Evaluating Biometric Security: Understanding the Impact of Wolves in Sheep's Clothing

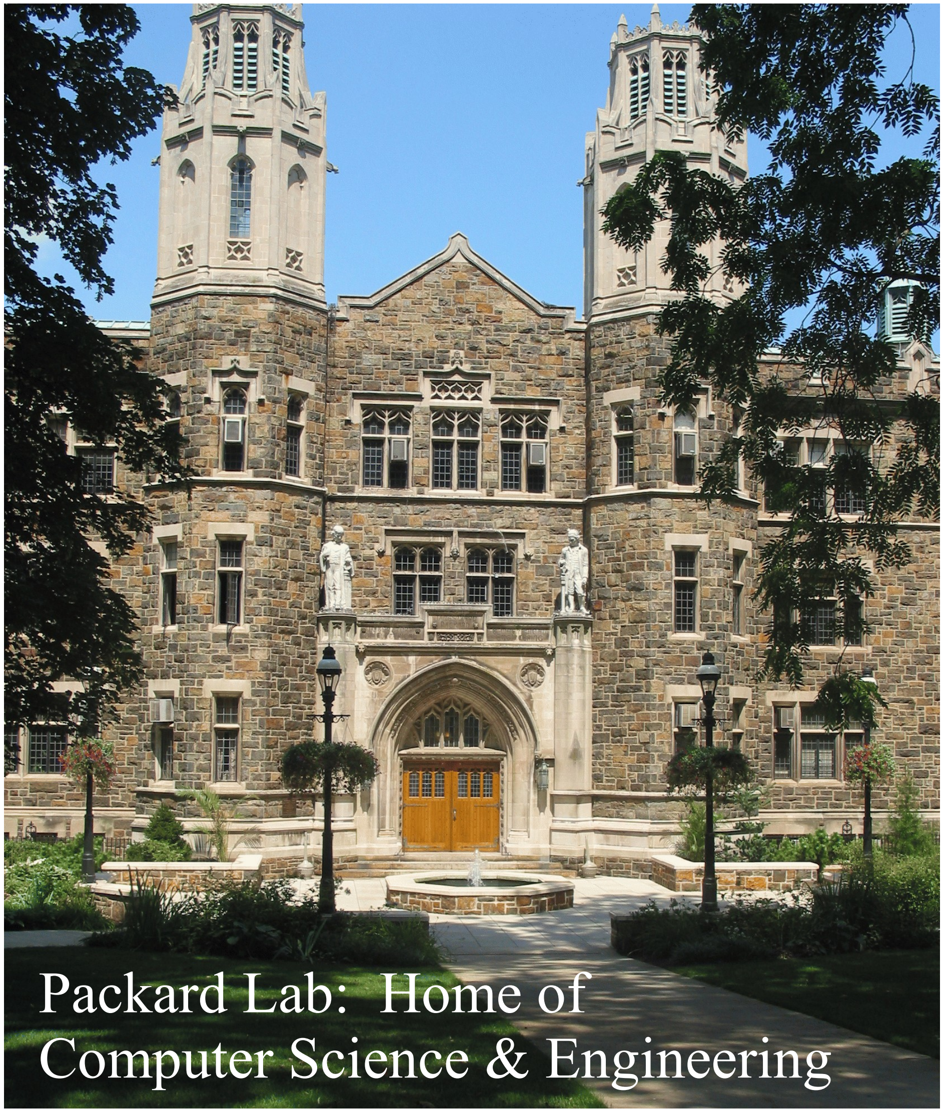
Daniel Lopresti[1], *Fabian Monrose*[2], *and Lucas Ballard*[2]

November 2006

[1] Lehigh University
Bethlehem, PA 18015, USA
lopresti@cse.lehigh.edu

[2] Johns Hopkins University
Baltimore, MD 21218, USA
{fabian,lucas}@cs.jhu.edu

Lehigh University

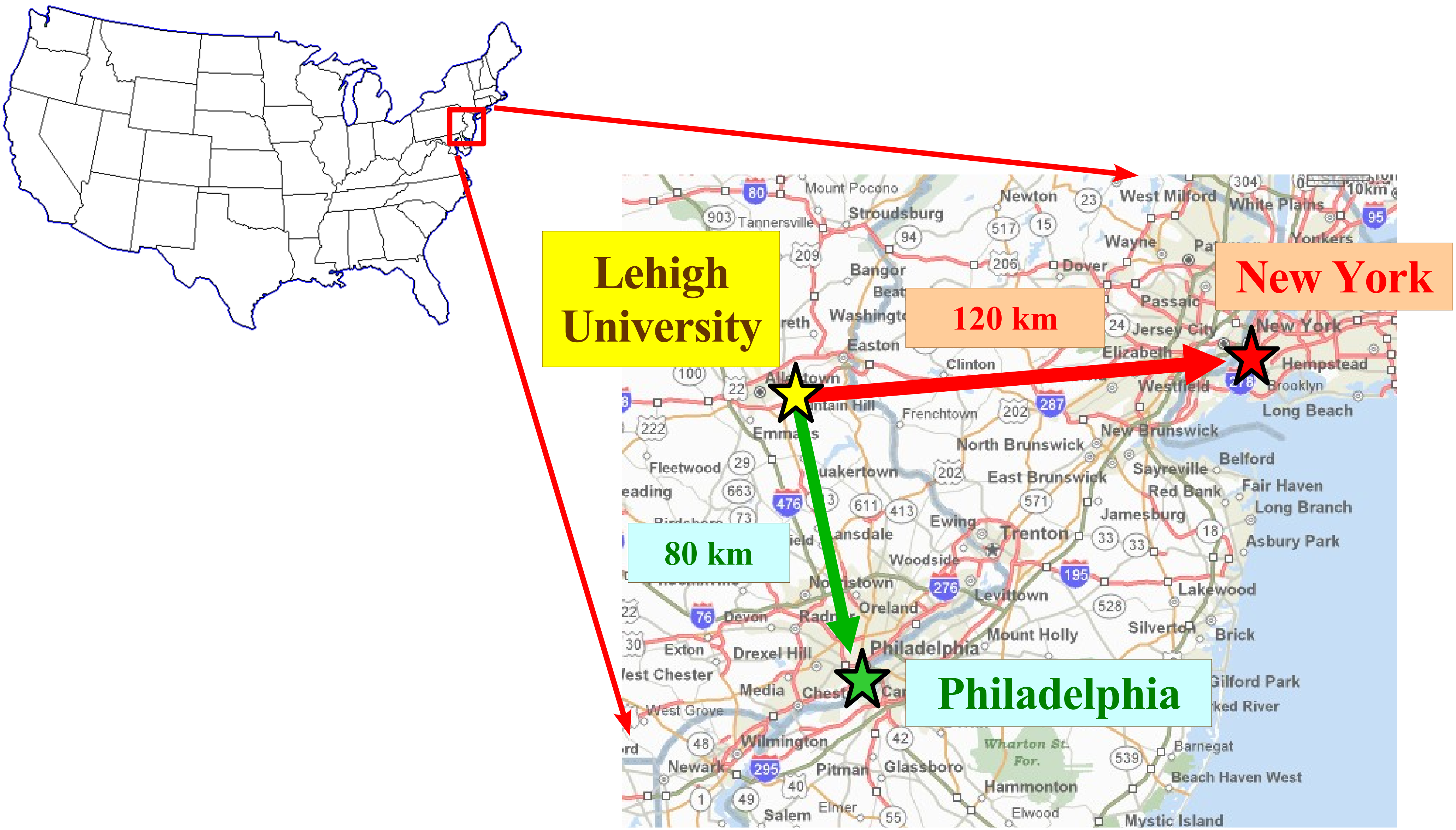


Packard Lab: Home of
Computer Science & Engineering

Key facts about Lehigh:

- A research university founded in 1865.
- Four colleges: Engineering, Arts & Sciences, Business, Education.
- Faculty = 441 full-time.
- Graduate students = 2,064.
- Undergraduates = 4,577.
- Three campuses spread over 1,600 acres (mountain side, wooded).
- Located in northeastern U.S. (about 1.5 hours from New York and Philadelphia, 3 hours from Washington, DC).
- Engineering College ranked in top 20% of Ph.D.-granting schools in U.S.
- University ranked in top 15% of U.S. national universities.

Lehigh University



Main Message

Prevailing methodologies for evaluating biometric security are inadequate in some important ways.

Current schemes:

- Fall far short of measuring real threats, and present a view of security that is too optimistic.
- Have arisen from pattern recognition research and allow for noisy inputs, but not for true adversaries.

Better model comes from computer security field: determined adversaries having time and resources.

Talk Overview

- Motivation
- Biometric Authentication / Key Generation
- Handwriting as an Exemplar Biometric
- Evaluating Security Under Determined Adversaries
- Generative Attacks on Handwriting Biometrics
- Conclusions and Recommendations

Motivation (Actually, Coincidence)

A scene from recent thriller *Mission Impossible 3*:

- Good guy (Tom Cruise) forces bad guy (Philip Seymour Hoffman) to read random-sounding text from index card ...

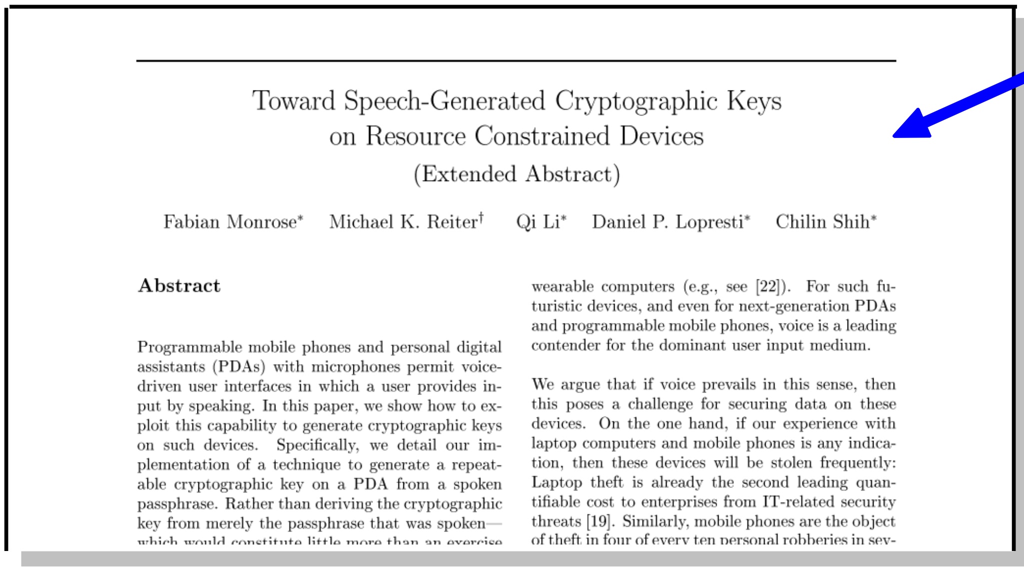
- ... which good guys use to compile a speech synthesizer that can perfectly mimic bad guy's voice.

Is this scenario plausible, or just science fiction?

Is Such a Threat Real?

Minus a few details, the threat as depicted is very real.

2002 paper describing same basic idea shown in movie



Toward Speech-Generated Cryptographic Keys
on Resource Constrained Devices
(Extended Abstract)

Fabian Monrose* Michael K. Reiter† Qi Li* Daniel P. Lopresti* Chilin Shih*

Abstract

Programmable mobile phones and personal digital assistants (PDAs) with microphones permit voice-driven user interfaces in which a user provides input by speaking. In this paper, we show how to exploit this capability to generate cryptographic keys on such devices. Specifically, we detail our implementation of a technique to generate a repeatable cryptographic key on a PDA from a spoken passphrase. Rather than deriving the cryptographic key from merely the passphrase that was spoken—which would constitute little more than an exercise

wearable computers (e.g., see [22]). For such futuristic devices, and even for next-generation PDAs and programmable mobile phones, voice is a leading contender for the dominant user input medium.

We argue that if voice prevails in this sense, then this poses a challenge for securing data on these devices. On the one hand, if our experience with laptop computers and mobile phones is any indication, then these devices will be stolen frequently: Laptop theft is already the second leading quantifiable cost to enterprises from IT-related security threats [19]. Similarly, mobile phones are the object of theft in four of every ten personal robberies in sev-

"Towards Speech-Generated Cryptographic Keys on Resource-Constrained Devices," F. Monrose, M. Reiter, Q. Li, D. Lopresti, and C. Shih, *Proceedings of the Eleventh USENIX Security Symposium*, August 2002, San Francisco, CA, pp. 283-296.

What is a Biometric?

- A *biometric* is a measure of a user's "unique" biological and/or physiological traits:

  E.g., iris, fingerprint, face.

- More specifically, a *behavioral biometric* measures how a user performs a given action:

  E.g., voice, handwriting, typing patterns, gait.

- We are studying security of behavioral biometrics.
- Applications to authentication and key-generation.

Typical Approach to Evaluation

Propose new biometric (or features or classifier), then:

- Assemble 10 (or 50 or 100) students in a room and collect appropriate measurements from them (or use existing database gathered for such purposes).
- Perhaps (but too rarely) let test subjects see inputs they are supposed to be forging.
- Examine FRR vs. FAR (false reject rate vs. false accept rate) curves and draw conclusions.

The Real World

The real world teaches us to be more paranoid:

- Some users better than others at creating forgeries.
- Adversaries will dedicate much time and effort to defeating your system ...
- ... and may even try to exploit advances in algorithms and computer hardware.

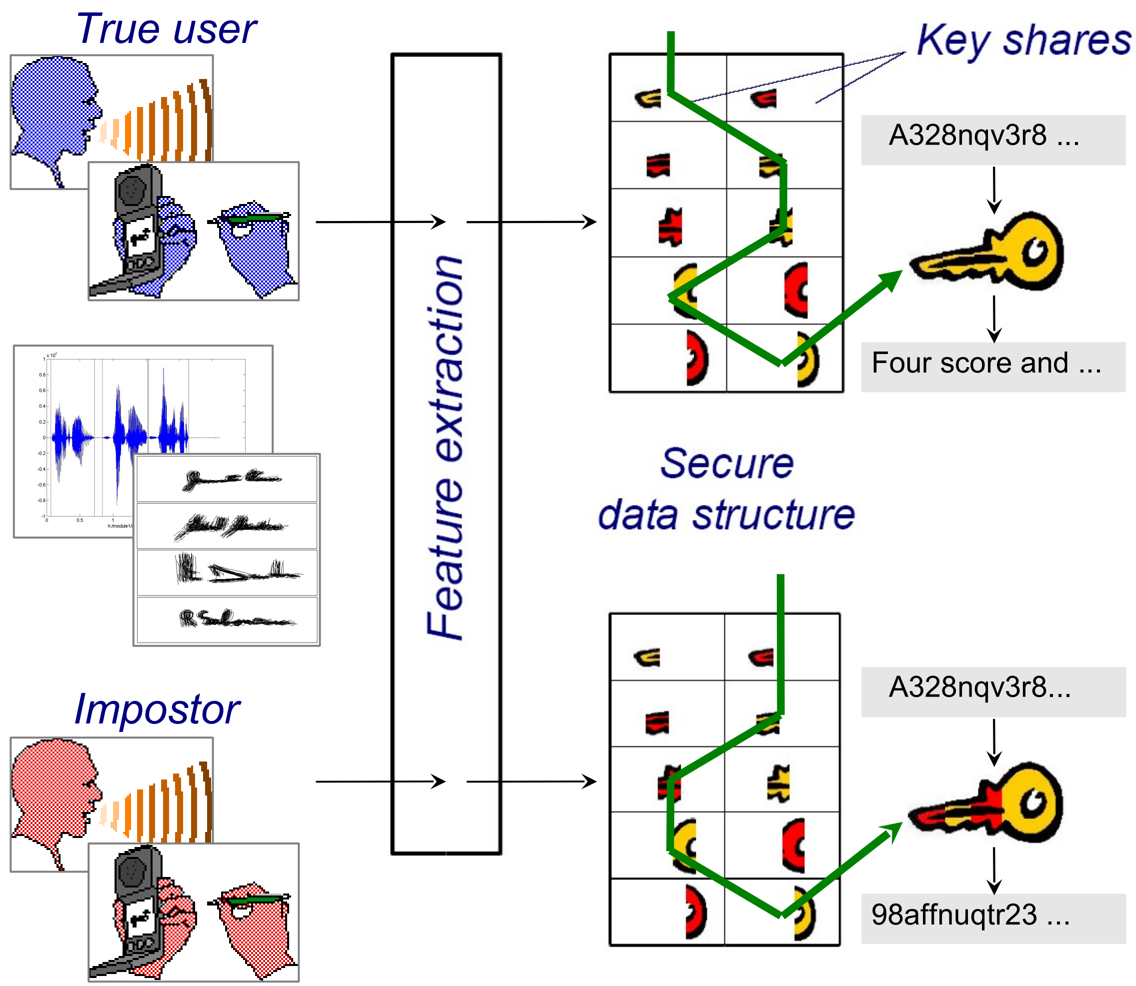
Wolf in sheep's clothing
(user who seems innocent, but who is determined to break system and has talent and resources to do so)

Authentication

- Task is to prove you are who you say you are.
- Passwords commonly used, but have low entropy (are easily guessed, as past research has shown).
- Biometrics are assumed to have high entropy and to be strong indicators of identity.
- Even better: combine biometrics with passwords (password hardening).

Key Generation via Biometrics

- Cryptographic key broken into shares and mixed with random data.
- Features extracted from user's speech or handwriting.
- Only input from true user will select correct shares to yield proper key.



Example Systems

- Cryptographic keys from voice [MRLW01, MRLS02].
- Private DSA keys (handwriting) [HC02].
- “Biometric hash” (handwriting) [VS04].
- Cryptographic keys from face [GN03, CZC04].
- Cryptographic keys from dynamic handwriting [KGNT05].
- Cryptography and biometrics (iris) [HAD06].
- Lots of work on “fuzzy extractors” (10+ papers).

Handwriting as a Biometric

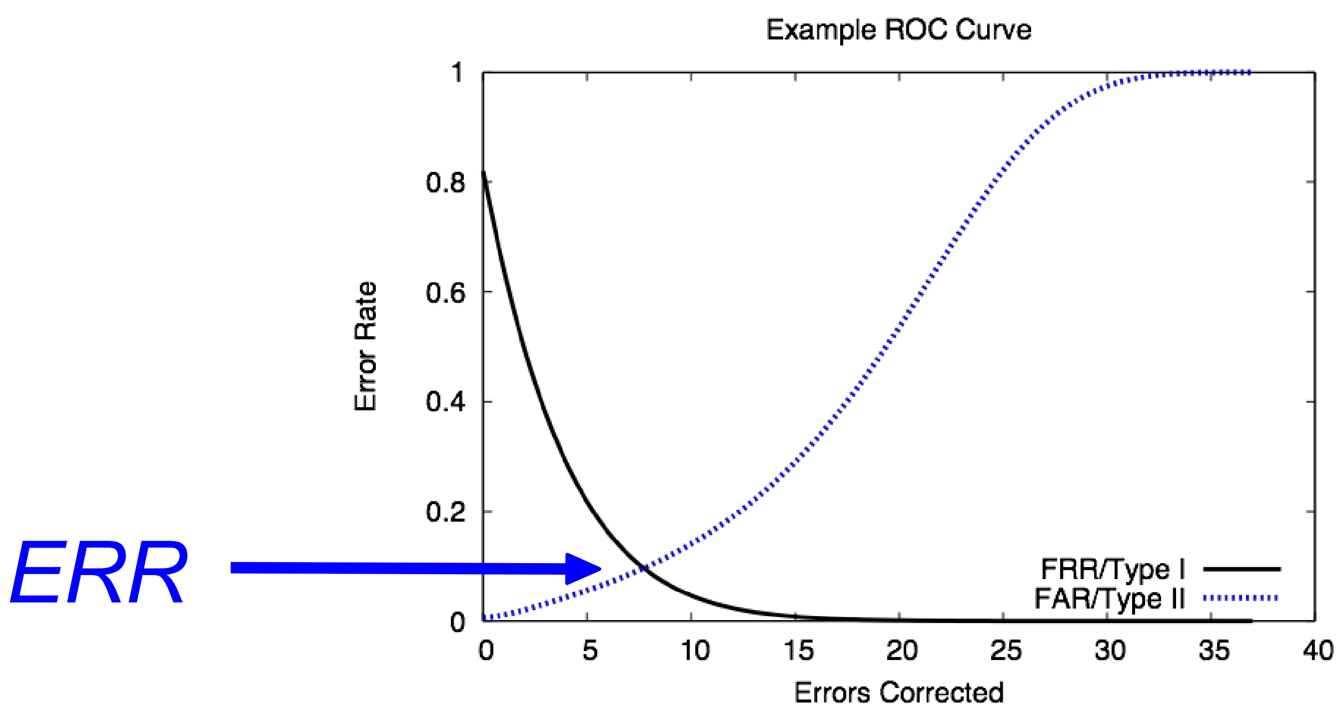
- Signatures have some well-known advantages:
  - » natural and familiar way of confirming identity,
  - » long-standing (legal) acceptance as identifiers,
  - » capture is less invasive than other biometrics.
- Not necessarily best choice for key generation or authentication, though.
- Our work focuses on writing of *passphrases*.
- Typical features used:

  offline width, height, aspect ratio, area,

  online pen up/down time, velocity, acceleration.

Security Analysis

- Receiver Operating Characteristic (ROC) curves
  - » False Reject Rate vs. False Accept Rate
  - » I.e., Type I / Type II errors
  - » Examine Equal Error Rate (EER)



Security Analysis

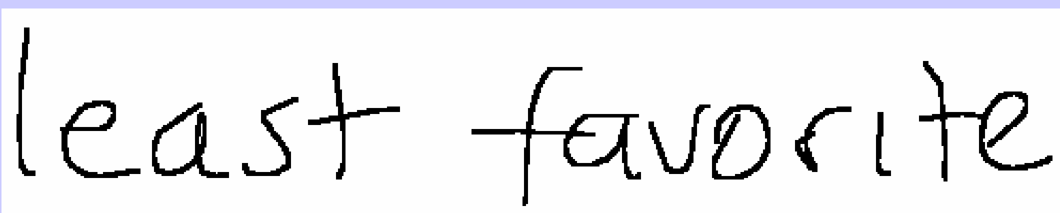
- Compute FRR by partitioning samples into two sets:
  - » use first set to make template,
  - » authenticate second set against template,
  - » repeat.
- Computing FAR is trickier. Must authenticate forgeries against template, but where to get them?
- Four criteria reflecting increasing knowledge:

Naïve → Naïve* → Static → Dynamic

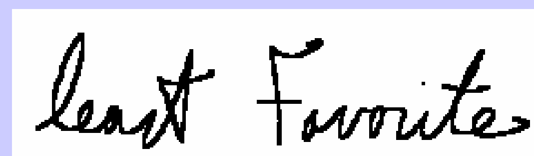
Naïve Forgeries

- Very common in the literature.
- Use other subjects' writing as it was naturally rendered to forge the target writer.

Target



Forgery

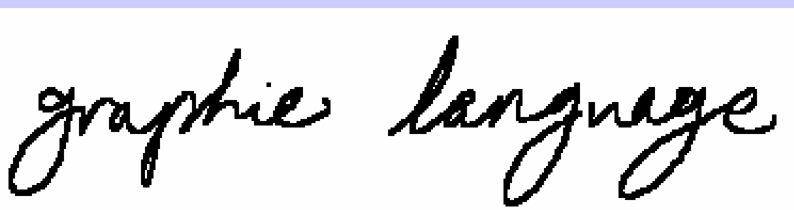


- Useful first step, but not a good test of security.

Naïve* Forgeries

- Similar to Naive, but only tests similar writing styles.
- Writing styles: Cursive, Mixed, Block.

Target



Forgery

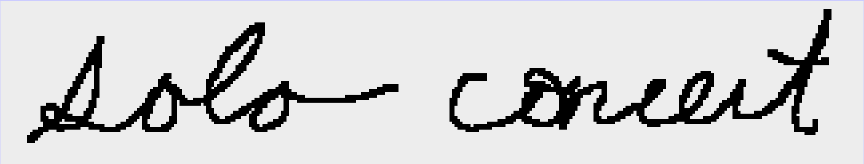
- Slightly better than simple Naive.

Static Forgeries

- Provide forgers with image of target passphrase.

Target

Forgery

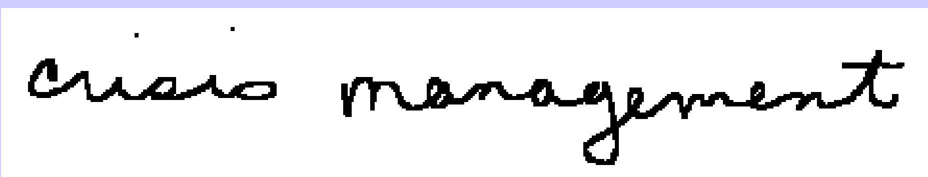


- Looks better!
- But what about temporal features?

Dynamic Forgeries

- Show users dynamic rendering of target passphrase.
- Allow multiple replays.

Target



Forgery

- For paranoid security analysis, this is what we need.

Experimental Analysis

Initial data collection:

- Study of approximately 50 users (11K+ samples).
- Each provided 10-20 renderings of 5 passphrases.
- Also wrote a parallel corpus of unrelated material.

Forgery data collection:

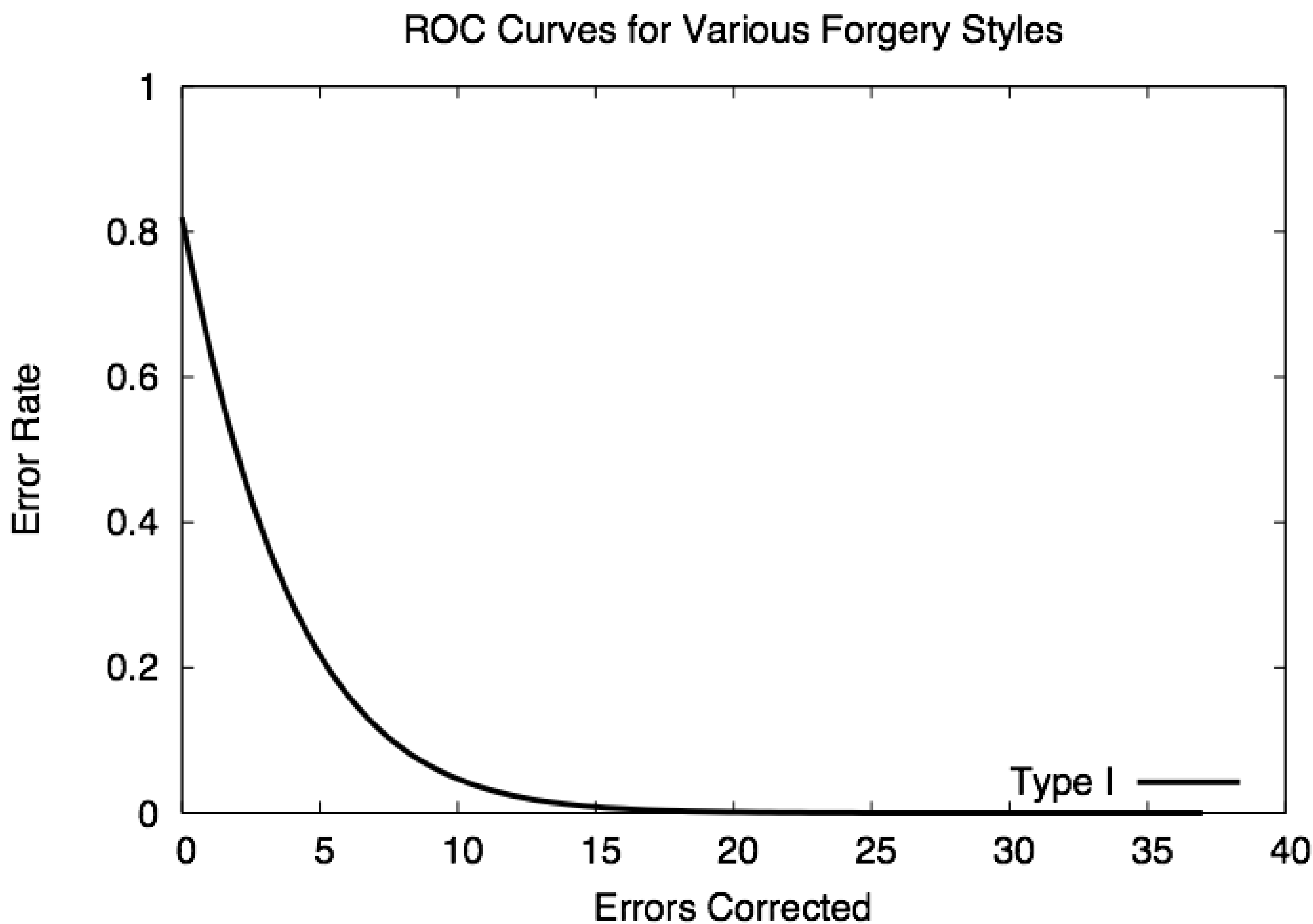
- 36 users each created 17 static, 17 dynamic forgeries.
- Forgery sessions took on average 1.5 hours.
- Evaluated quality of forgeries on a per-style basis.

Target System for Evaluation

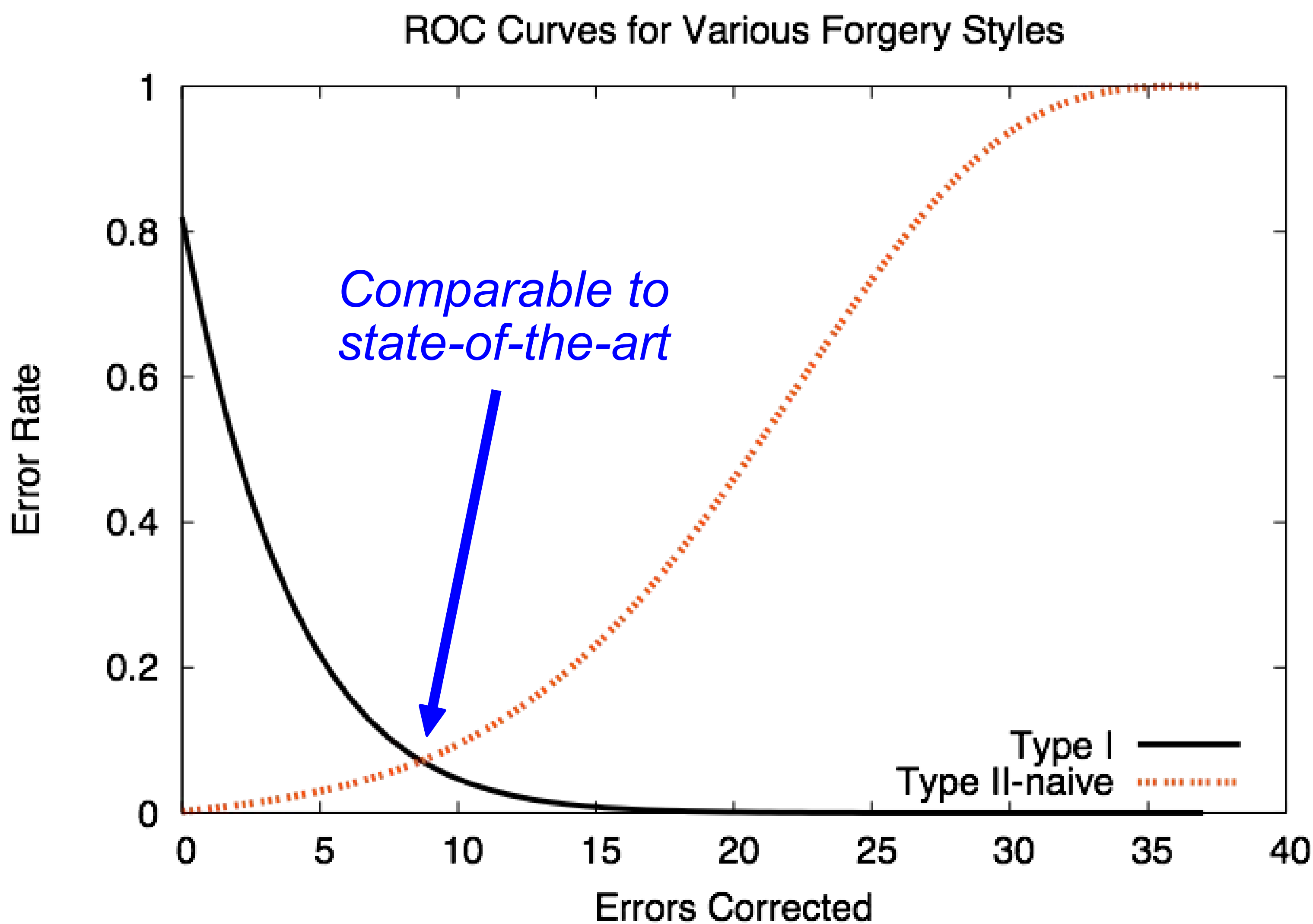
Need a real biometric system to test:

- Adapted from "Biometric Hash" of [VS04].
- Selected 36 (out of 144) best features:
  - » 13 static features,
  - » 23 dynamic features.
- "Best" = most secure in resistance to forging.
- Correlation with feature entropy unknown.

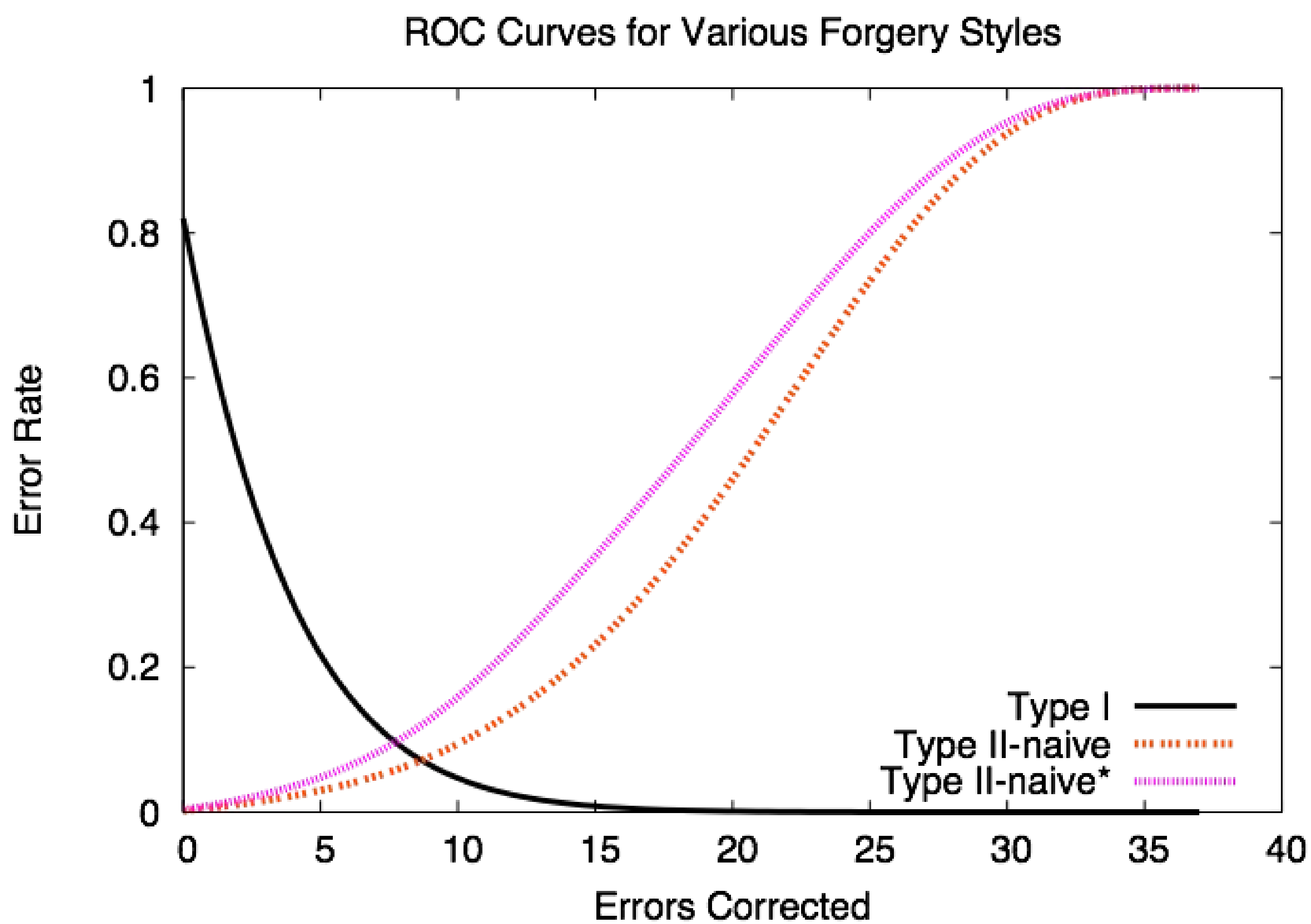
False Reject Rate



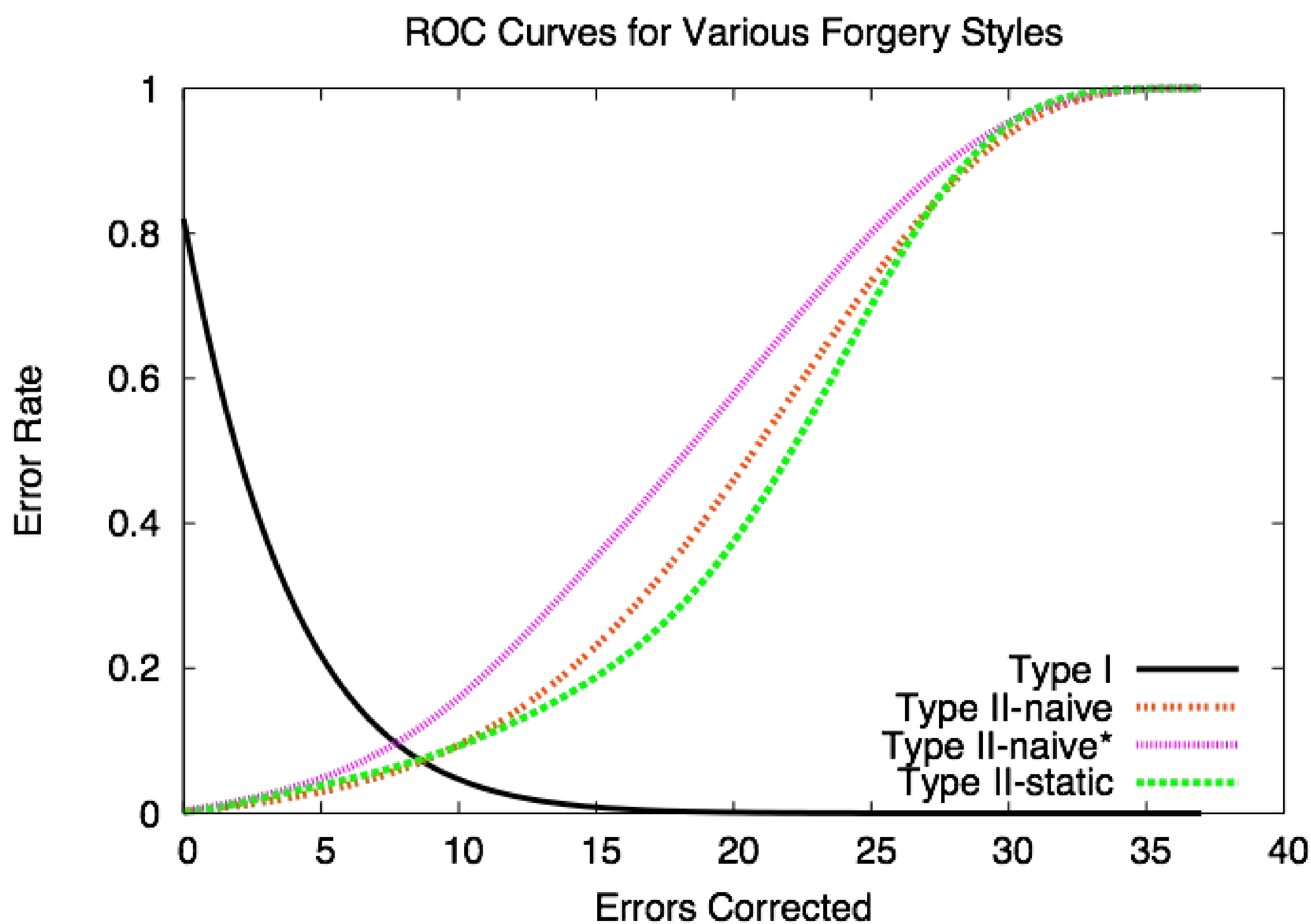
Equal Error Rate for Naïve Forgeries



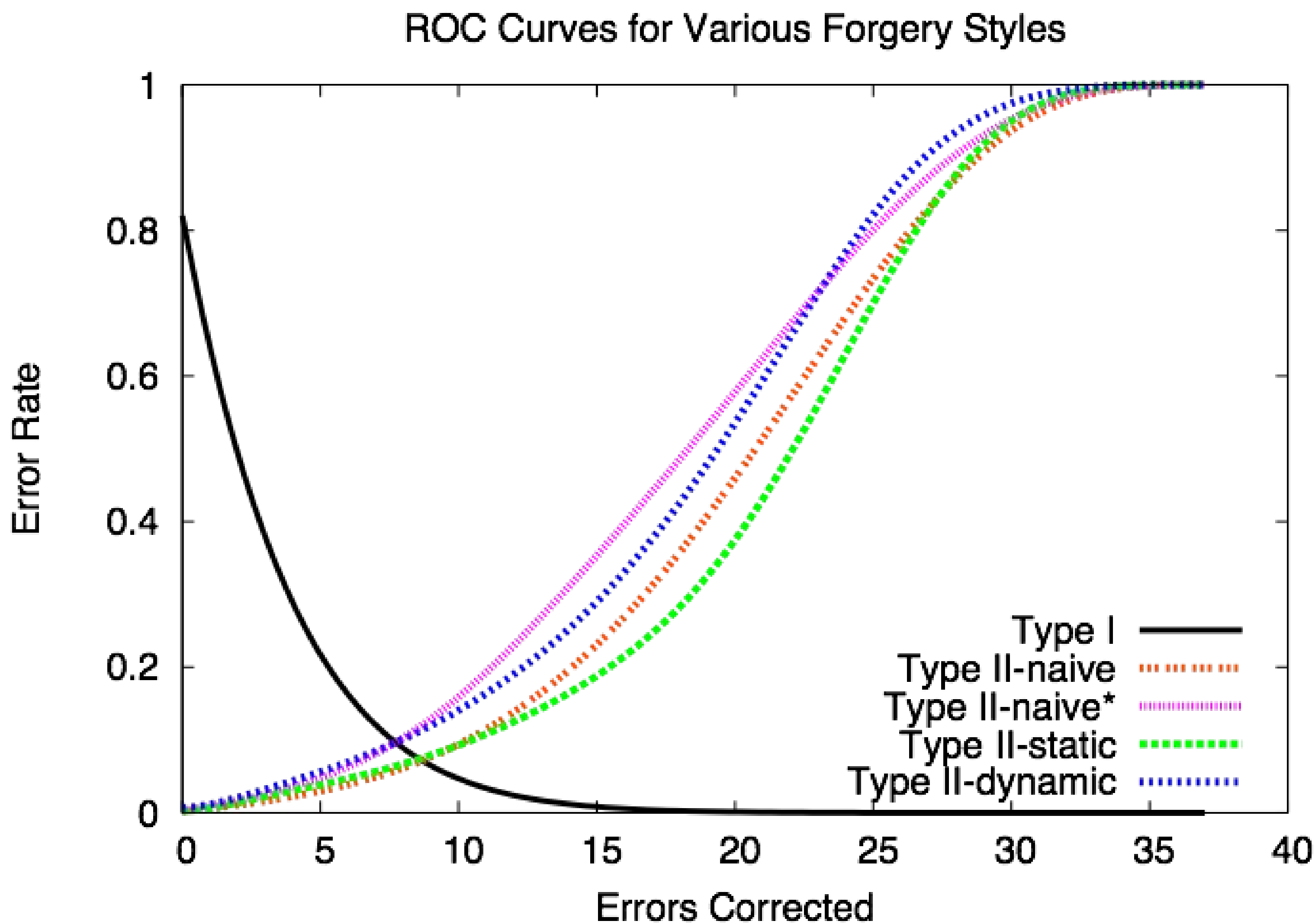
Equal Error Rates + Naïve* Forgeries



Equal Error Rates + Static Forgeries

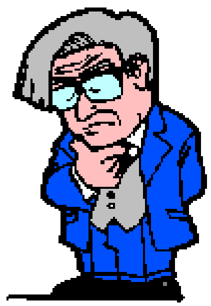


Equal Error Rate for All Forgeries

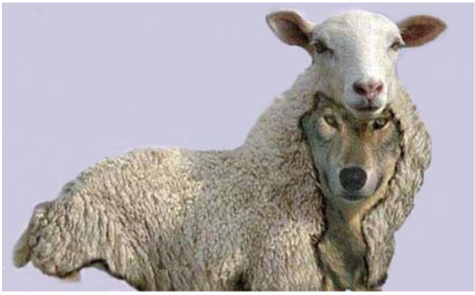


Good Measure of an Adversary?

- Are these threat models realistic?
  Naive? Static? Dynamic?
- Real adversaries are:
  » skilled,
  » knowledgeable,
  » motivated.



What happens when considering more realistic adversaries?



→ *Enter wolves in sheep's clothing*

Experimental Procedure

- Choose 9 strong forgers from Round I. Select forgers who exhibit tendency to succeed with particular writing style.
- Teach these forgers basics of how a system for generating biometric hash from handwriting works.
- Provide incentives for best-quality forgeries (gift certificates for iTunes, amazon.com, etc.).

Skill

Knowledge

Motivation

Experimental Procedure

- Choose 9 strong forgers from Round I.

o exhibit tend

icular writing

ers basics of h

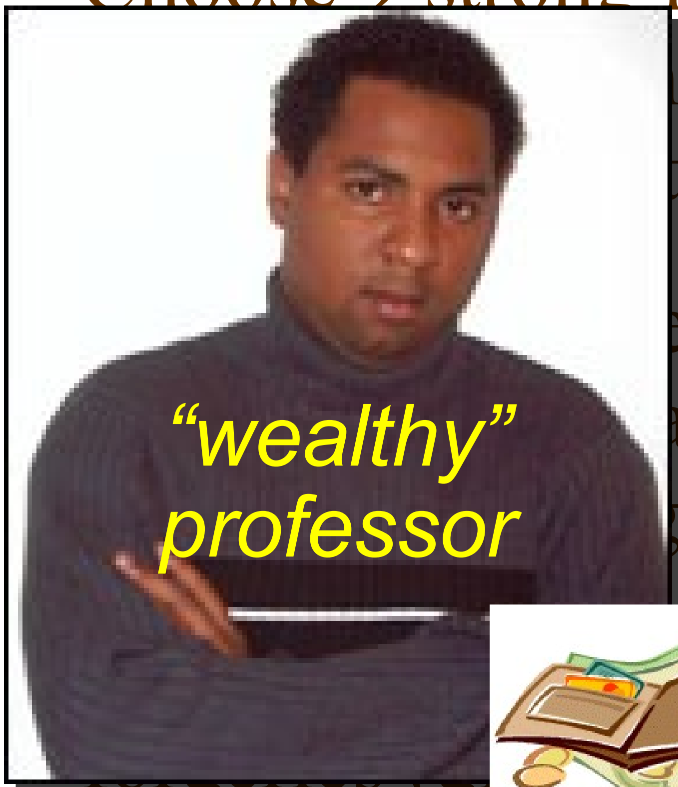
ting biometri

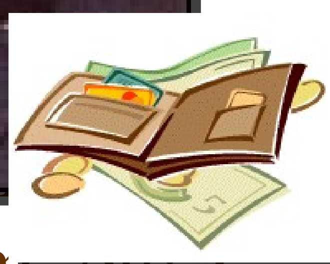
works.

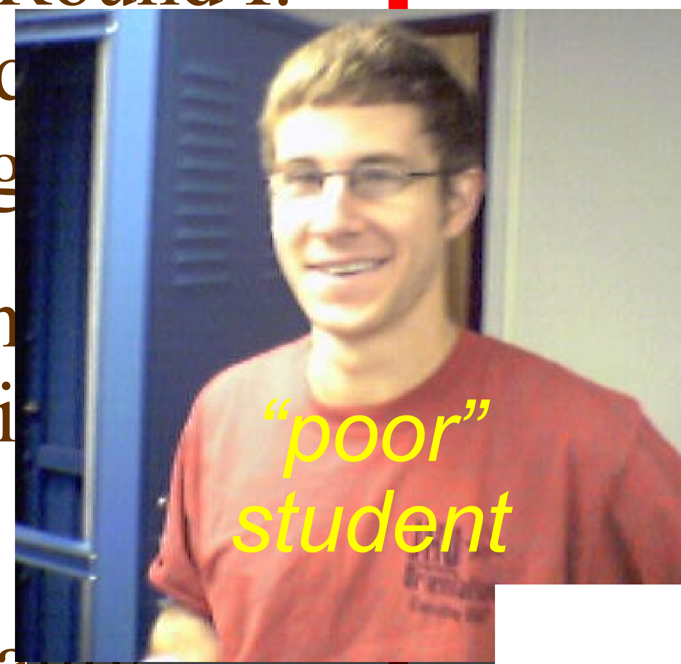
r best-quality

ates for iTunes,

amazon.com, etc.).



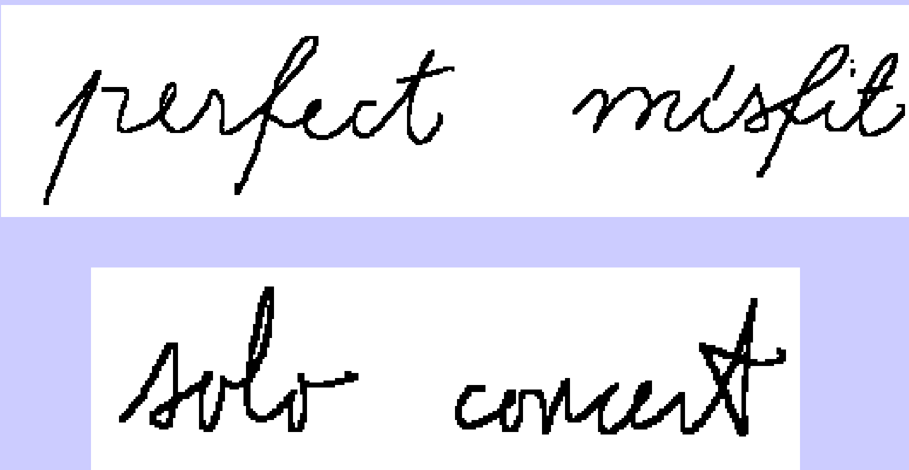
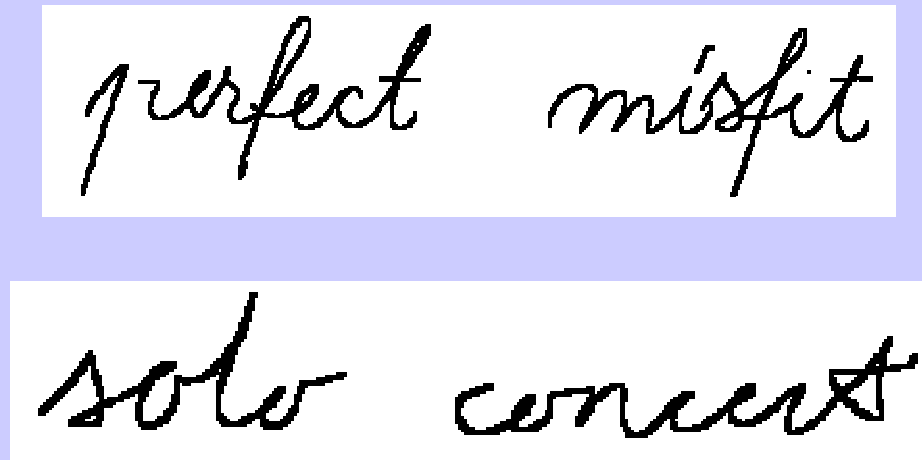




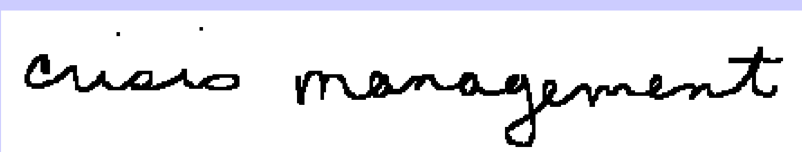
dge

M

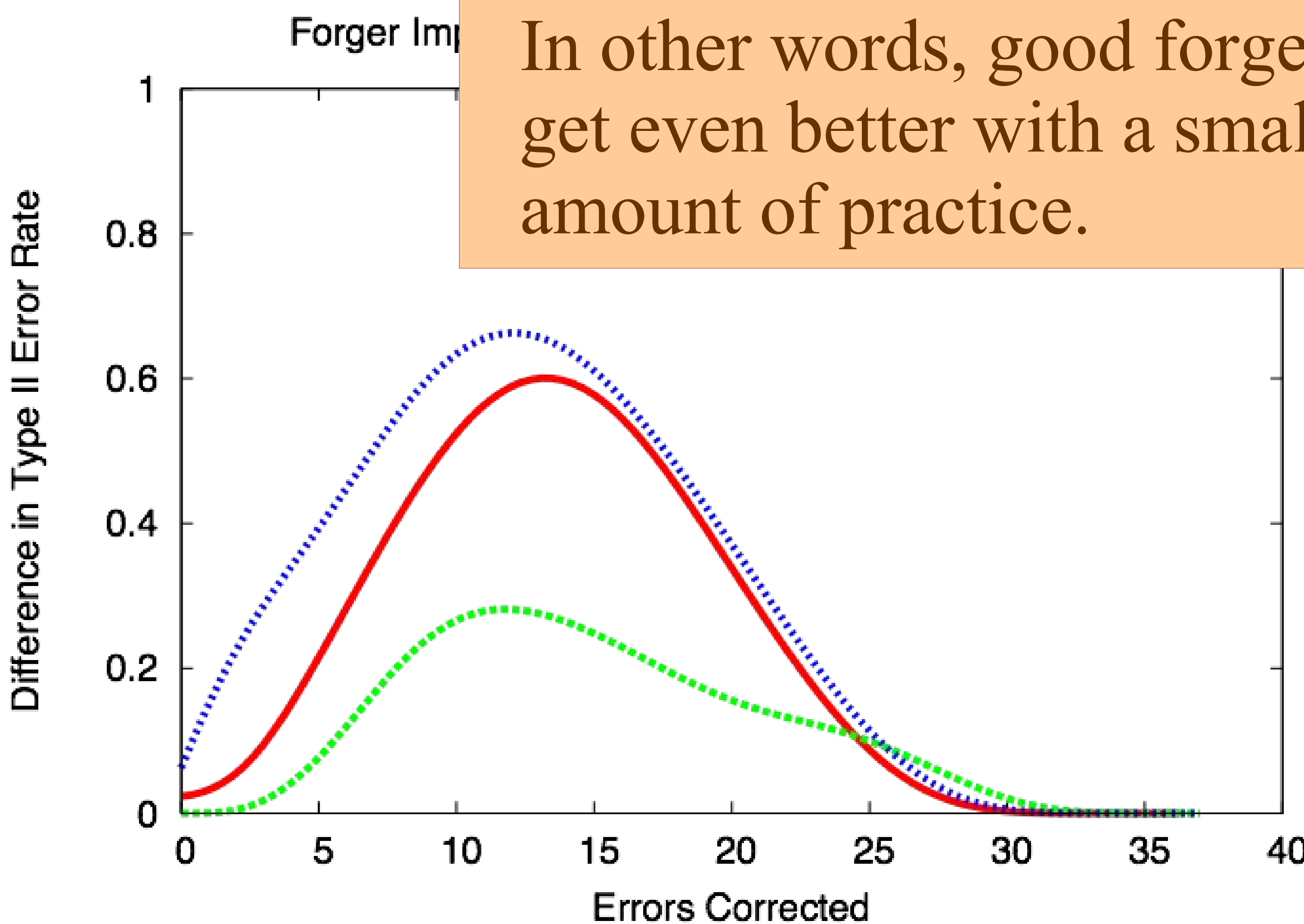
Examples of Skilled Forgeries

Targets	Forgeries
	

Comparison to unskilled case

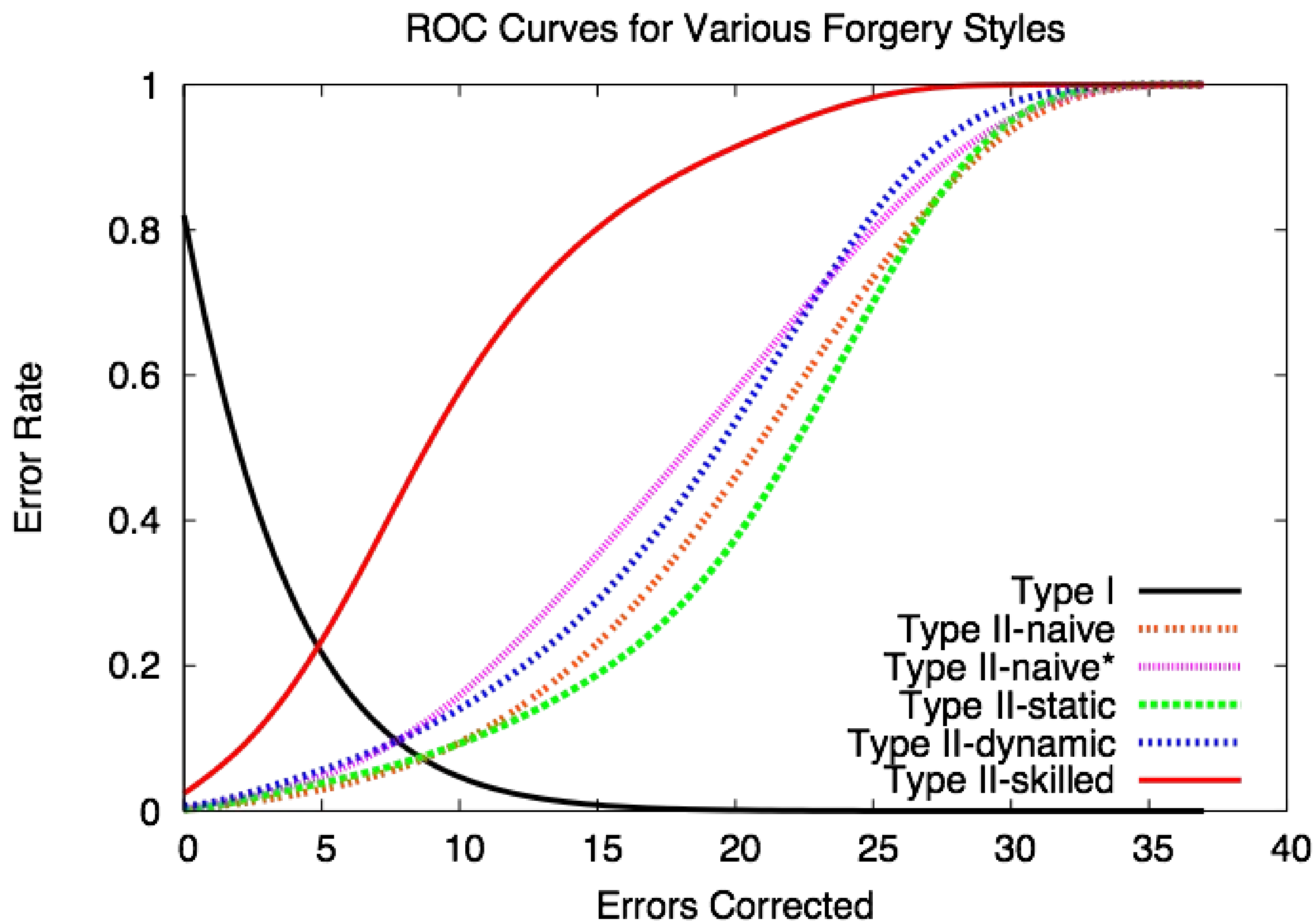


Grooming Sheep into Wolves

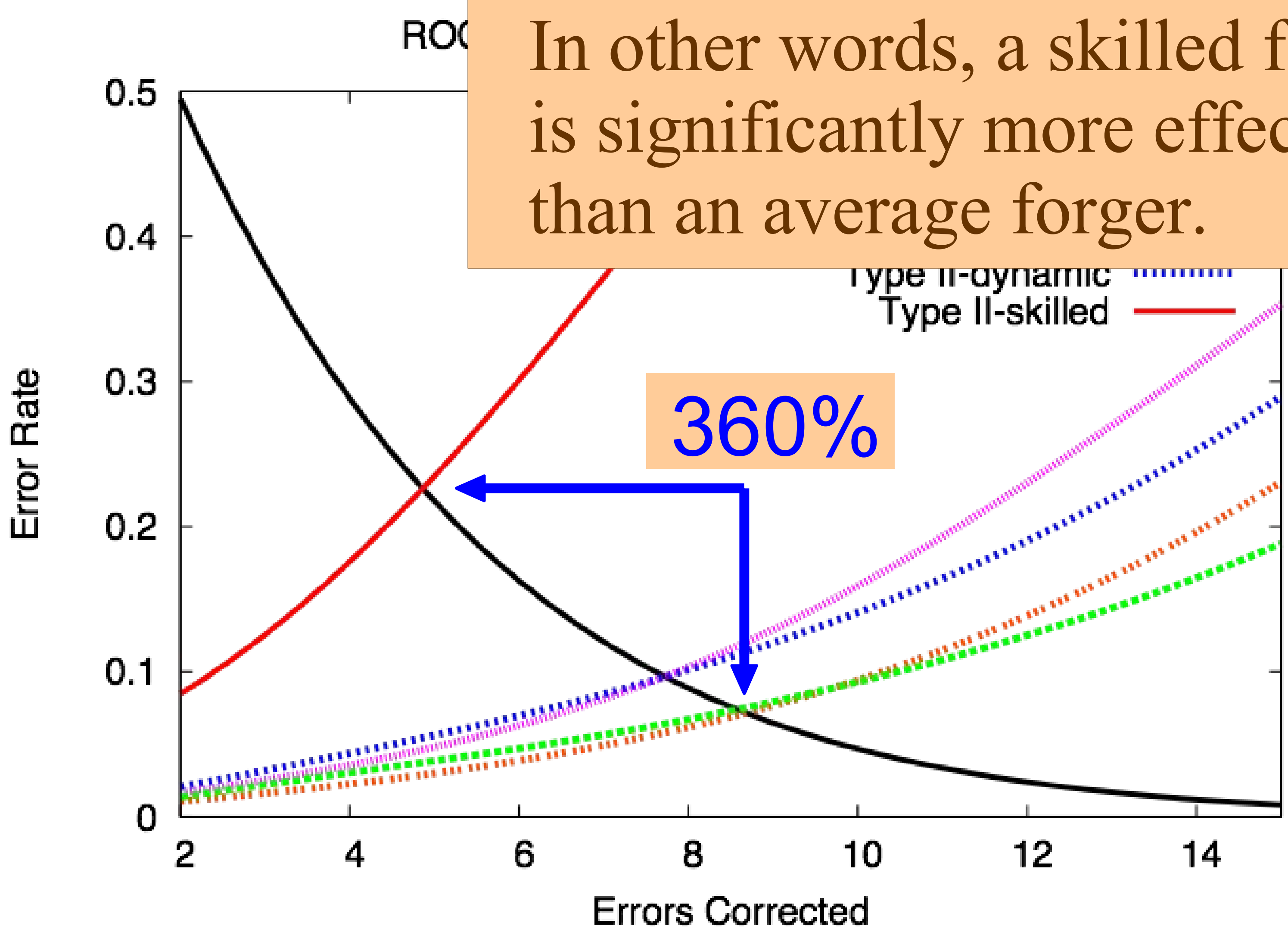


In other words, good forgers get even better with a small amount of practice.

Equal Error Rates + Skilled Forgers



Net Improvement for Skilled Forgers



In other words, a skilled forger is significantly more effective than an average forger.

Another Threat: Generative Models

- Use information gleaned about a user from various sources in attempt to synthesize his/her biometric.
- Assume adversary has access to:
  » knowledge of target user's writing style,
  » general population statistics for that style,
  » samples of user's handwriting from other contexts.
- Combine this information to create a good forgery.

A Semi-Automated Adversary

- Input:
  - » general population statistics (corpus),
  - » static samples from target user.
- Key step: infer velocity from static samples.
- Output: guess of target user's biometric.

Concatenative Handwriting Synthesis

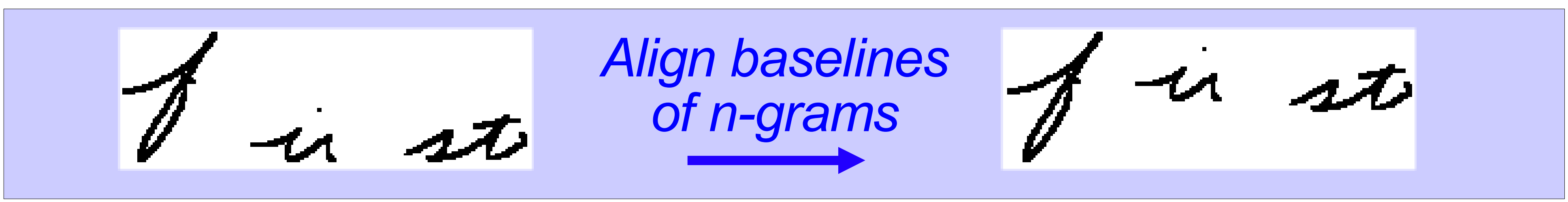
- Create velocity profiles using population statistics.
- Obtain static samples from target user.
- Trace samples onto tablet to:
  - » obtain electronic representation,
  - » guess stroke order/direction.
- Infer velocity using statistical models.
- Use concatenative synthesis to create forgeries.

Synthesis Algorithm

- Select n-grams from writing from different context such that:

$$g_1 \,||\, g_2 \,||\, g_3 \,||\, \ldots \,||\, g_k = \textit{passphrase}$$

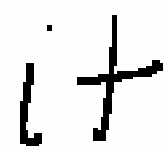
- Motivated by concatenative technique for text-to-speech synthesis (recall *Mission Impossible 3*).
- Shift the signals for each n-gram to generate a meaningful representation:

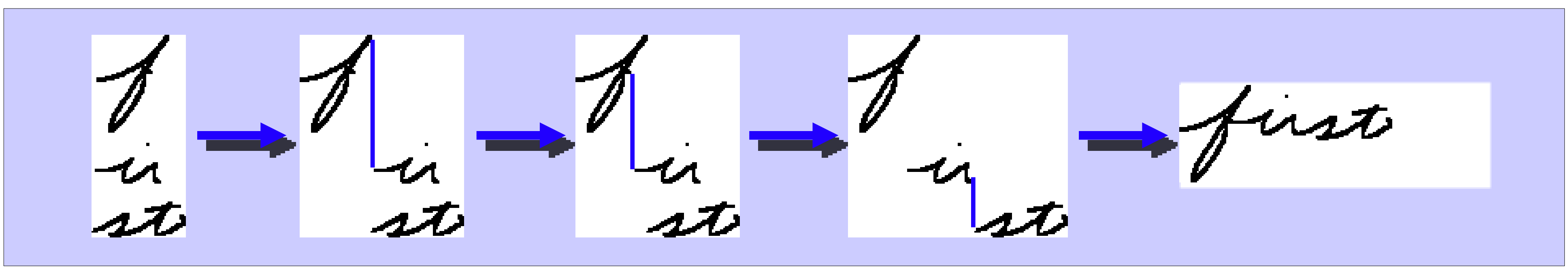


Connectivity via Population Statistics

- Connection statistics: $P_c(i, j, c_1, c_2)$
- Probability that stroke i of c_1 is connected to c_2, given that c_1 is rendered with j strokes.
- *E.g.*, $P_c(1, 2, \mathrm{i}, \mathrm{s}) \approx 1$ for cursive writers

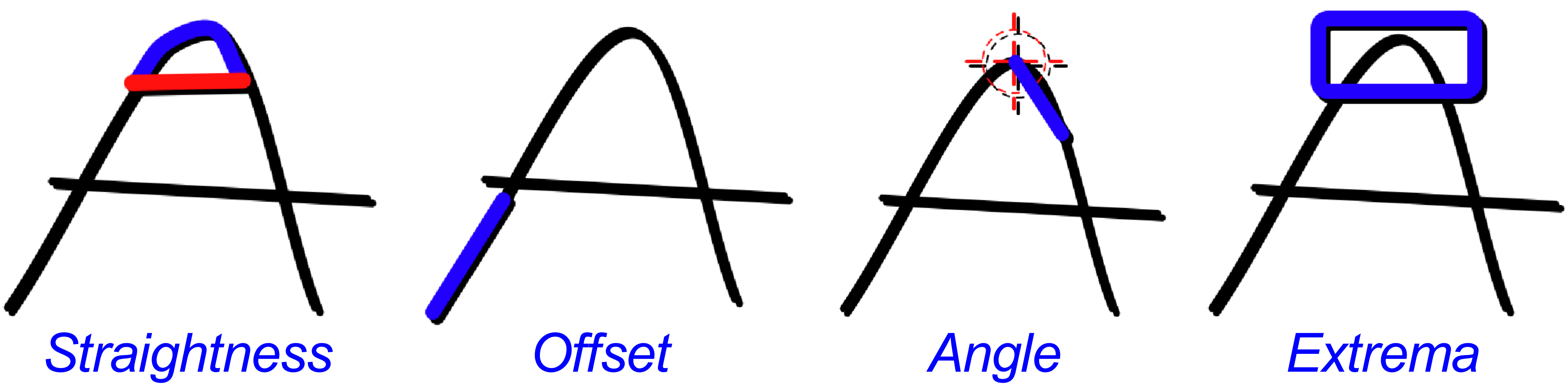
  $P_c(1, 2, \mathrm{i}, \mathrm{t}) \approx 0$ for block writers



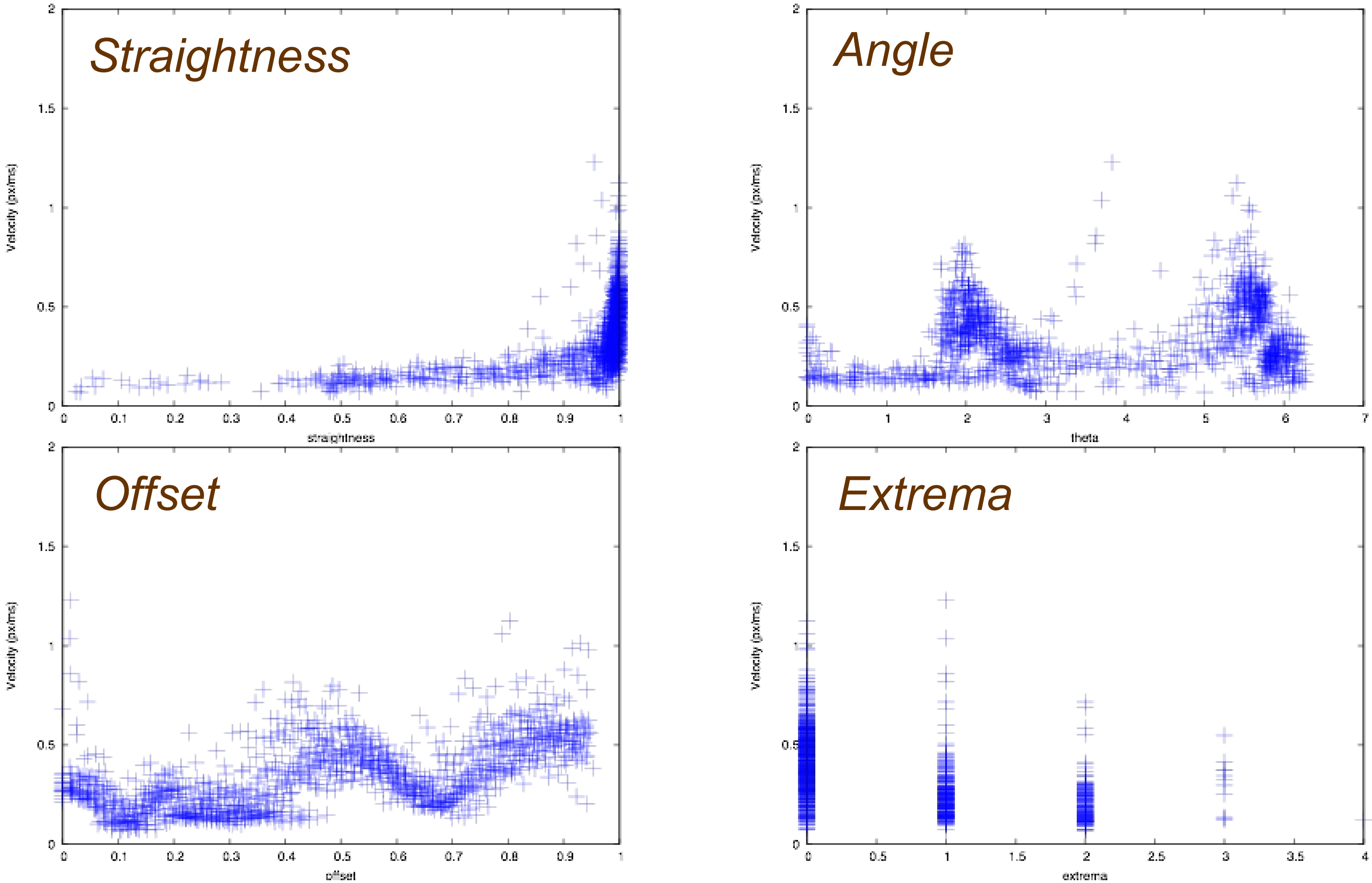


Velocity Statistics

- Group statistics on a per-stroke basis. E.g., "A" corresponds to two groups.
- Need "sufficient statistics" indicative of pen velocity.
- CANNOT be a function of distance between points.
- Examined 9 measures, selected 4 most-representative.



Population Velocity Statistics

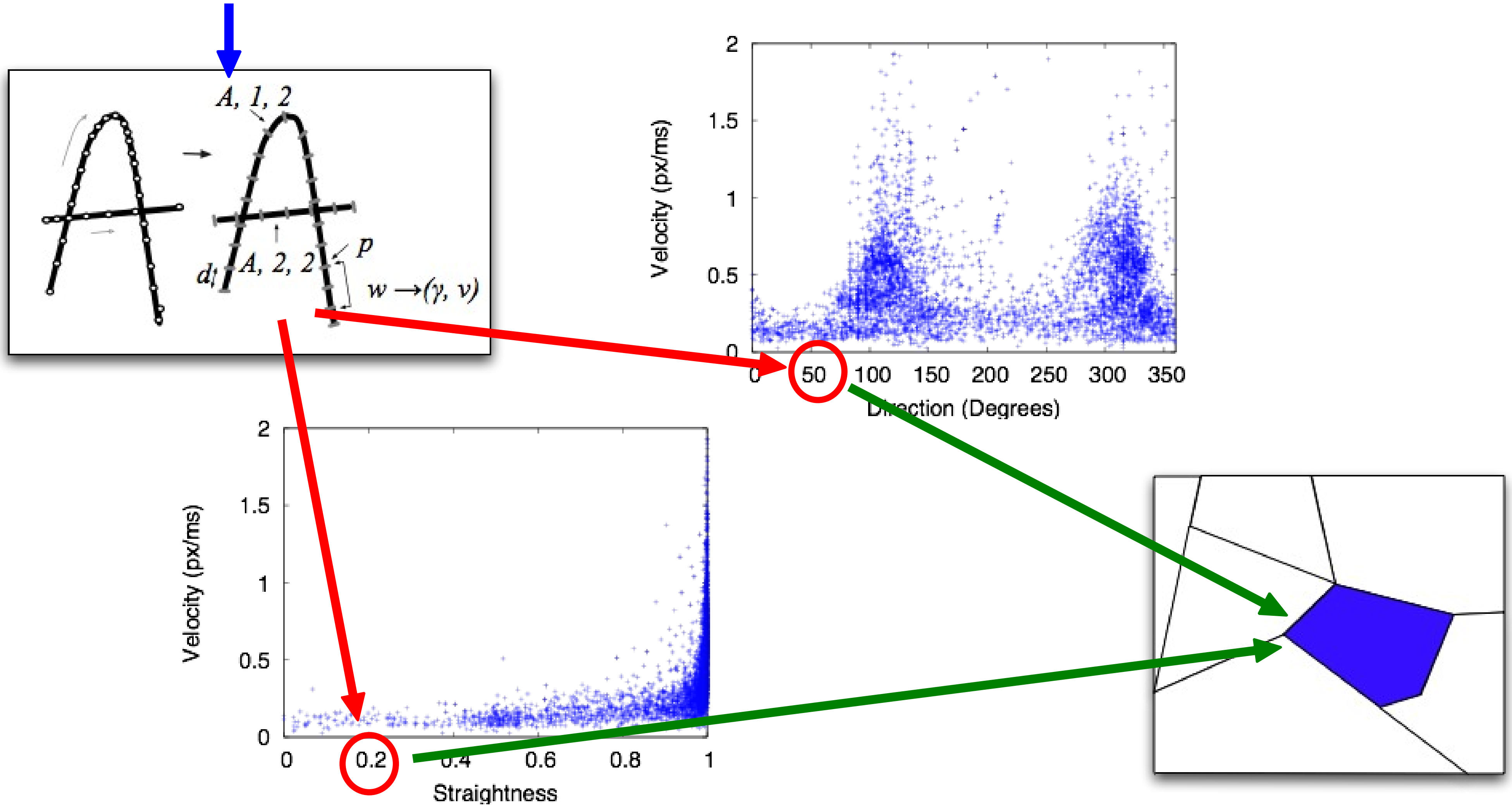


Velocity Profiles

- Take writers from similar style as target user.
- Compute statistics across each stroke.
- Assign a vector, velocity pair $\langle \gamma, \nu \rangle$ to each window.
- Partition vector space using *k*-means.
- Assign representative velocity to each partition.

Grouping Similar Windows

"GRAPHIC LANGUAGE"



Guessing the Biometric

- Trace sample by hand, then re-sample automatically:
  - » provides stroke order and direction,
  - » *x,y* positions.
- Infer velocities:
  - » For window ω_1, compute $\omega_1 \rightarrow \langle\gamma,?\rangle$.
  - » Use *k*-nearest neighbors to find closest partitioning and assign velocity at centroid.

Guessing the Biometric

- Combine samples to create a forgery:

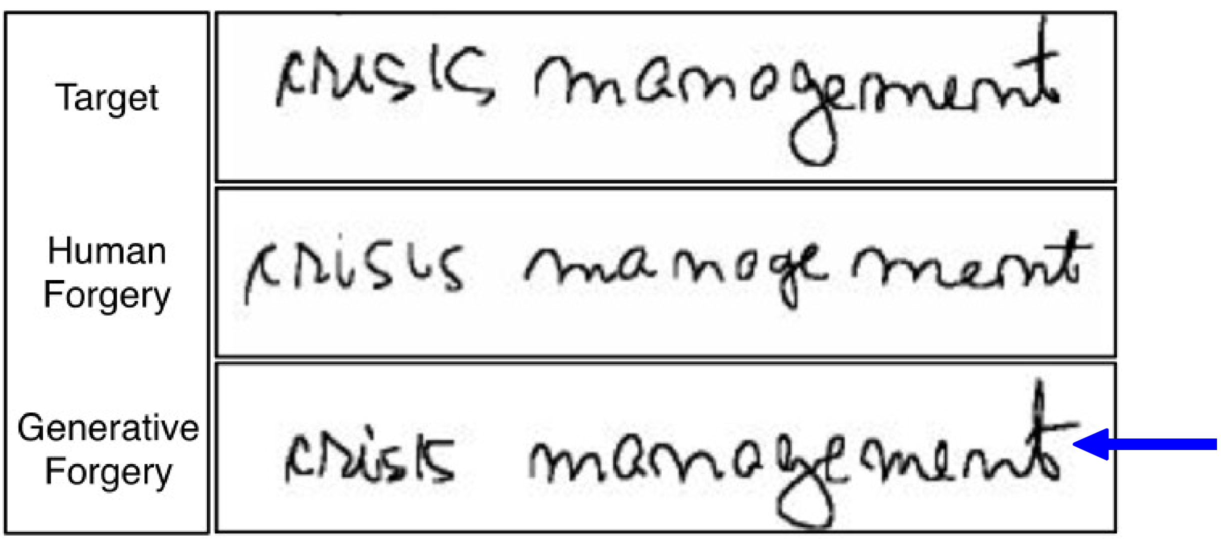
  se + *cre* + *t* = *secret*

- Use population statistics to estimate:

  spacing, inter-sample stroke ordering / stroke connections, pen-up time, velocities.

Experimental Procedure

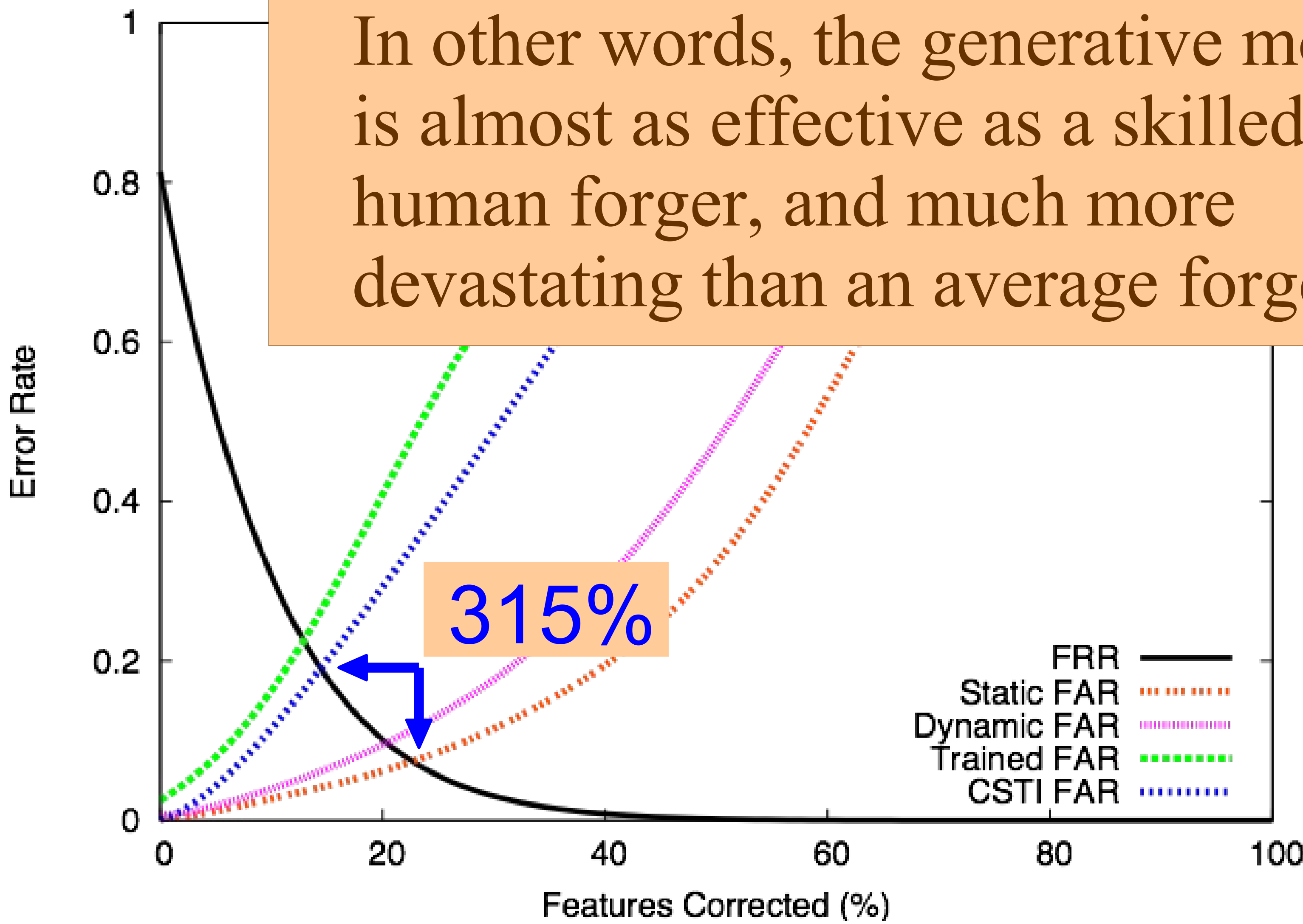
- Employ concatenative synthesis to forge passphrases.
- On average:
  - » each n-gram was less than 2 characters long,
  - » used < 7 writing samples to generate each forgery.



Population statistics good, but not perfect

Generative Attack vs. Skilled Forgers

In other words, the generative model is almost as effective as a skilled human forger, and much more devastating than an average forger.



Summary

- Current evaluation methodologies over-estimate biometric security in certain cases. Must consider:
  - » skilled adversaries,
  - » automated attacks.
- Trained students are decent forgers. (Watch out!)
- Careful evaluation is time-consuming.

Extensions

- Generative forgeries with access to less information (e.g., pieces of paper stolen from trash).
- Using human-traced samples to infer stroke direction.
- Adapting these techniques to test other proposed schemes for key-generation.
- Study human ability to distinguish forgeries (early results suggest we fall short of machines).
- Develop more rigorous evaluation paradigms.

The End

Thank you! Questions?

References

- BML06 - L. Ballard, F. Monrose, D. Lopresti. “Biometric Authentication Revisited: Understanding the Impact of Wolves in Sheep’s Clothing.” *Proceedings of the 15th Annual Usenix Security Symposium*. 2006.
- CZC04 - Y.J. Chang, W. Zhung, T. Chen, “Biometrics-Based Cryptographic Key Generation.” *Proceedings of the International Conference on Multimedia and Exposition*. 2004.
- GN03 - A. Goh, D.C.L. Ngo, “Computation of Cryptographic Keys from Face Biometrics.” *Proceedings of Communications and Multimedia Security*. 2003.
- HAD06 - F. Hao, R. Anderson, J. Daugman, “Combining Crypto with Biometrics Effectively.” To appear. 2006.
- HC02 - F. Hao, C. Wah, “Private Key Generation from On-Line Handwritten Signatures.” *Information Management & Computer Security*. 2002.
- KGNT05 - Y.P. Kuan, A. Goh, D. Ngo, A. Teoh. “Cryptographic Keys from Dynamic Hand-signatures with Biometric Security Preservation and Replaceability.” *Proceedings of the Fourth IEEE Workshop on Automatic Identification Advanced Technologies*. 2005.
- MRLW01 - F. Monrose, M. Reiter, Q. Li, S. Wetzel, “Cryptographic Key Generation from Voice.” *Proceedings of the IEEE Conference on Security and Privacy*. 2001
- VS04 - C. Vielhauer, R. Steinmetz, “Handwriting: Feature Correlation Analysis for Biometric Hashes.” *EURASIP Journal on Applied Signal Processing*. 2004