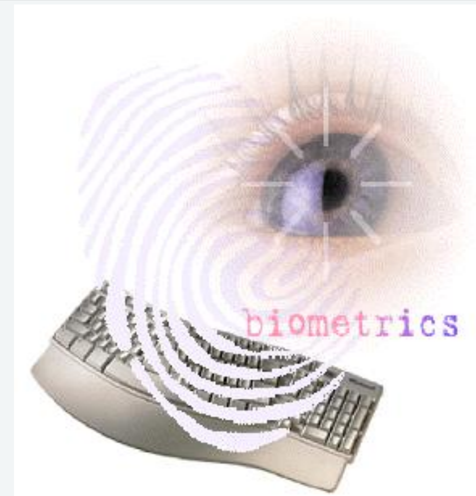
DAI-Labor
TU Berlin

Continuous User Verification through Behavior Biometrics



Dipl.-Inform Arik Messerman

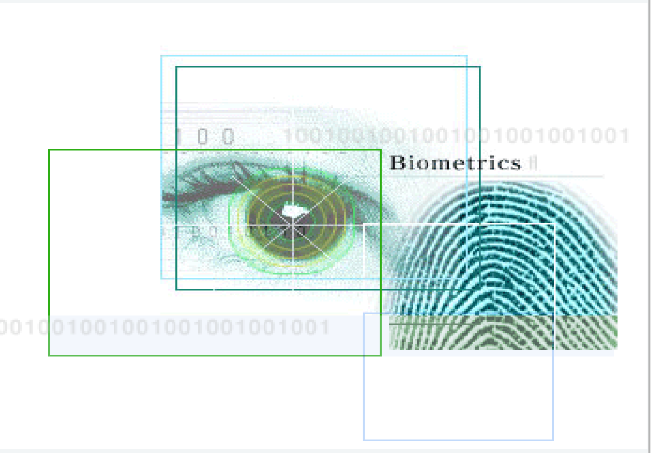
CC SEC
Security

Agenda

1. Introduction, Motivation
2. Focus
3. Deeper view to a Keystroke Dynamic Approach
4. Milestones, Discussion

Introduction to Biometrics

- Physiological Biometric
  - Passive approach
    - Measure distinct traits that humans have
    - Do not vary over time
  - Iris scans, retina scans, fingerprints, DNA, …

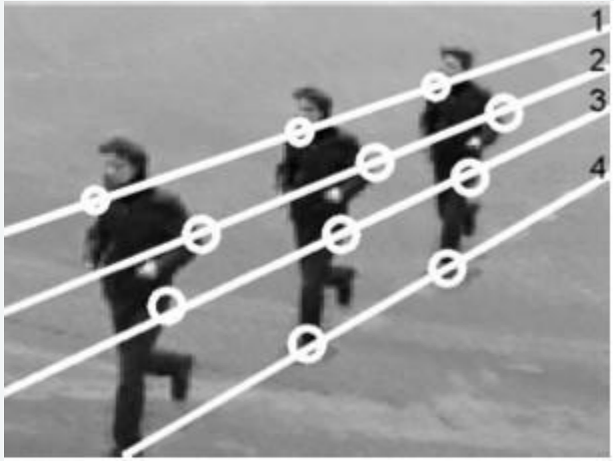


- Behavioral Biometric
  - Active approach
    - Measure performed tasks
    - Do vary over time
  - Types of behavioral Biometrics
  - Each subdivision has its own characteristics in terms of
    - usage, deploy ability, user acceptance, quality, …

Motivation (Behavioral Biometrics)

DAI-Labor
TU Berlin

- Humans can be verified traditionally by / through …
  - Knowledge (passwords, PINs, ..)
  - Ownership (software/security token, ID card, …)
  - Inherence   (fingerprint, voice, interaction, …)
    - In most cases: Physiological Biometrics
- Risk for traditional solutions:
  - Object is verified, object == actor ?
- Additional security layer is required
  →behavioral biometrics
  - Further requirements



Agenda

DAI-Labor
TU Berlin

1. Introduction, Motivation
2. Focus
3. Deeper view to a Keystroke Dynamic Approach
4. Milestones, Discussion

Focus: Transparent Continuous Verification

- Focus: "Development and trustable evaluation of reactive, transparent and free-action-based continuous user verification solutions with low error rates under real-time environments and conditions with minimum and user-friendly requirements to stakeholders through Keystroke Dynamic approaches in the field of Behavioral Biometrics."
- User verification
  - Free text → Most solutions: Fix
  - Continuous → Most solutions: Initial
  - Transparent → Most solutions: Defined action to perform required
  - Low error rates → Many solutions: Evaluation under unreal environments
  - Short response times → Many solutions: Not really considered
  - User model update → Most solutions: Static enrollment
  - Large user data → Most solutions: Evaluation based on a very limited amount of data
  - Comparable evaluations (Open DB) → In the field of Keystroke Dynamics not given
- Deployment
  - With minimum effort into real-time environments
  - Without any additional hardware-equipment

Agenda

DAI-Labor
TU Berlin

1. Introduction, Motivation
2. Focus
3. Deeper view to a Keystroke Dynamic Approach
4. Milestones, Discussion

Edit distance calculation (Free text)*

DAI-Labor
TU Berlin

Database of user u:

ab	ll	ff	ce	ef	by	gk	gl	ew	kl	mn	op	oz	qr	th	uv	wx	nt	yz
10ms	11ms	14ms	15ms	16ms	17ms	19ms	20ms	20ms	23ms	24ms	27ms	27ms	28ms	30ms	33ms	35ms	36ms	41ms

New typing sample: „i will buy a new table ...open the door, ... efficient“

* based on: Daniele Gunetti and Claudia Picardi. **Keystroke analysis of free text.** ACM Trans. Inf. Syst. Secur., 8(3):312-347, 2005.

Distance calculation*

Database of user u:

ab	**ll**	**ff**	ce	ef	**uy**	gk	gl	**ew**	kl	mn	**op**	oz	qr	**th**	uv	wx	**nt**	yz
10ms	11ms	14ms	15ms	16ms	17ms	19ms	20ms	20ms	23ms	24ms	27ms	27ms	28ms	30ms	33ms	35ms	36ms	41ms

New typing sample: „i will buy a new table …open the door, … efficient"

* based on: Daniele Gunetti and Claudia Picardi. **Keystroke analysis of free text.** ACM Trans. Inf. Syst. Secur., 8(3):312-347, 2005.

Distance calculation*

Database of user u:

ab	ll	ff	ce	ef	gk	gl	by	ew	kl	mn	op	qz	qr	th	uv	wx	nt	yz
10ms	11ms	14ms	15ms	16ms	17ms	19ms	20ms	20ms	23ms	24ms	27ms	27ms	28ms	30ms	33ms	35ms	36ms	41ms

New typing sample: „i will buy a new table …open the door, … efficient“

ab	ll	ff	th	uy	ew	op	nt
14ms	16ms	18ms	19ms	21ms	28ms	31ms	38ms

* based on: Daniele Gunetti and Claudia Picardi. **Keystroke analysis of free text.** ACM Trans. Inf. Syst. Secur., 8(3):312-347, 2005.

Distance calculation*

Database of user u:

ab	ll	ff	uy	ew	op	th	nt
10ms	11ms	14ms	17ms	20ms	27ms	30ms	36ms

New typing sample: „i will buy a new table …open the door, … efficient“

ab	ll	ff	th	uy	ew	op	nt
14ms	16ms	18ms	19ms	21ms	28ms	31ms	38ms

Distance:

$(1+1+1+3)/(0.5 \times 8^2) = 0.1875$

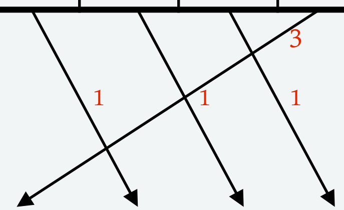
* based on: Daniele Gunetti and Claudia Picardi. **Keystroke analysis of free text.** ACM Trans. Inf. Syst. Secur., 8(3):312-347, 2005.

Distance calculation*

DAI-Labor
TU Berlin

Database of user u:

ab	ll	ff	uy	ew	op	th	nt
10ms	11ms	14ms	17ms	20ms	27ms	30ms	36ms



Distance:
(1+1+1+3)/(0.5 [illegible] = 0.1875

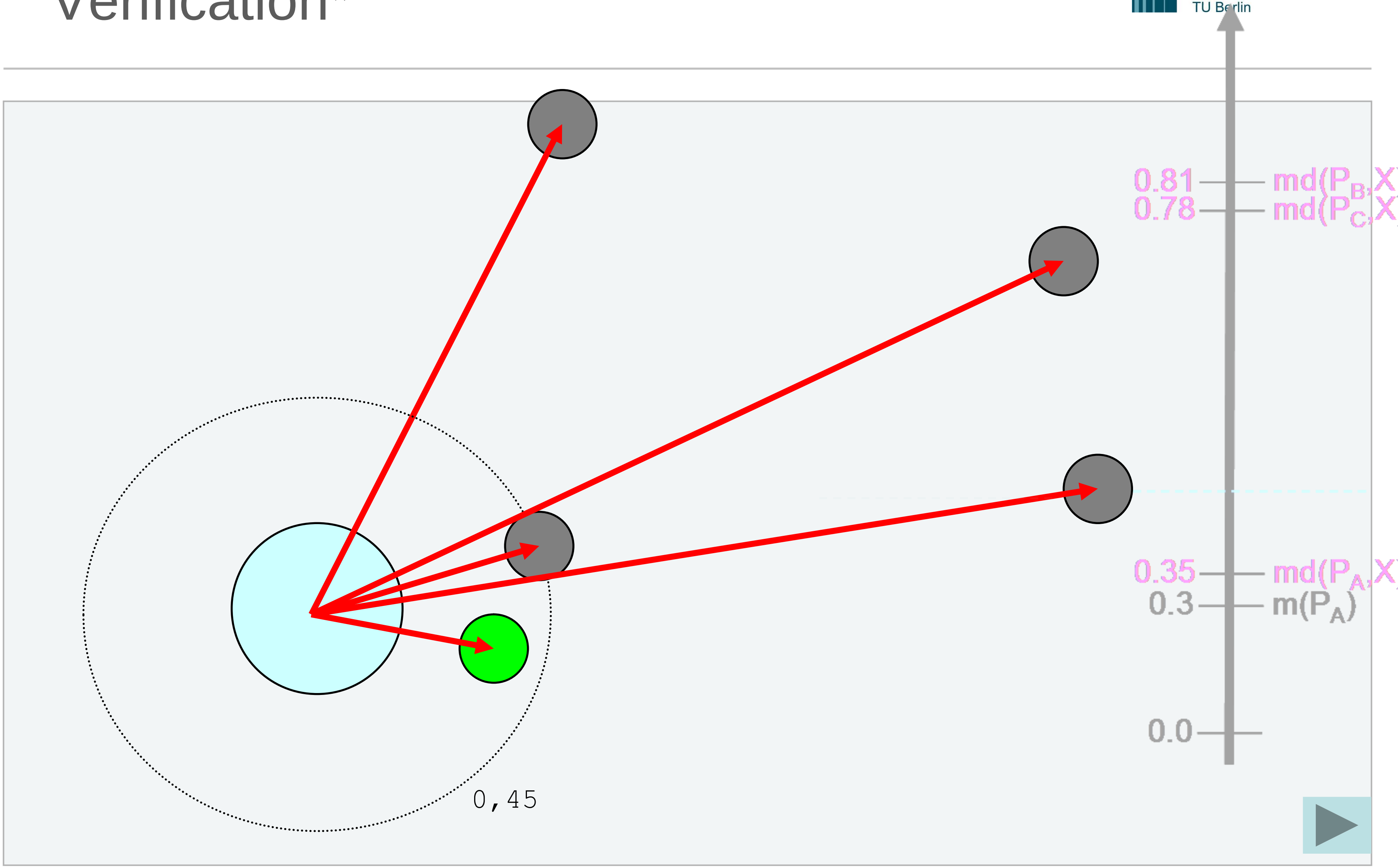
New typing sample:

ab	ll	ff	th	uy	ew	op	nt
14ms 28	16ms 32	18ms 36	19ms 38	21ms 42	28ms 56	31ms 62	38ms 76

- Pattern of a user can be regarded as an array with values
- Calculation of the distance between patterns from the user data base and new one is to reduce to the calculation of the position of elements in permutations

* based on: Daniele Gunetti and Claudia Picardi. **Keystroke analysis of free text.** ACM Trans. Inf. Syst. Secur., 8(3):312-347, 2005.

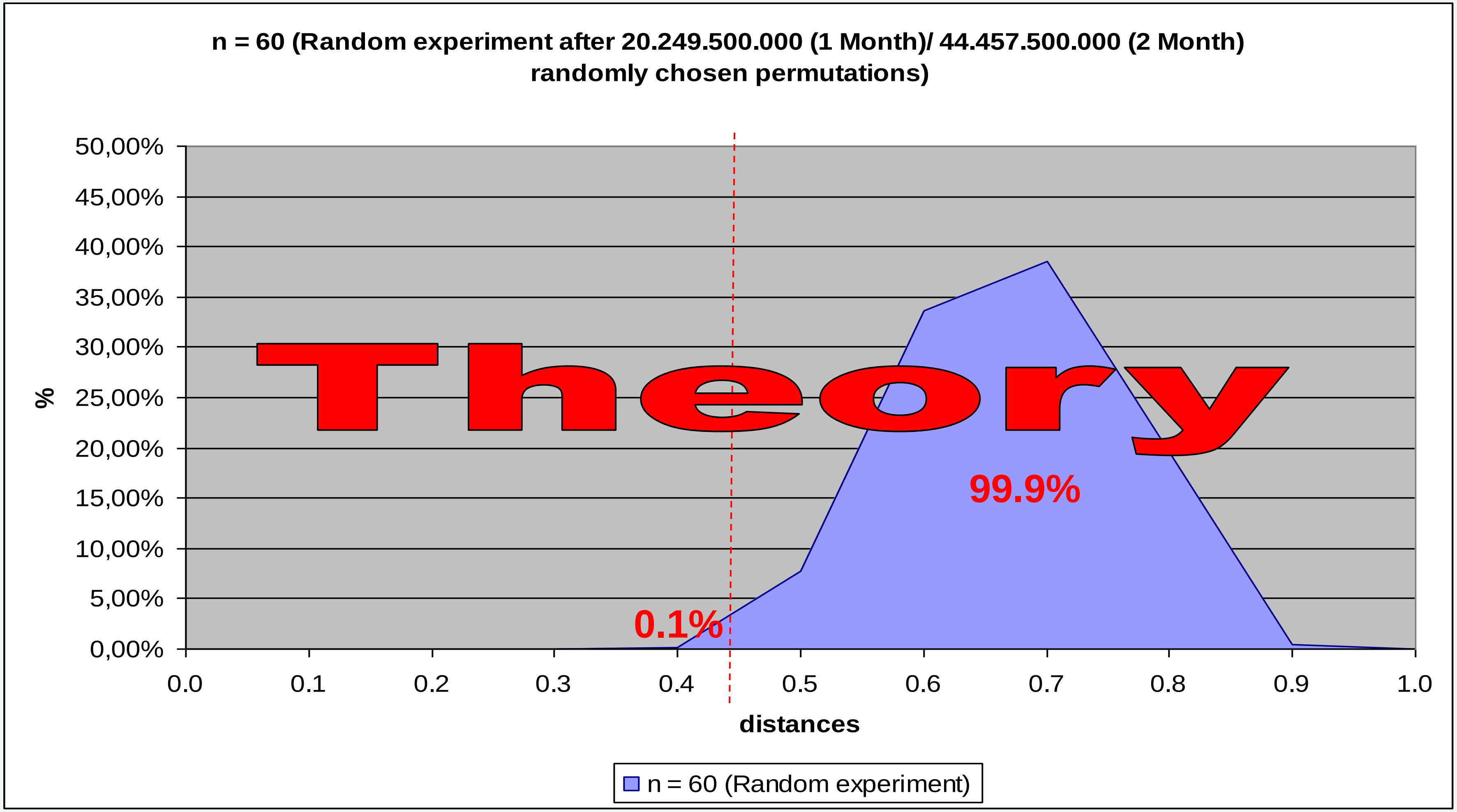
Verification*



* based on: Daniele Gunetti and Claudia Picardi. **Keystroke analysis of free text.** ACM Trans. Inf. Syst. Secur., 8(3):312-347, 2005.

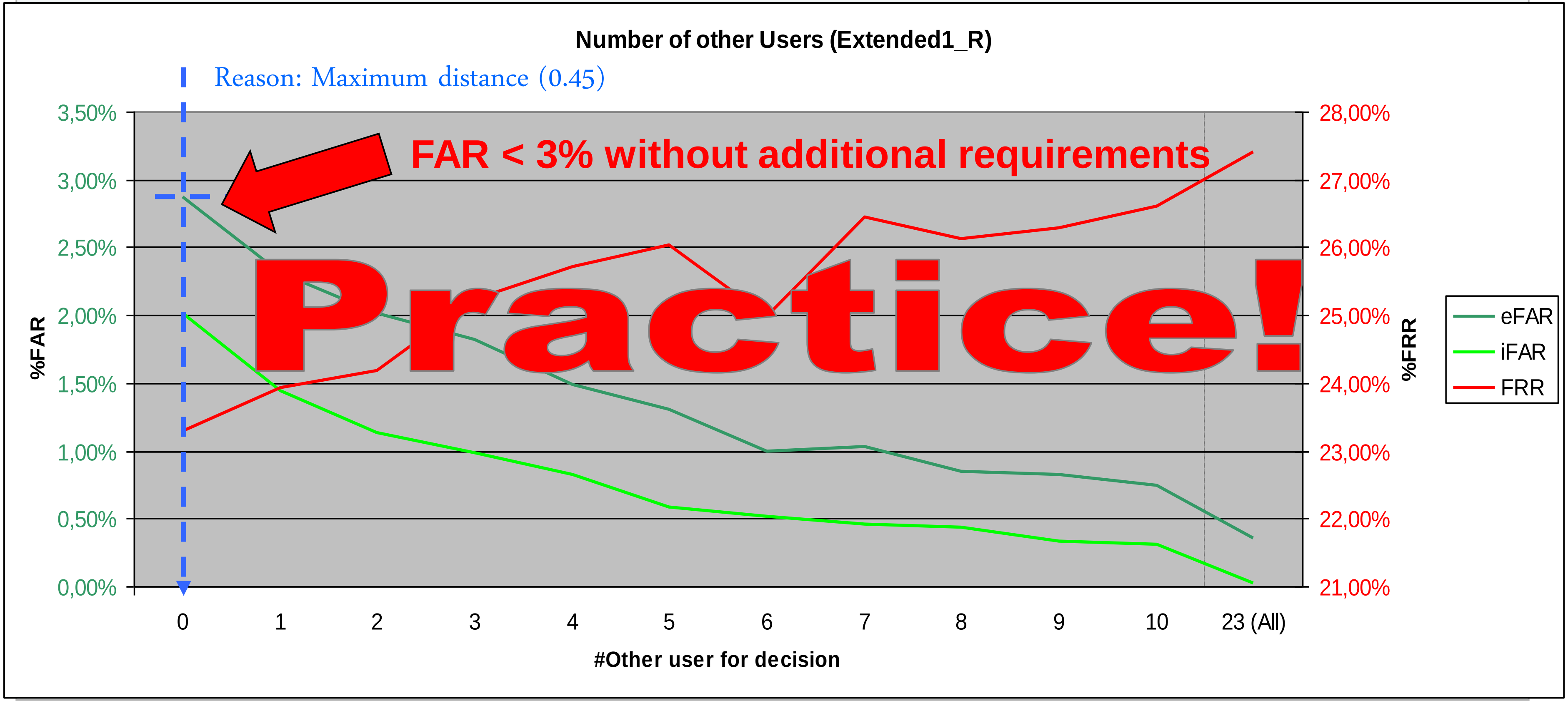
n = 60 → n! = 60! = 8,321 x 10^{81} (after 2 Month on 8 CPU)

DAI-Labor TU Berlin



Experiment with real user data

SQL-Dump Date:	AbV 20090409 0422.sql
Number of Events per Verification:	50
Event buffer size (maximum):	400
Number of other users for decision:	1-23
Event vector size:	700
Number of full required vectors:	5
Number of total used vectors:	7
Legitimate User:	23
Attacker:	17/23
Dynamic	yes



Bioinformatics approach

- Distance calculation → Edit distance
- Evolution-Theory:
  - Combinations of amino acids are specified through sequences of nucleotides in DNA → Genes
    - Edit distances between DNA, RNA or protein strings
    - Protein: Sequence of units = amino acids
  - Example: glyceraldehyd3-phosphate dehydrogenase (GADPH) protein
    - `Fly:        GAKKVIISAPSAD-APM—F`
    - `Human:      GAKRVIISAPSAD-APM-F`
    - `Yeast:      GAKKVVSTAPSS-TPM—F`
  - How closely related are two strings which represent the amino acid sequence of a particular gene between two species?
  - From a computer science perspective this issue is one of pattern matching and search

Bioinformatics approach

From a computer science perspective this issue is one of pattern matching and search

- Idea:
  - Apply the huge amount practical and theoretical research that have been successfully developed in bioinformatics to the task of authentication/verification [/Anomaly detection]
- Other distance metrics:
  - Levenshtein distance:
    - Levenshtein distance between two strings is given by the minimum number of operations needed to transform one string into the other, where an operation is an insertion, deletion, or substitution (???) of a single character
  - Hamming, Euclidian, Cayley, Ulam, Spearman's Footrule, Spearman's rank correlation, Kendall's tau …
  - From a computer science perspective this issue is one of pattern matching and search

Agenda

1. Introduction, Motivation
2. Focus
3. Deeper view to a Keystroke Dynamic Approach
4. Milestones, Discussion

Milestones (1)

- Parallel work in project “Activity-based Verification” (until 06.2010)
  - Work packages for DAI reflect next steps of my dissertation
    - e.g. new distance metrics/approaches
- IEEE ISI 2009: “Identity Theft, Computers and Behavioral Biometrics” (Jun 09), Dallas, TX
- Survey journal: Draft version (45 pages)
  - State of the art, deeper discussions, review and novel views in the field of B.B.
- Verification & Evaluation Service Platform paper: Pre-draft version
  - Conception and Implementation of a generic platform was made
- Several verification approaches
  - Focus: Continuous Free Text Verification

Milestones (2)

- Web mail application
  - Enable transparent collection of behavioral data
  - Large dataset of user behavior
  - Currently: 52 users, ~5000 ‘KeyDown’ events, Goal: ~100 users
- Theoretical/Scientific work paper: Paper planned (Start Sept. 09)
- Smart Senior: "Erkennung von Notsituationen im häuslichen Umfeld durch sensorbasierte Analyse von Verhaltensanomalien" Paper planend (Start Sept. 09)
  - Adapt knowledge made in the field of Behavioral Biometrics to anomaly detection
- Bachelor/Diploma thesis
  - 1. Adoption of existing (own) methods to login verification (Start 08.2009)
  - 2. Generic evaluation engine of AbV verification methods (Start 07.2009)
  - 3. Adoption of existing (own) methods to Smart Phone environments (Start 08.2009)

Some References

DAI-Labor
TU Berlin

- Moskovitch R., Feher C., Messerman A., Kirschnick N., Mustafić T., Camtepe A., Löhlein B., Heister U., Möller S., Rokach L., Elovici Y. **Identity Theft, Computers and Behavioral Biometrics.** IEEE Intelligence and Security Informatics, 2009. ISI '09., 2009
- K. Revett. **Behavioral Biometrics.** Wiley, J, 2008. ISBN: 978-0-470-51883-0.
- K. Revett, P.S. Magalhaes, and H.D. Santos. **Developing a keystroke dynamics based agent using rough sets.** In 2005 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology. University of Technology of Compiegne, 2005.
- A. Peacock, X. Ke, and M. Wilkerson. **Typing Patterns: A Key to User Identication.** IEEE SECURITY & PRIVACY, 2:40{47, 2004.
- Daniele Gunetti and Claudia Picardi. **Keystroke analysis of free text.** ACM Trans. Inf. Syst. Secur., 8(3):312{347, 2005.
- Hu, J, Gingrich D., Sentosa, A. **A k-Nearest Neighbor Approach for User Authentication through Biometric Keystroke Dynamics**, Communications, 2008. ICC '08. IEEE International Conference, pages 1556-1560
- K. Hempstalk, **Continuous Typist Verification using Machine Learning**, PhD, University of Waikato, New Zealand, Department of Computer Science, 2009
- R. Bolle, Ratha N., and Pankanti S. **Performance evaluation in 1:1 biometric engines**, volume 3338/2005, chapter Biometrics, pages 27{46. Springer Berlin / Heidelberg, 2005. IBM Thomas J Watson Research Center, Yorktown Heights, NY 10598, ETATS-UNIS. 9. Kuan-Ta Chen and Li-Wen
- G. Bartolacci, M. Curtin, M. Katzenberg, N. Nwana, S.H. Cha, and CC Tappert. **Long-Text Keystroke Biometric Applications over the Internet**, 2005.

Contact

DAI-Labor TU Berlin

Dipl.-Inform. Arik Messerman

Researcher
Competence Center Security

+49 (0) 30 / 314 – 74 047
+49 (0) 30 / 314 – 74 003
arik.messerman@dai-labor.de

www.dai-labor.de

DAI-Labor · Technische Universität Berlin · Sekretariat TEL 14
Fakultät IV - Elektrotechnik und Informatik
Ernst-Reuter-Platz 7 · D -10587 Berlin