{\sum_{k=1}^{M_b} \sum_{l=1}^{M_c} \omega_{kl} p_k + (1 - \omega_{kl}) q_l}$$

Chernoff Information

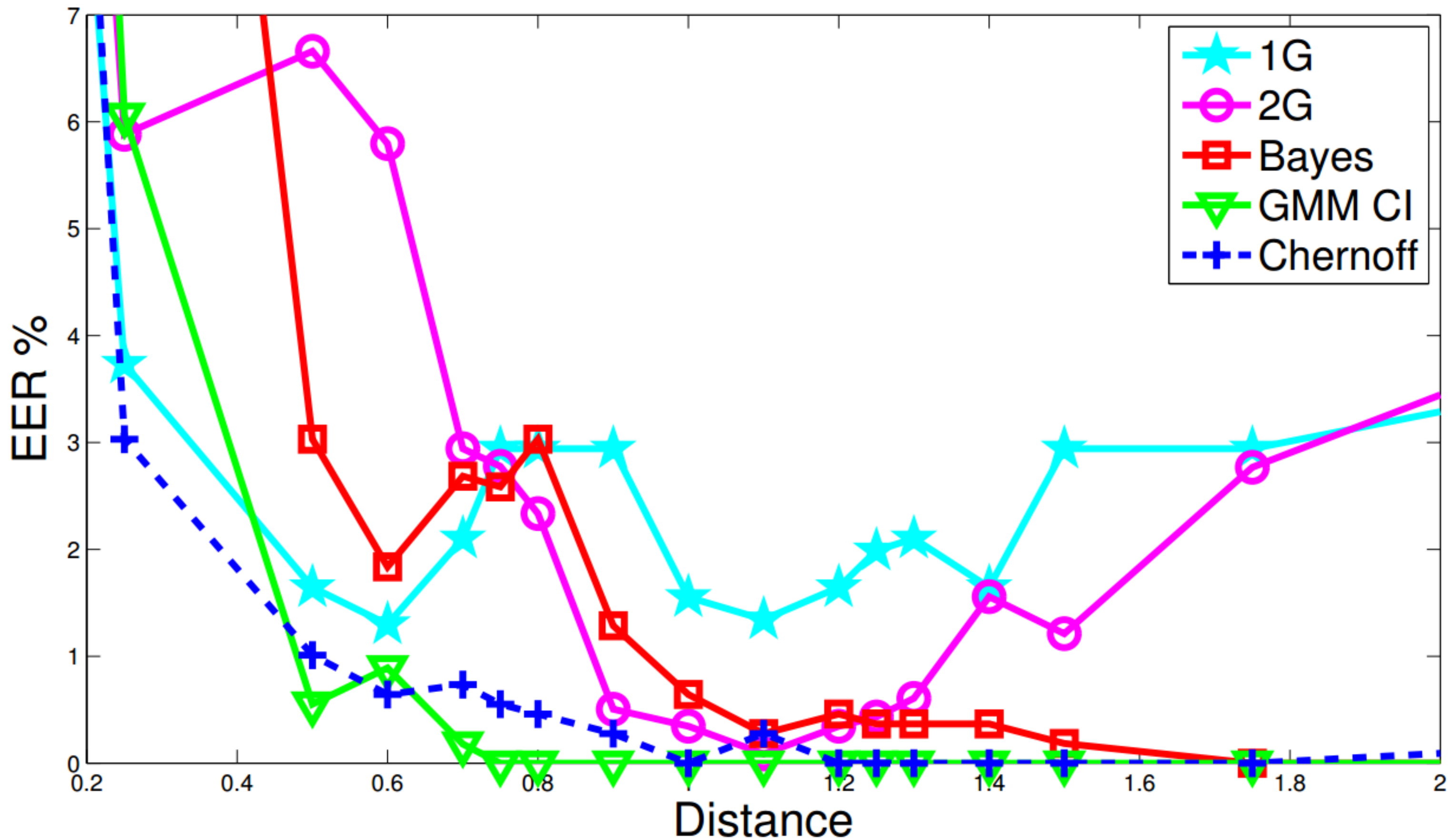
$$A_{ij}^{-1} = \omega B_i^{-1} + (1 - \omega) C_j^{-1}$$

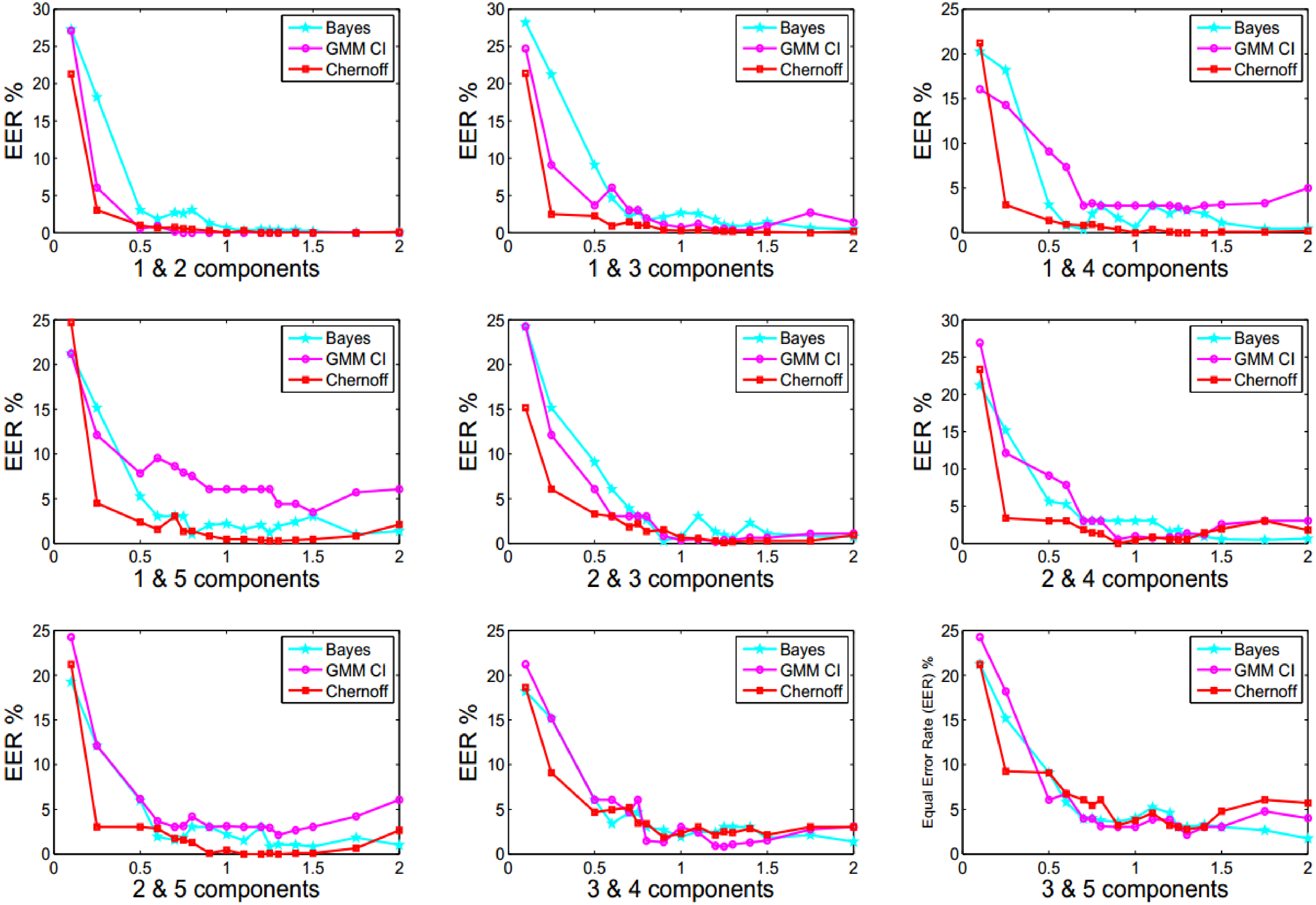
$$a_{ij} = A_{ij}\left(\omega B_i^{-1} b_i + (1 - \omega) C_j^{-1} c_j\right)$$

$$r_{ij} = \frac{p_i^{\omega} q_j^{(1-\omega)}}{\sum_{k=1}^{M_b} \sum_{l=1}^{M_c} p_k^{\omega} q_l^{(1-\omega)}}$$

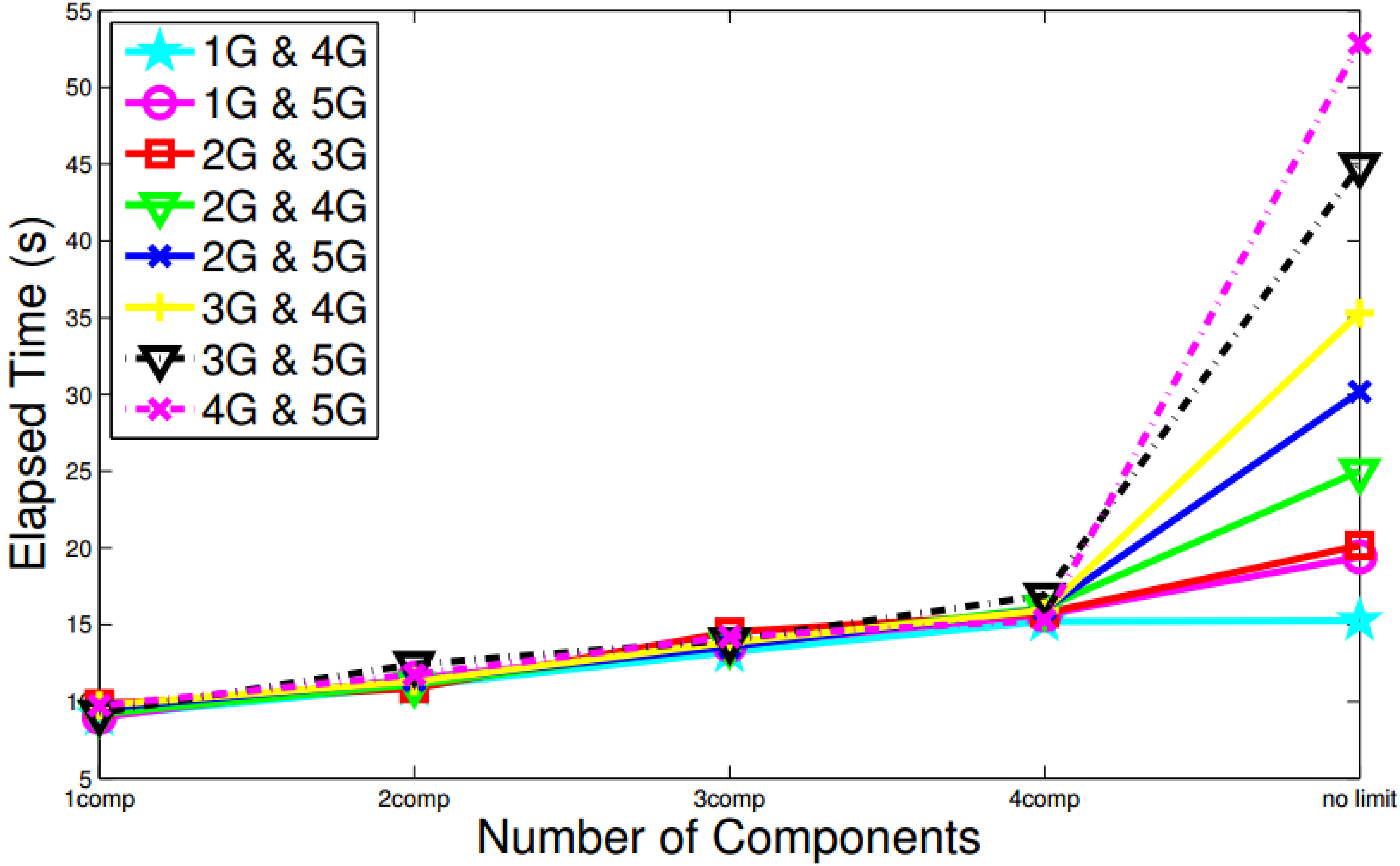
Fusion Results

- Lower error rates
- Regular trend-lines
- Robust classifier




Bayes
GMM CI
Chernoff
EER %
Equal Error Rate (EER) %
1 & 2 components
1 & 3 components
1 & 4 components
1 & 5 components
2 & 3 components
2 & 4 components
2 & 5 components
3 & 4 components
3 & 5 components
0
0.5
1
1.5
2
5
10
15
20
25
30

Time Performance



(2) SVM as the Classification Algorithm

Feature alignment method

- Each observation holds a single row
- Observations from different features in different columns
- Rest cells filled with 0

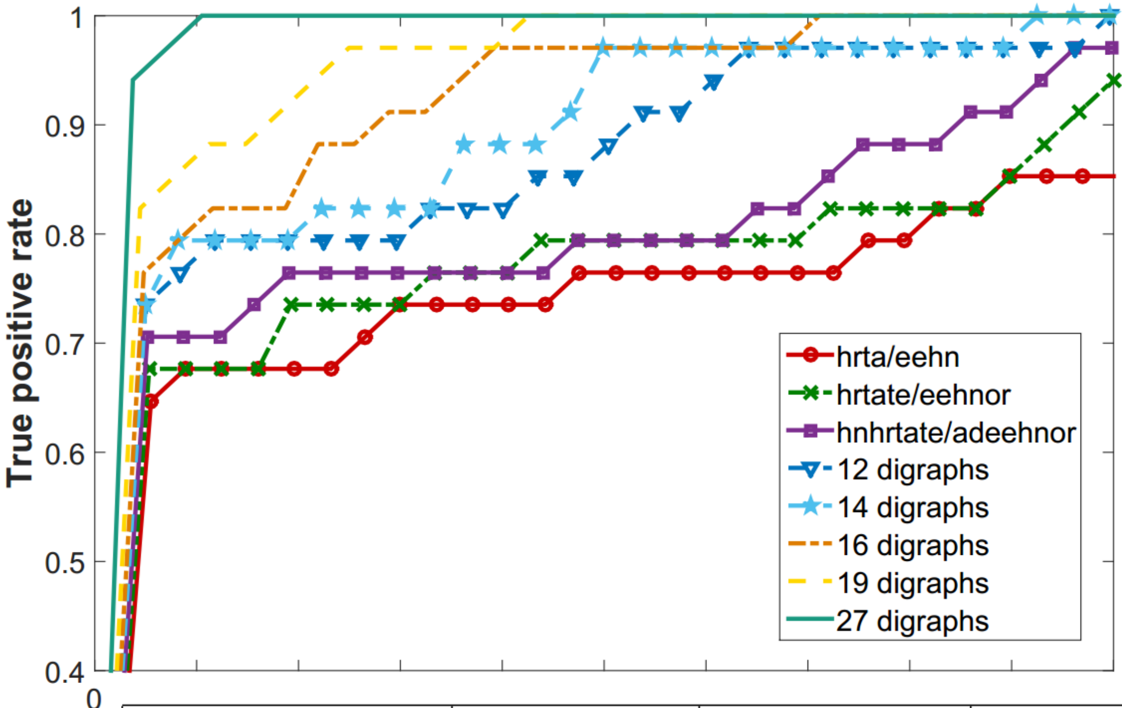
Feature matrix

Feature 1	Feature 2	$\cdots$	Feature N
f_1^1	0	$\cdots$	0
$\vdots$	$\vdots$	$\cdots$	$\vdots$
f_1^k	0	$\cdots$	0
0	f_2^1	$\cdots$	0
0	$\vdots$	$\cdots$	$\vdots$
0	f_2^k	$\cdots$	0
		$\cdots$	
0	0	$\cdots$	f_n^1
$\vdots$	$\vdots$	$\cdots$	$\vdots$
0	0	$\cdots$	f_n^k

Experiment Setting

- Data partition
  - Training sets (80%) and Testing set (20%)
- SVM packages on MATLAB
- Model trained with genuine data (positive class)
- One vs. All testing strategy
- Optimal kernel scale

SVM Results



Fi

Digraph set	Kernel Scale	Training Time	Testing Time	AUC	EER %
hrta/eehn	0.31	0.12	0.0077	0.9947	2.94
hrtate/eehnor	0.36	0.20	0.0128	0.9973	2.58
hnhrtate/adeehnor	0.46	0.28	0.0169	0.9979	2.94
12 digraphs	0.54	0.55	0.0275	1	0
14 digraphs	0.56	0.67	0.0379	1	0
16 digraphs	0.65	0.84	0.0448	1	0

(3) Transfer Learning as Classification Algorithm

Authenticate an enrolled user in a different environment with least amount of re-training

- Knowledge acquired in previous session is transferred via parameters that contain classifier info
- There is a source system and a target system
- Use two different adaptive SVMs with linear and Gaussian kernels
- Source profile works as a regularizer of target profile in the SVM cost function
- Uses a small no. of samples from the target system

Transfer learning in other domains

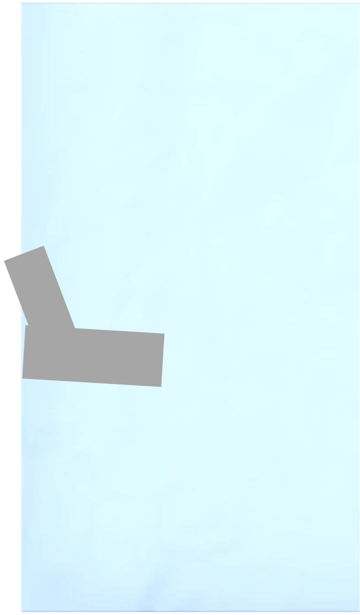
- Concept drift in data mining
- Incremental learning
- Cross-domain learning

Intra-User Variability

John

Normal conditions

John



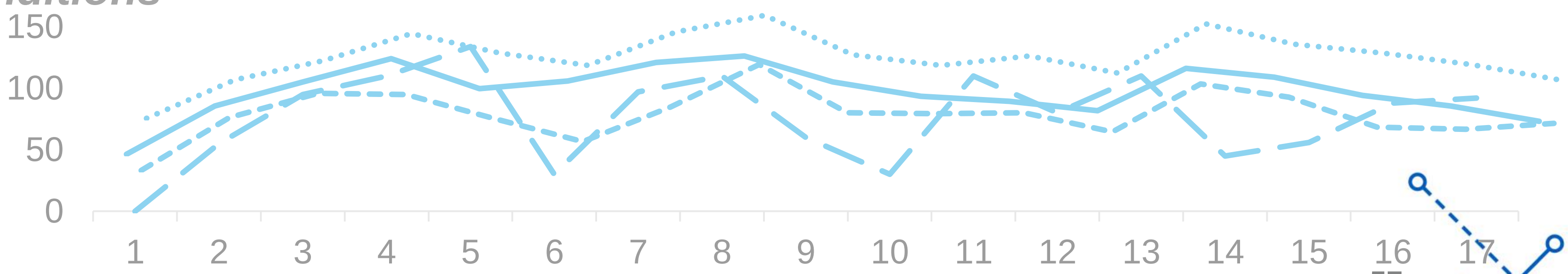
Keyboard

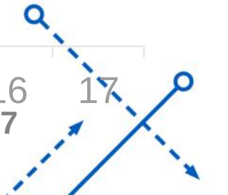
John

Time

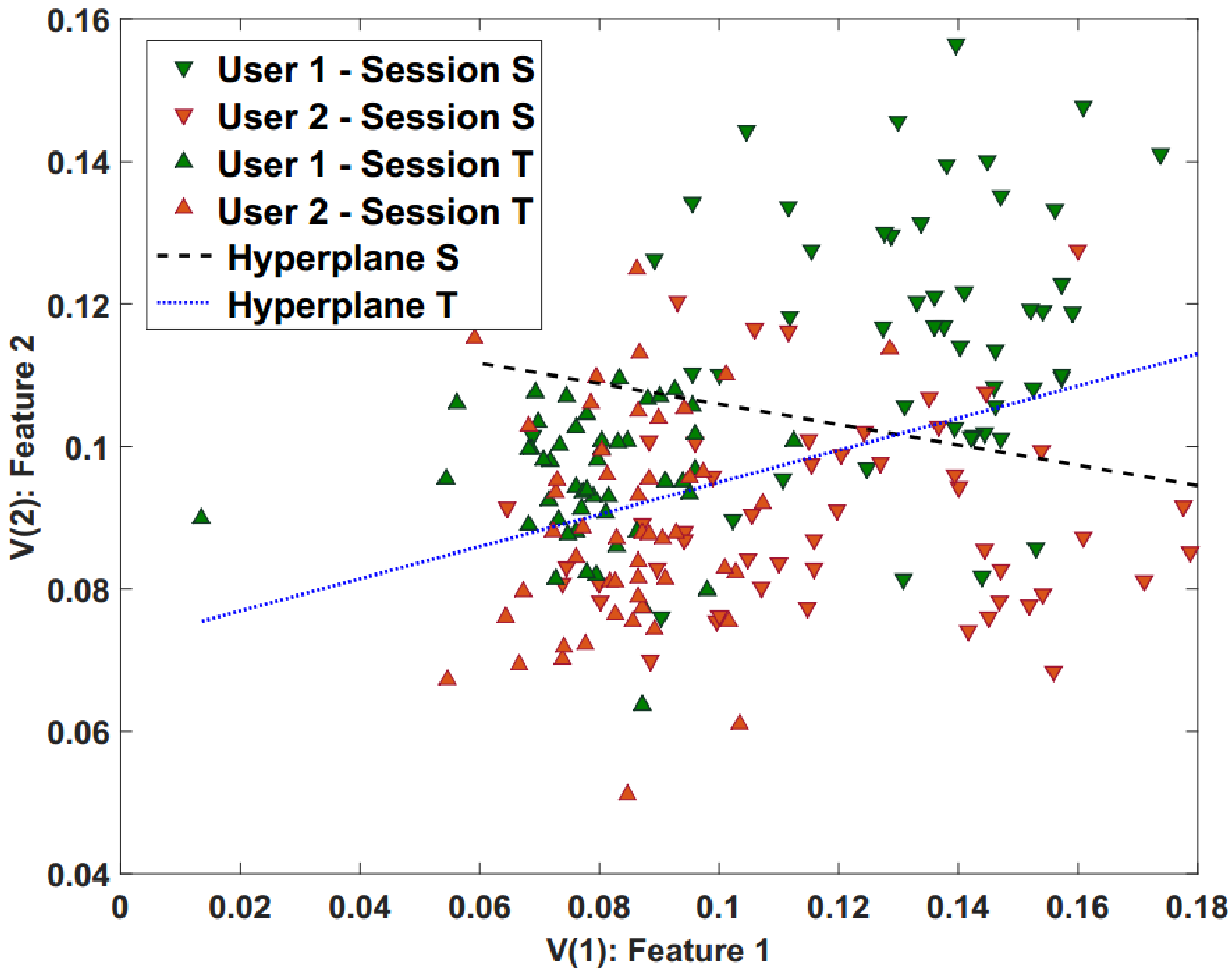
John

Angry

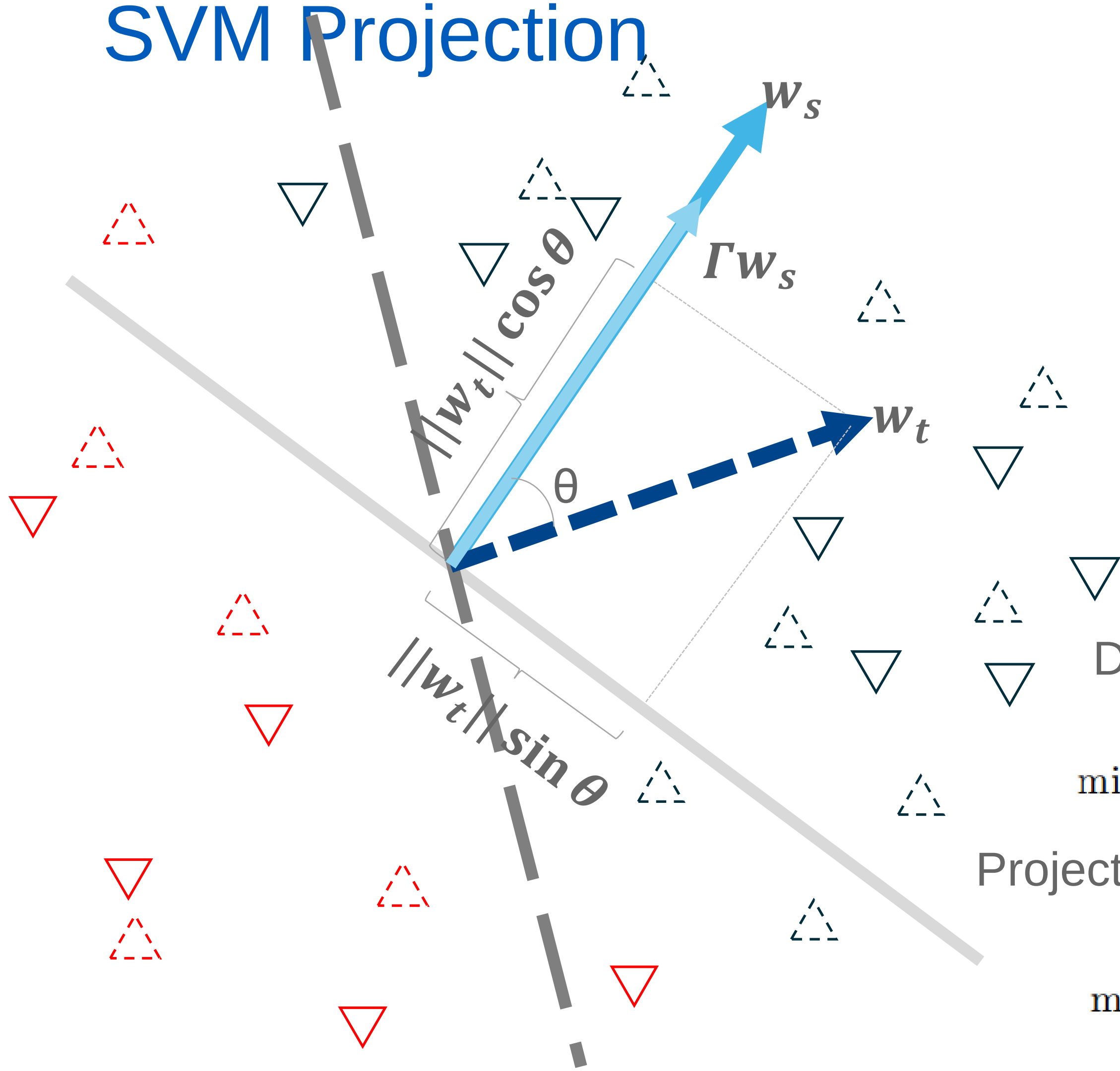




Separating Hyperplanes



SVM Projection



Classic SVM

$$\min \quad J(w,b,\xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i$$
$$\text{s.t.} \quad y_i\left(w^T x_i + b\right) \geq 1 - \xi_i$$
$$\xi_i \geq 0, \quad i = 1, \cdots, l$$

A-SVM [1]

$$\min \left(\frac{1}{2}\|w_t - \Gamma w_s\|^2 + C\sum_{i=1}^{l}\xi_i\right)$$

Deformable Adaptive SVM [2]

$$\min \left(\frac{1}{2}\|w_t - \Gamma f(w_s)\|^2 + \lambda * \Delta + C\sum_{i=1}^{l}\xi_i\right)$$

Projective Model Transfer SVM [2]

$$\min \left(\frac{1}{2}\|w_t\|^2 + \Gamma\|Pw_t\|^2 + C\sum_{i=1}^{l}\xi_i\right)$$

[1] J. Yang, R. Yan, and A. G. Hauptmann. Adapting SVM Classifiers to Data with Shifted Distributions. *Seventh IEEE International Conference on Data Mining Workshops (ICDMW 2007)*, pages 69–76, 2007.
[2] Y. Aytar and A. Zisserman. Tabula rasa: Model transfer for object category detection. Proceedings of the IEEE International Conference on Computer Vision, pages 2252–2259, 2011.

Results*

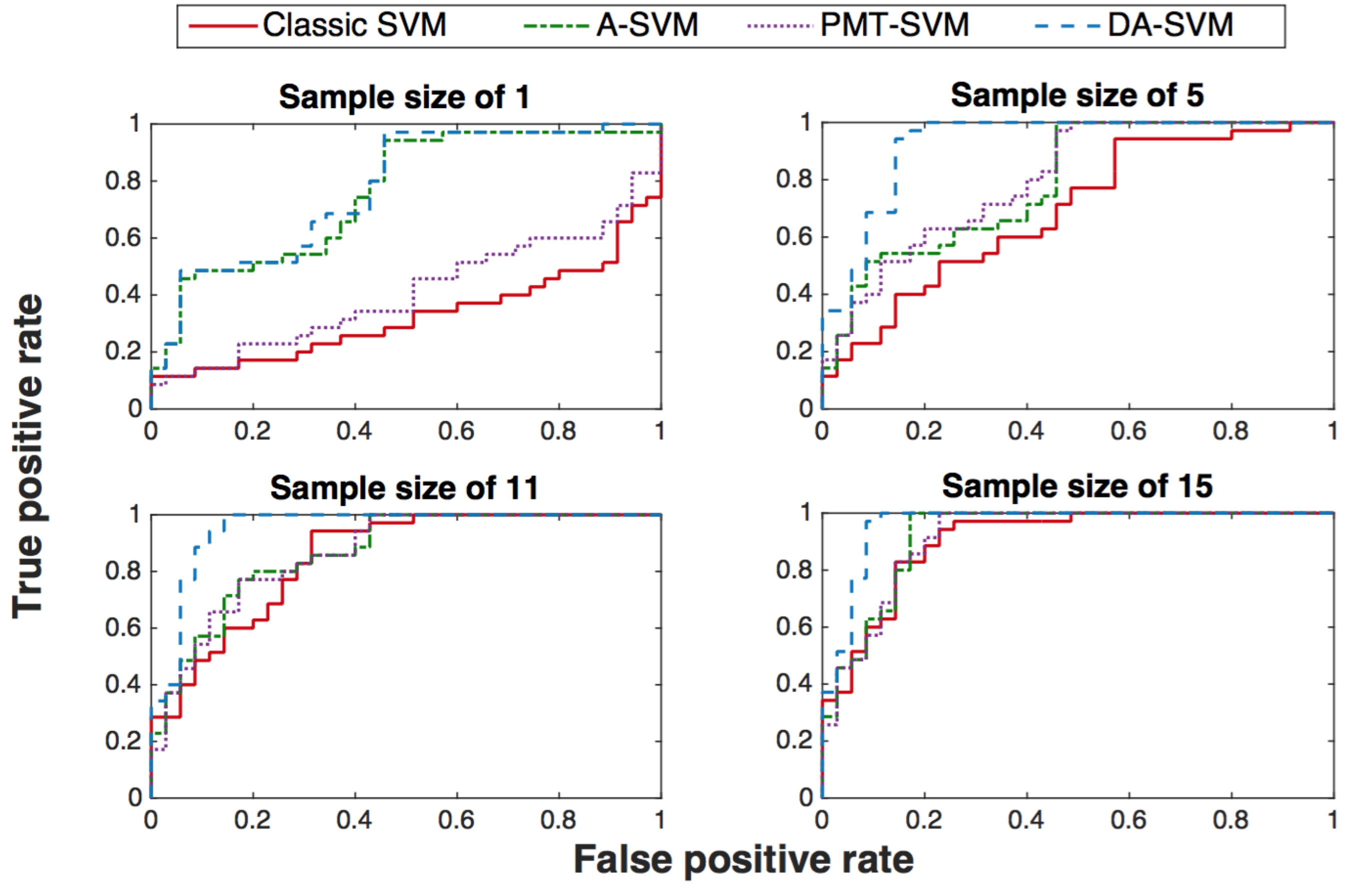


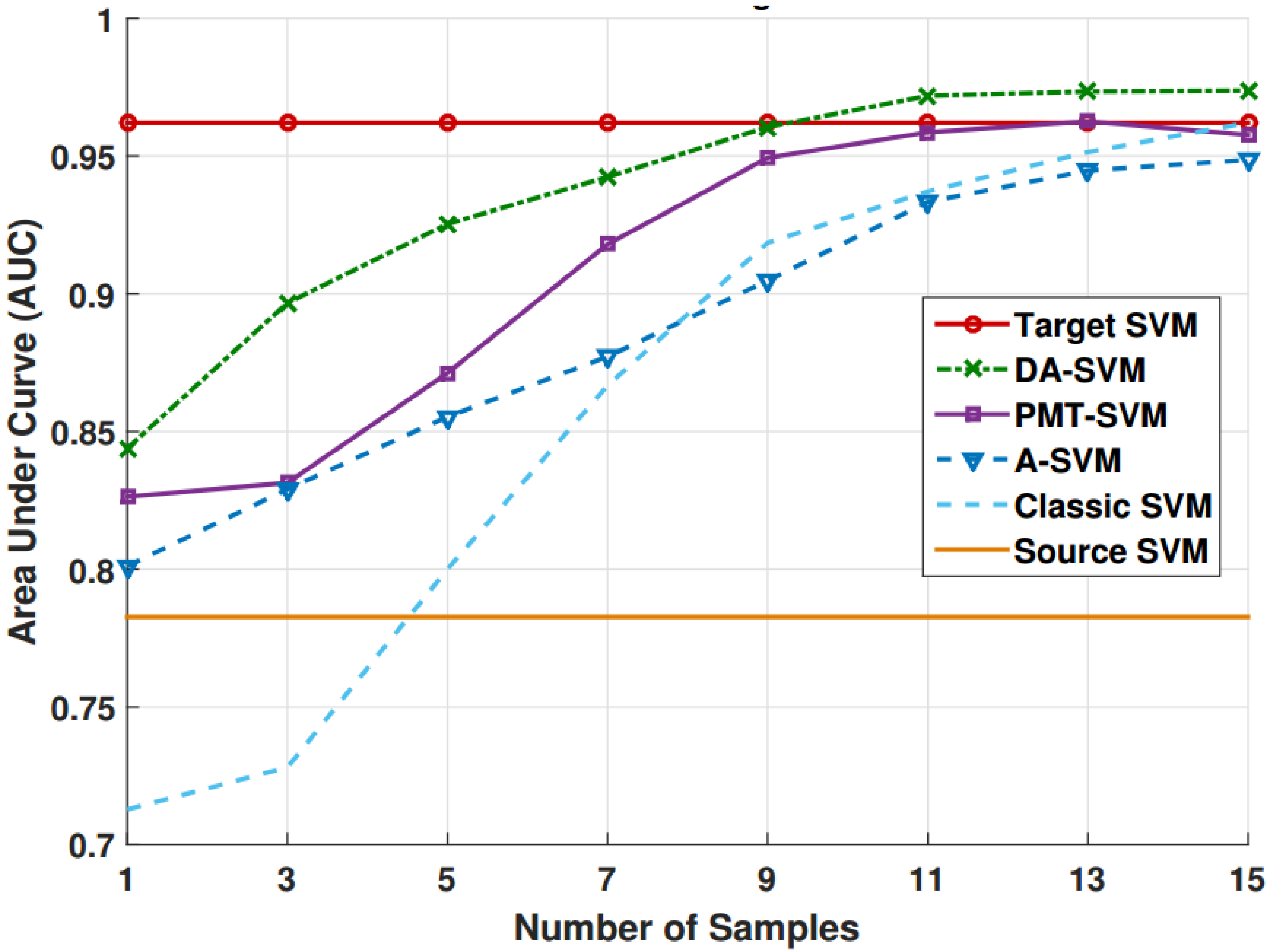
Figure 4: ROC with various step sizes

Sample Size:	1	5	11	15
Classic SVM	**71.28** ± 11.39	**80.01** ± 9.01	**93.71** ± 4.55	**96.20** ± 3.42
A-SVM	**80.11** ± 2.61	**85.54** ± 3.49	**93.33** ± 2.33	**94.86** ± 1.84
PMT-SVM	**82.64** ± 9.00	**87.12** ± 6.95	**95.86** ± 2.74	**95.76** ± 2.75
DA-SVM	**84.39** ± 4.03	**92.53** ± 3.42	**97.18** ± 1.10	**97.37** ± 1.01

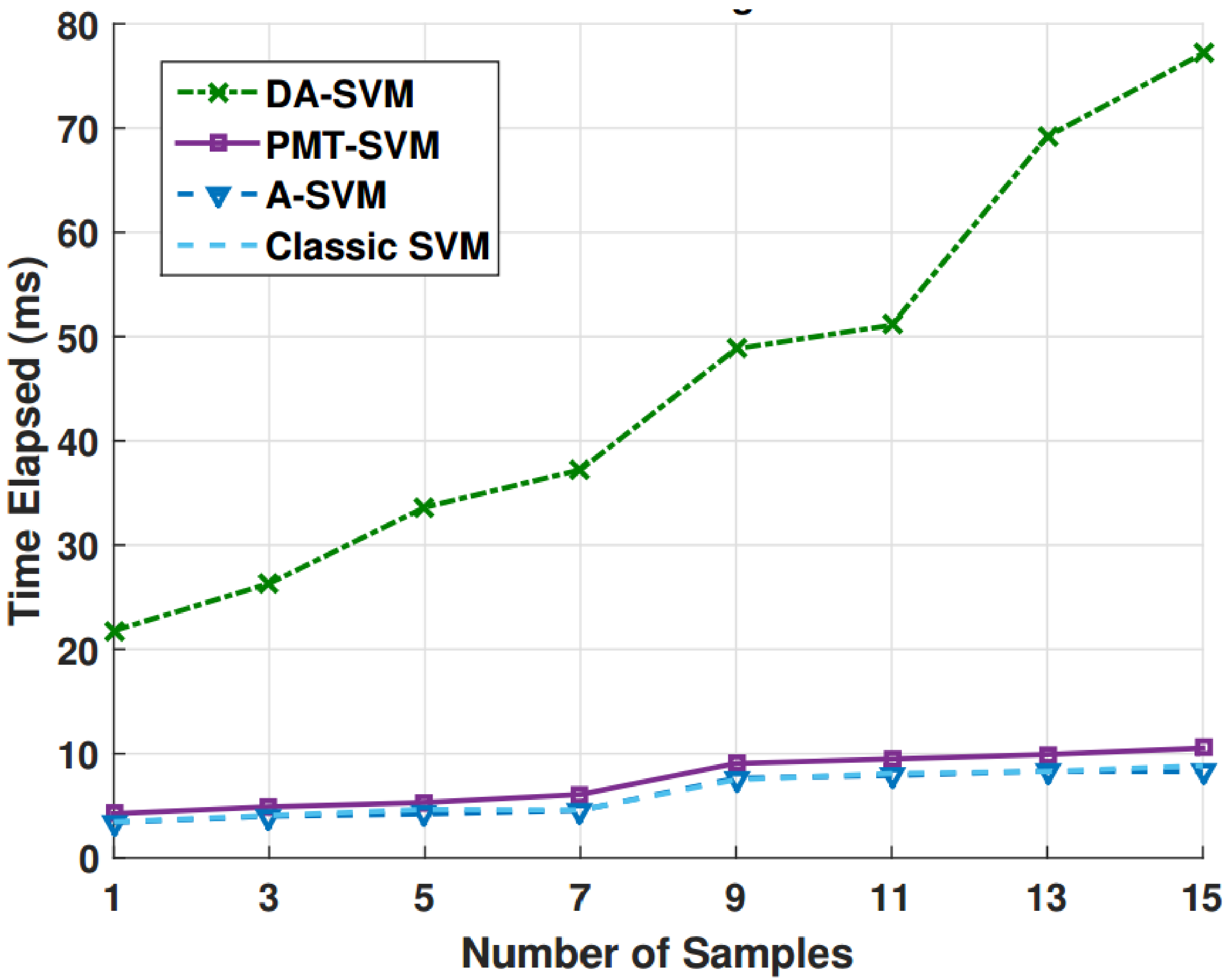
AUC values are multiplied with 100 for higher precision

*Killourhy, K. S., & Maxion, R. A. (2009).
Comparing anomaly-detection algorithms for keystroke dynamics. In *Dependable Systems & Networks, 2009. DSN'09. IEEE/IFIP International Conference on*

Comparison of SVM Algorithms



Performance



Some Observations

- GMM provides improvement over single Gaussian
- Fusion can improve the accuracy
- SVM can be utilized for long-text data efficiently
- Intra-user variability is important
  - Can be addressed by using transfer learning

Outline of the Talk

Introduction

- General approach to continuous authentication

Keystroke dynamics and mouse movements

- Feature selection
- Methodology - Gaussian model, SVM, transfer learning
- Datasets and anonymization

Results

- GMM, SVM, transfer learning

Research directions

- Secondary features
- Deep learning
- Adversarial learning
- Extension to network of smart devices

Secondary Features – Achieving More with Less

Punctuations

- Period
- Comma
- …

Functional keys

- Shift
- Backspace

Number and others

- 1, 2, 3 …
- Dash

Compare with primary features

- Primarily from 26 letters (A to Z)

Feature Extraction

Dwell time

- Period, Comma, Tab, Space, Enter, Backspace, Arrow keys, Number keys, Dash

Flight time

- Period – Space and Comma – Space
- Shift – [a-z 0-9]
- Ctrl – [a-z]

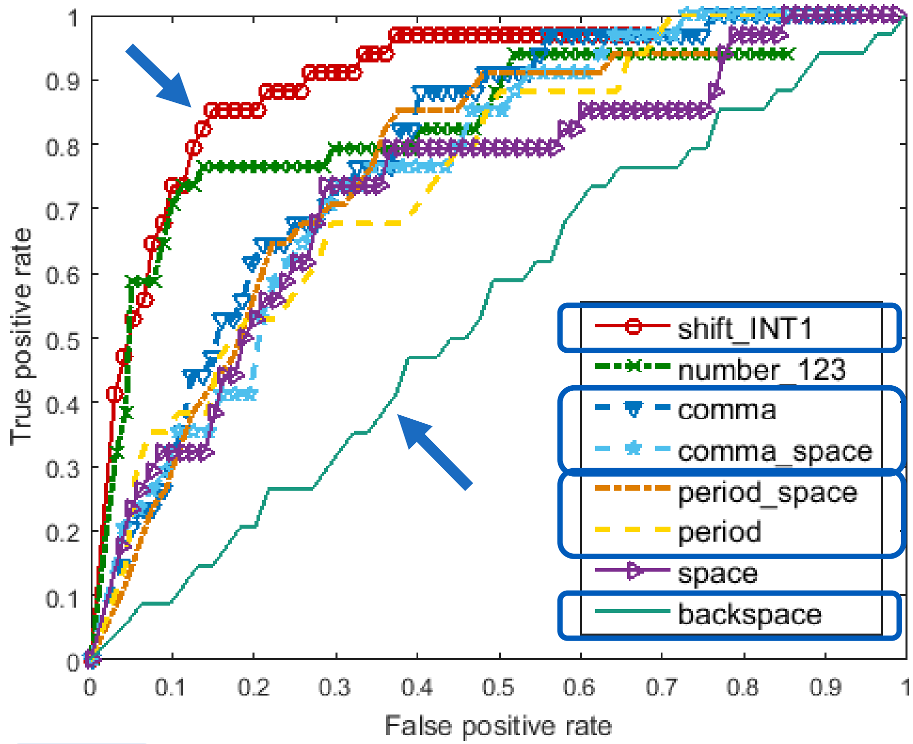
Available Secondary Features in the Clarkson's dataset

Feature	Feature # Type	Average # records	# Occurrence (out of 34)
Backspace	Dwell	1010.47	34
Space	Dwell	2535.85	34
Number 1	Dwell	64.03	34
Number 2	Dwell	36.23	31
Number 3	Dwell	32.35	31
Shift_I	Flight	89.03	31
Shift_N	Flight	33.5	28
Shift_T	Flight	22.6	30
Shift_1	Flight	29.75	32
Comma	Dwell	118.38	34
Comma_Space	Flight	116.38	34
Period	Dwell	162.68	34
Period_Space	Flight	155.71	34
Dash	Dwell	19.53	30
LeftArrow	Dwell	30.94	16
RightArrow	Dwell	33.15	13

Single Feature Evaluation

- 8 features (groups)
- Best −> Shift group
- Worst −> Backspace
- Comma > Comma-Space
- Period < Period-Space

ROC curves



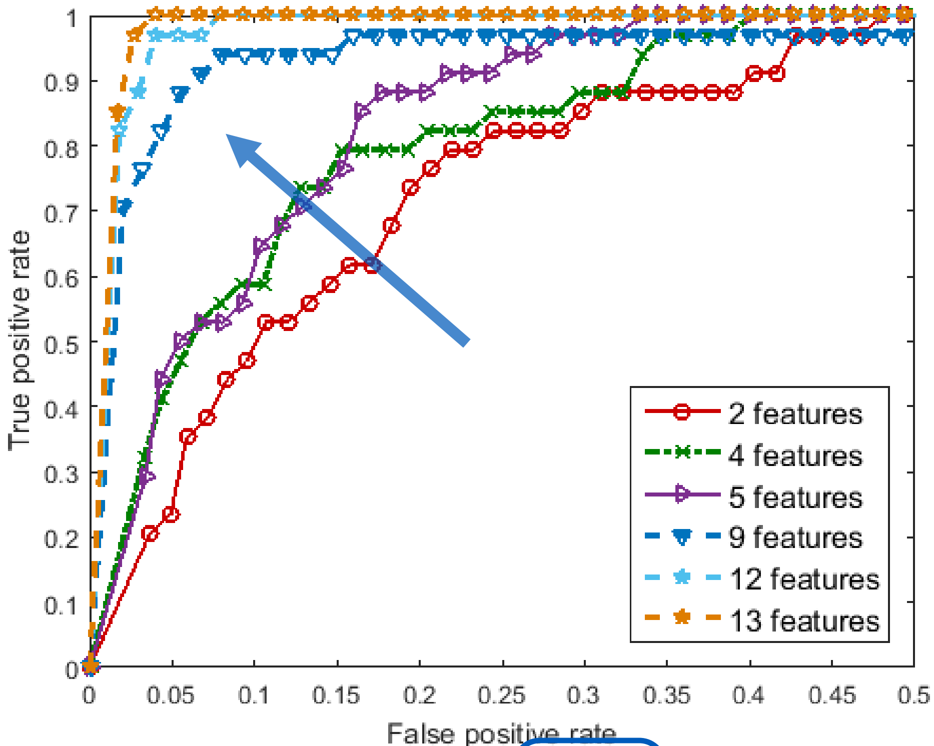
	Backspace	Space	Num(123)	Shift(INT1)	Comma	Comma-Space	Period	Period-Space
Kernel Scale	0.5	0.25	0.01	0.15	0.5	0.45	0.3	0.01
AUC	0.5401	0.7307	0.8323	0.8991	0.7877	0.7599	0.7493	0.7636
EER (%)	47.5	27.8	23.53	14.71	28.88	29.41	32.35	29.41

Overall Evaluation

- 2 : Comma and Period-Space
- 4 : + Left & Right arrow
- 5 : + Dash
- 9 : + Shift [I N T 1]
- 12 : + Number [1 2 3]
- 13 : + Space

  (Data Sampling)

ROC curves



# Features	2	4	5	9	12	13
AUC	0.8521	0.8932	0.9088	0.953	0.9897	0.9937
EER (%)	21.57	19.61	16.22	6.95	3.83	2.94

Comparison with Primary Features

Study	# users	EER (%)
Atam et al. [8]	43	8.77
Killourhy et al. [12]	51	10.2
Giot et al. [7]	100	6.96
Gabriel et al. [1]	24	1.57
Rahman et al. [13]	50	10
Kaneko et al. [10]	51	0.84
Ceker et al. [4]	30	0.08
Our work	34	2.94

Deep Learning

- Current solutions in keystroke dynamics
  - Use timing information between the keys separately (digraph, trigraph, n-graphs), fusion
  - Trial and error works, but unwieldy
  - Computational complexity increases exponentially
- Scaling up (no. of users) would mean lower accuracy
- CNNs can provide a deeper architecture and unify ML techniques by consolidating the power of various features
- CNN has been successfully applied in vision, speech, NLP

Adversarial Learning

Attack scenarios

- Adjust attack based on the feedback on where the typing was different from the legitimate users
- Synthetic forgery attacks designed to mimic the legitimate users based on their typing profiles

Possible solution

- Combine multiple biometrics features
  - E.g., Keystroke dynamic and mouse movement

Extension to Smartphone Environment

- Portable mobile devices have become ubiquitous
  - Sensitive data, business usage
  - Owner may leave this device unlocked
  - There is security risk
- What features can be extracted?
  - User activities on the mobile devices touch-screen - clicks, speed, angles of movement, number of clicks during a session, pressure on the touchscreen
- Accelerometer, rotation vector and orientation sensor to generate the feature vectors
- We can apply a variety of ML algorithms in this context

Publications

- Yan Sun and Shambhu Upadhyaya, “Secure and privacy preserving data processing support for active authentication”, *Information Systems Frontie*rs 17, no. 5 (2015): 1007-1015.
- Hayreddin Çeker and Shambhu Upadhyaya, “Enhanced recognition of keystroke dynamics using Gaussian mixture models“, *IEEE MILCOM*, 2015.
- Hayreddin Ceker and Shambhu Upadhyaya, "Enhanced Recognition of Keystroke Dynamics in Long-Text Data“, *IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS*) 2016.
- Hayreddin Ceker and Shambhu Upadhyaya, "Adaptive Techniques to Address Intra-User Variability in Keystroke Dynamics“, *IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS)* 2016.
- Yan Sun, Hayreddin Ceker and Shambhu Upadhyaya, “Shared Keystroke Dataset for Continuous Authentication”, *8th IEEE International Workshop on Information Forensics and Security (WIFS)* 2016.
- Yan Sun, Hayreddin Ceker and Shambhu Upadhyaya, “Anatomy of Secondary Features in Keystroke Dynamics – Achieving More with Less”, *IEEE International Conference on Identity, Security and Behavior Analysis (ISBA*), 2017.
- Hayreddin Ceker and S. Upadhyaya, "Transfer Learning in Long-Text Keystroke Dynamics", *IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, 2017.