Attacks on Online Handwriting Biometrics

Daniel Lopresti [1]
Dishant Patel [1]

[1] Computer Science & Engineering
Lehigh University
Bethlehem, PA 18015, USA

Fabian Monrose [2]
Lucas Ballard [2]

[1] Computer Science
Johns Hopkins University
Baltimore, MD 21218, USA

lopresti@cse.lehigh.edu

Motivation

Data becoming more portable (PDA's, cell phones, laptops, etc.) – theft is a growing concern.

Why aren't passwords enough?

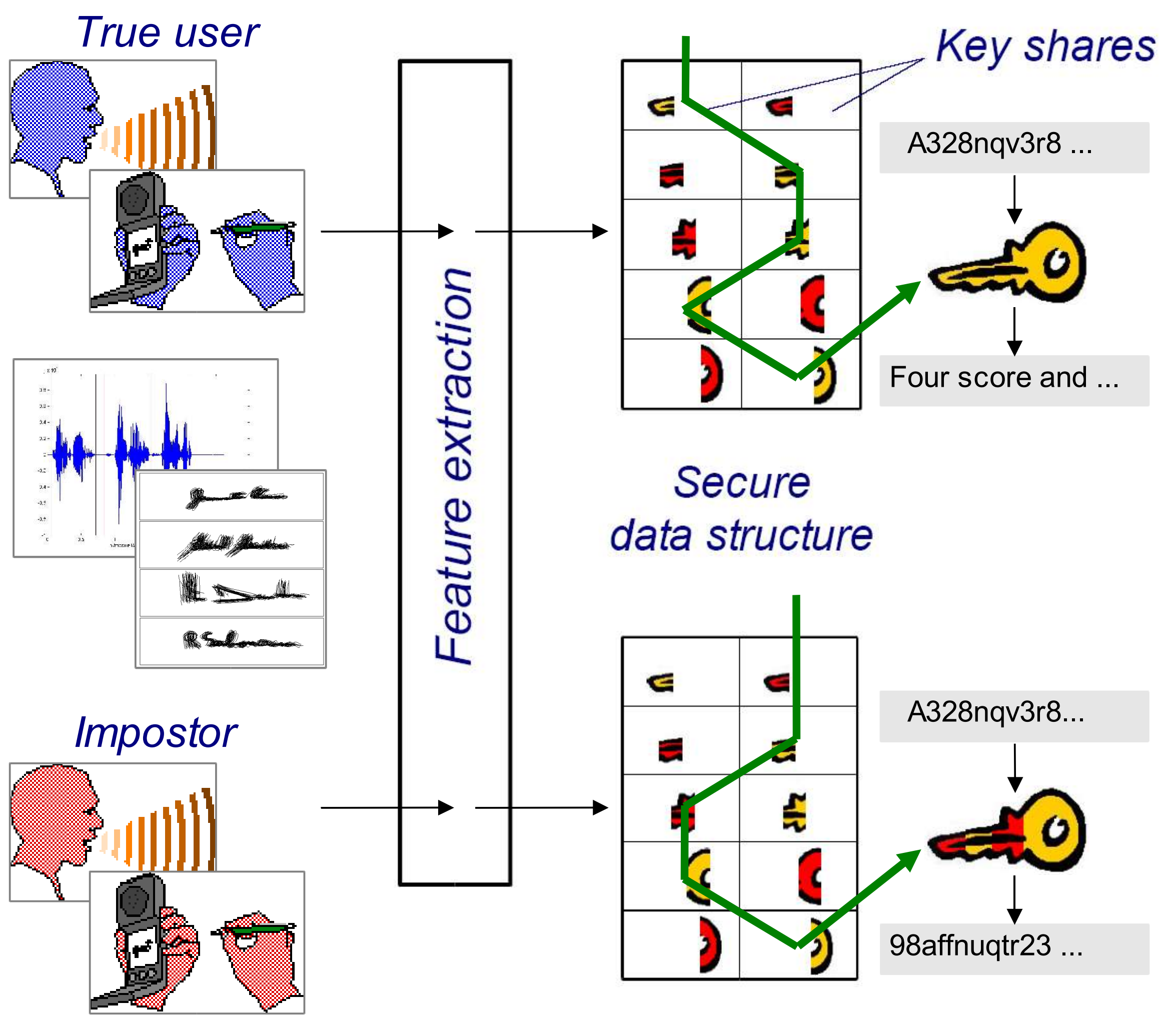
- Very easy to "crack."
- Thief can disassemble and reverse-engineer device.

Two-pronged solution:

- Biometrics in place of (or in addition to) passwords.
- Secure data structure to encrypt information.

Using Biometrics to Protect Data

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

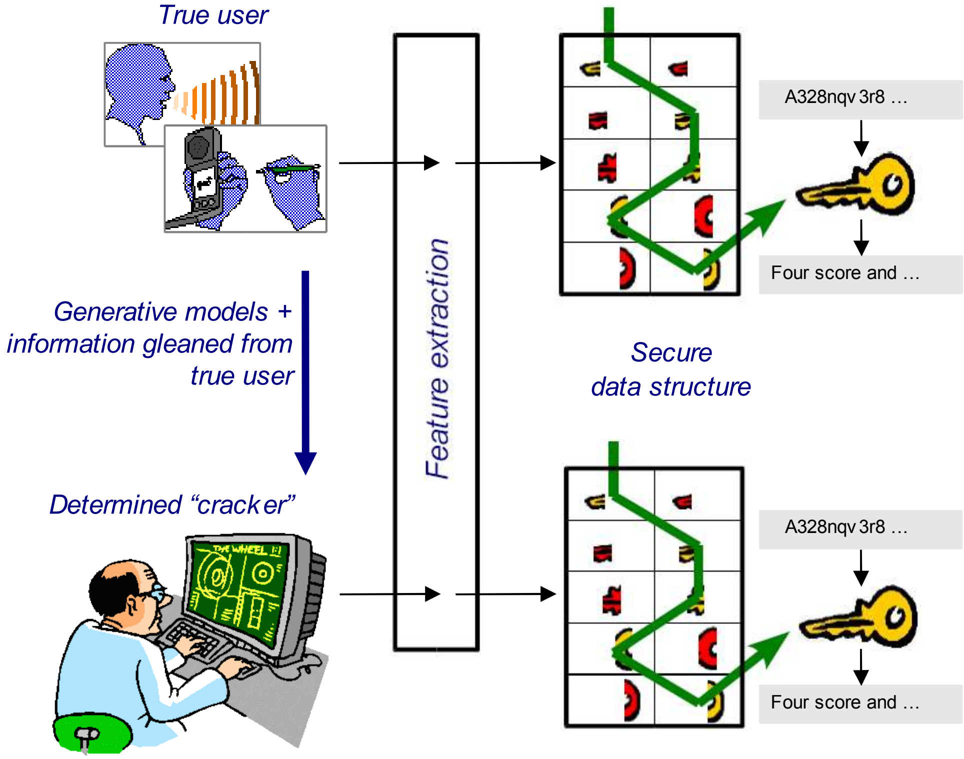


"Towards Speech-Generated Cryptographic Keys on Resource-Constrained Devices," F. Monrose, M. Reiter, Q. Li, D. Lopresti, and C. Shih, *Proceedings of the Eleventh USENIX Security Symposium*, August 2002, San Francisco, CA, pp. 283-296.

Using Biometrics to Protect Data

Biometrics may be vulnerable:

- Generative models can mimic human behavior.
- If successful, some systems breakable.



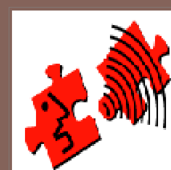
Our work:

- Identify potential attacks.
- Analyze risk.

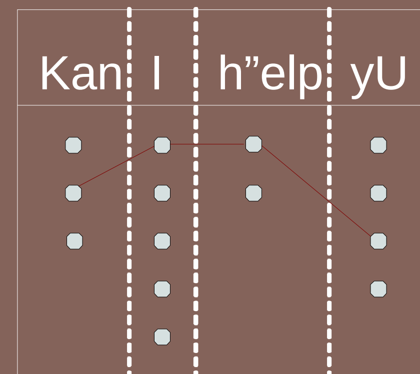
Use our experience to improve biometric security.

Concatenative Attack on Speech

Speech

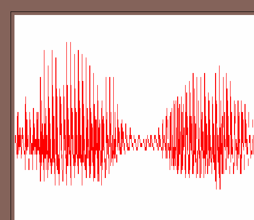
Can I help you →  → Kan I h"elp yU →

search



kan
"elp
yU

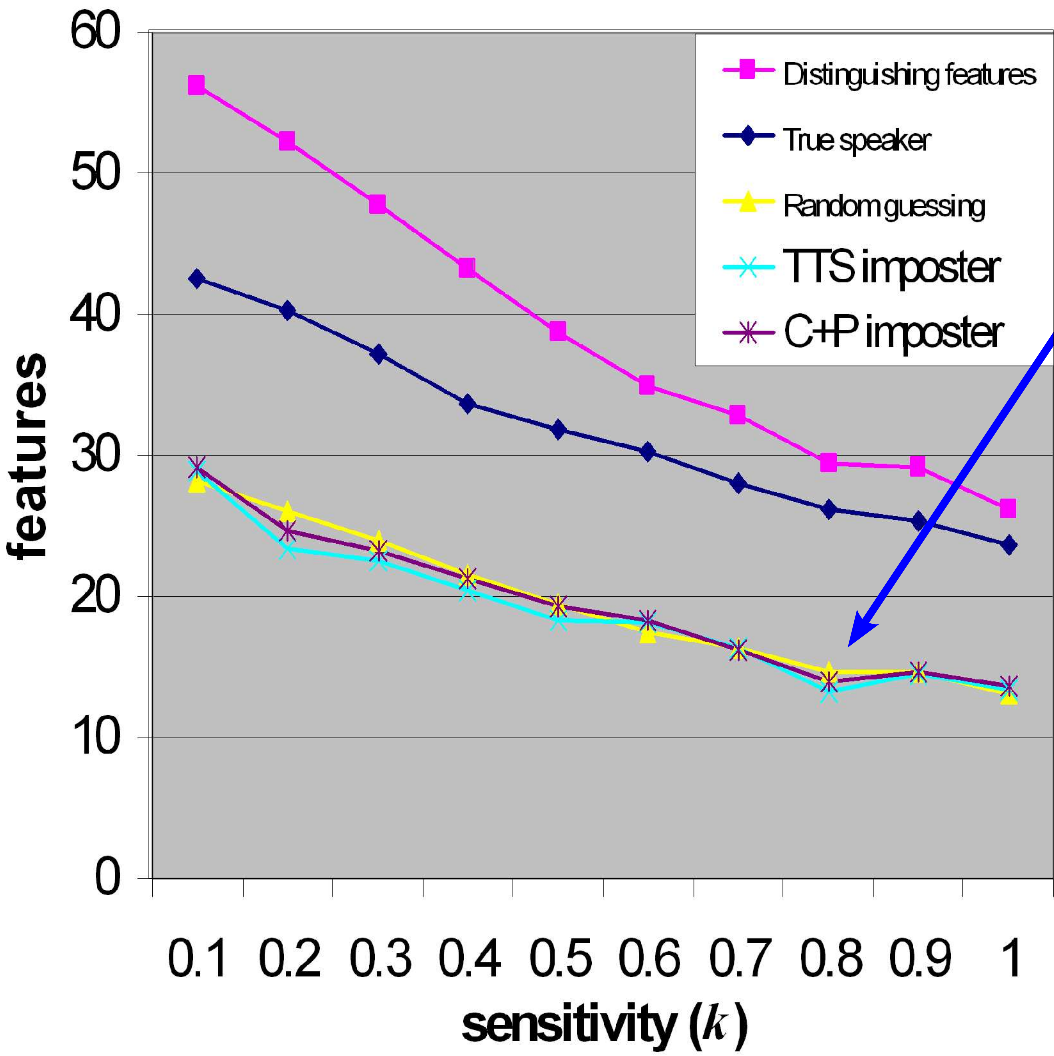
corpus



forgery

Presumes attacker has access to corpus of prerecorded speech. (Hack voice mail, record target with hidden mike, etc.)

Results of Text-to-Speech Attacks*



TTS is no better than random guessing. Why?

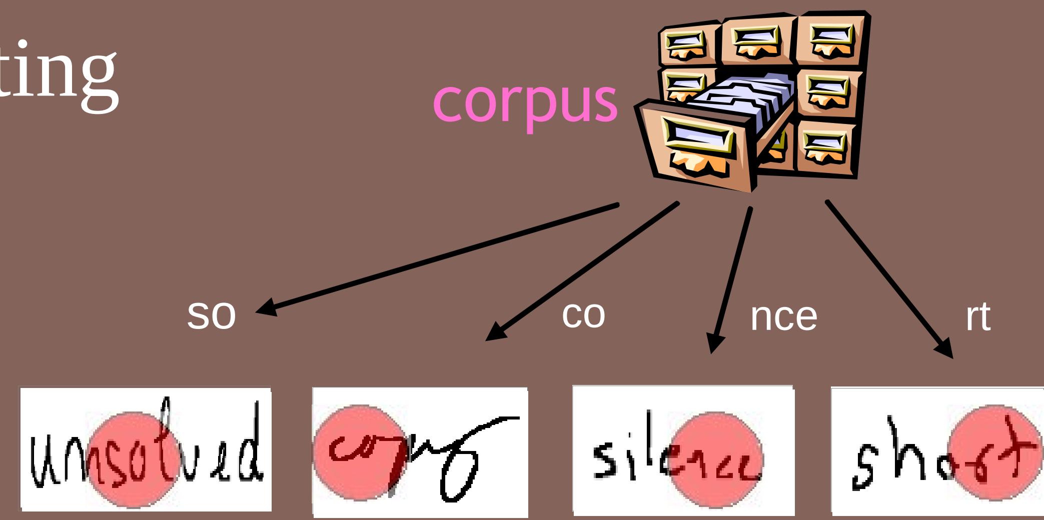
- Speech synthesis too immature at this point.
- We just didn't have enough data.

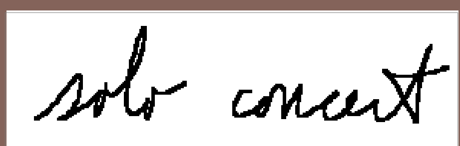
Either way, we expect attacks to become more worrisome over time.

* "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.

Concatenative Attack on Handwriting

Situation with handwriting is analogous.

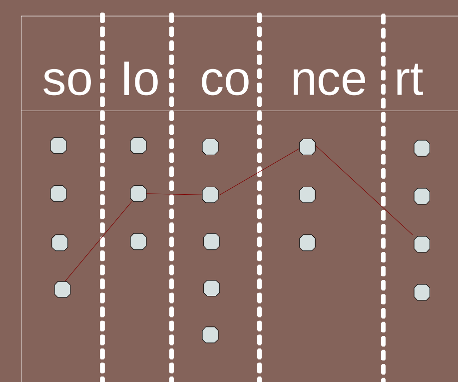




target

forgery

Handwriting



search

Investigations

In case of speech, we found concatenative attacks did no better than random guessing. Is same true for handwriting biometrics?

Models we studied

Class 1	different user, different passphrase.
Class 2	different user, true passphrase.
Class 3	true user, different passphrase.
Class 4	concatenation attack (true password constructed from unrelated writing).
Class 5	true user, true passphrase (as baseline).

"The Effectiveness of Generative Attacks on an Online Handwriting Biometric," Daniel Lopresti and Jarret Raim, *Proceedings of the Conference on Audio/Video Based Person Authentication*, July 2005.

Biometric Hash from Handwriting

Studied published technique by Vielhauer, et al. for converting handwriting into secure 24-element hash.

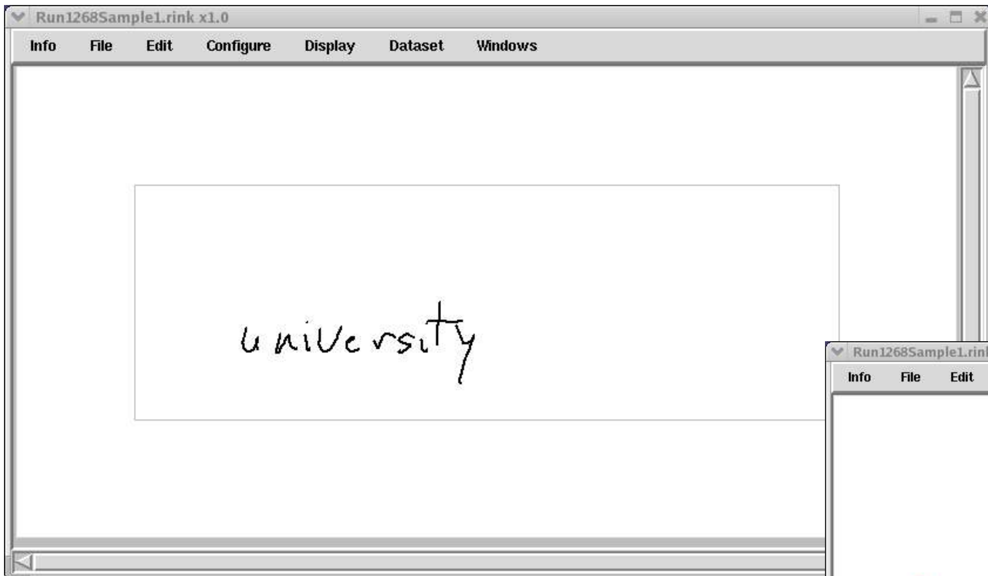
Features extracted from each sample:

1. Number of strokes
2. Total writing time (ms)
3. Total number of samples (points)
4. Sum of all local (x,y) minima and maxima
5. Aspect ratio (x/y) * 100
6. Pen-down / total writing time * 100
7. Integrated area covered by x signal
8. Integrated area covered by y signal
9. Average writing velocity in x
10. Average writing velocity in y
11. Average writing acceleration in x
12. Average writing acceleration in y
13. Effective writing velocity in x
14. Effective writing velocity in y
15. Integrated area under x, segment 1
16. Integrated area under x, segment 2
17. Integrated area under x, segment 3
18. Integrated area under x, segment 4
19. Integrated area under x, segment 5
20. Integrated area under y, segment 1
21. Integrated area under y, segment 2
22. Integrated area under y, segment 3
23. Integrated area under y, segment 4
24. Integrated area under y, segment 5

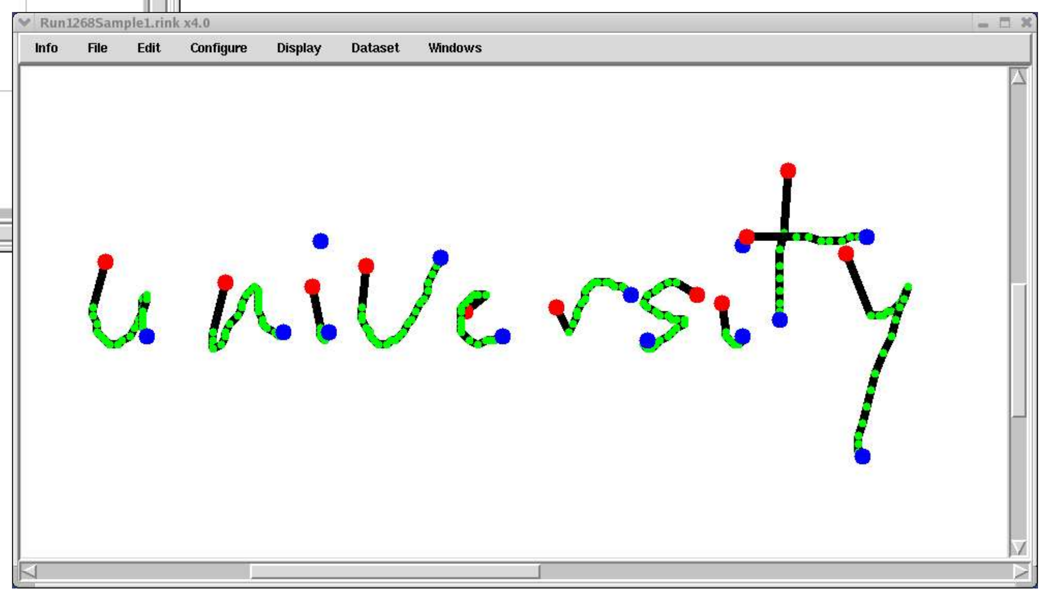
“Biometric Hash based on Statistical Features of Online Signatures,” Claus Vielhauer, Ralf Steinmetz, and Astrid Mayerhofer, *Proceedings of the Sixteenth International Conference on Pattern Recognition*, vol. 1, August 2002, pp. 123-126.

Handwriting Features #1

Snapshots of our tool for ink capture written in Tcl/Tk:

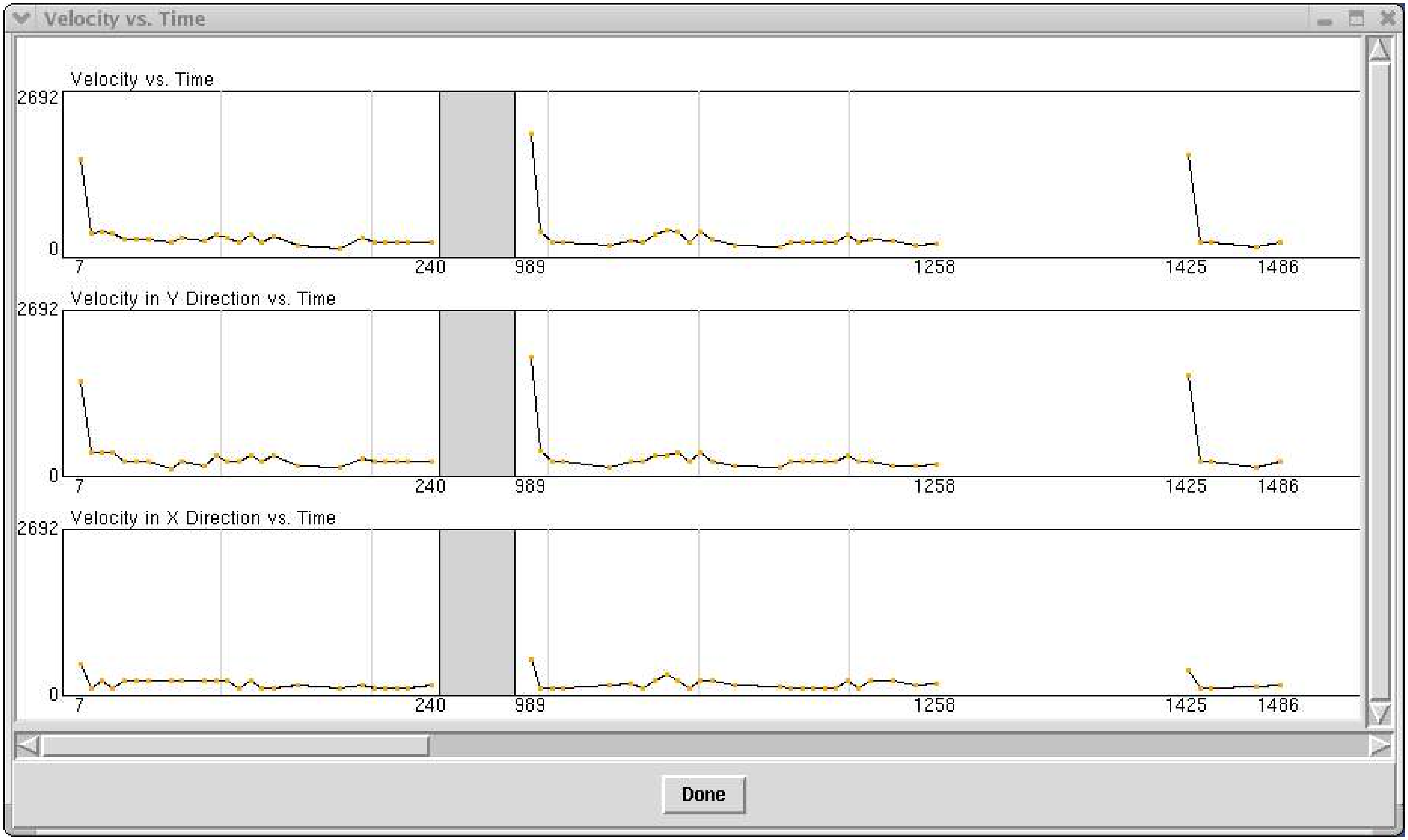


Passphrase



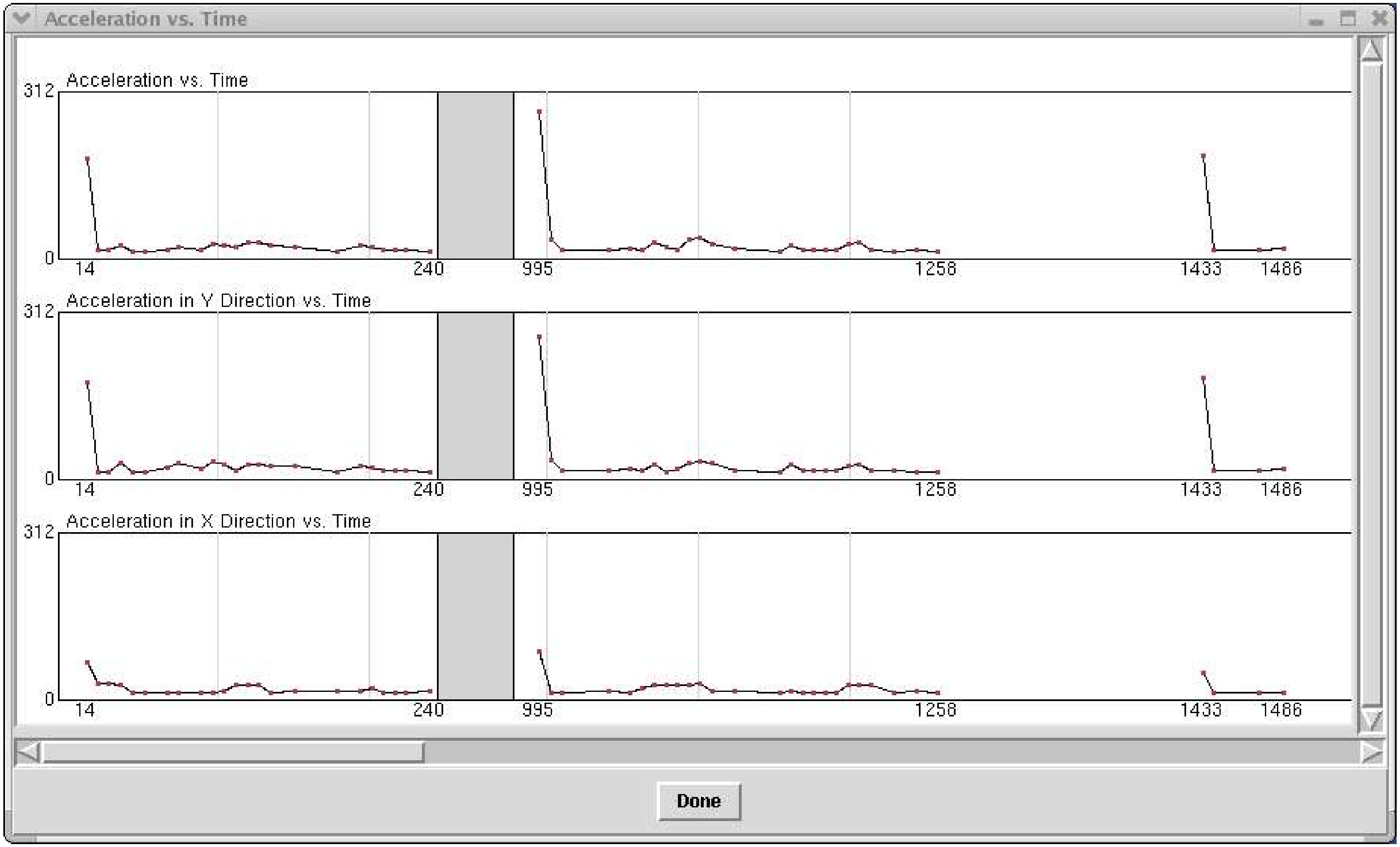
Sampled points

Handwriting Features #2



Snapshot of velocity profiles

Handwriting Features #3



Snapshot of acceleration profiles

Typical Performance Evaluation

Traditional approach: conduct study using human subjects (naive and/or skilled "forgers") and report False Reject Rate (FRR) and False Accept Rate (FAR).

- E.g., Vielhauer, et al. used 10 subjects who provided six samples and also tried to forge writing of other subjects based on static image.
- Average FRR was measured to be 7.0%.
- Average FAR was measured to be 0.0%.

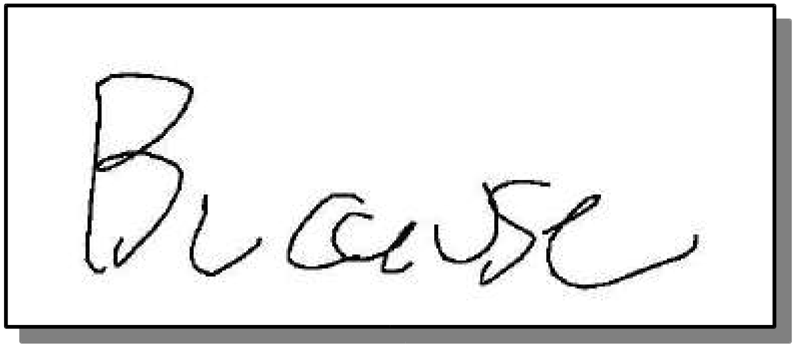
This model misses the more ominous threat.

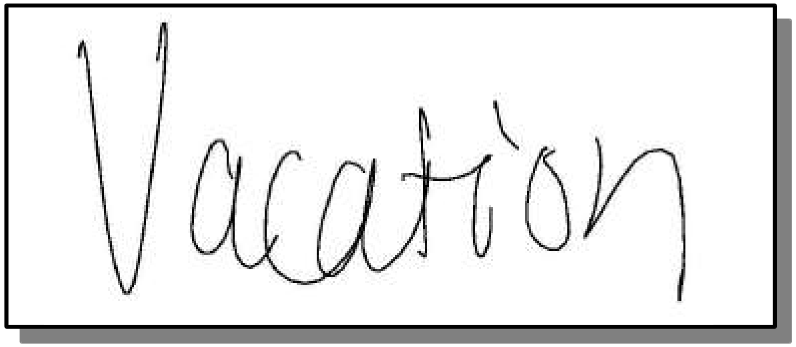
"Biometric Hash based on Statistical Features of Online Signatures," Claus Vielhauer, Ralf Steinmetz, and Astrid Mayerhofer, *Proceedings of the Sixteenth International Conference on Pattern Recognition*, vol. 1, August 2002, pp. 123-126.

Our Test Data

- Two writers each wrote four different passwords 20 or more times using Wacom Intuos tablet.
- Additional samples collected independently to support concatenative attacks.
- Dataset is small, but we are not trying to prove biometric is secure: we are studying its weaknesses.

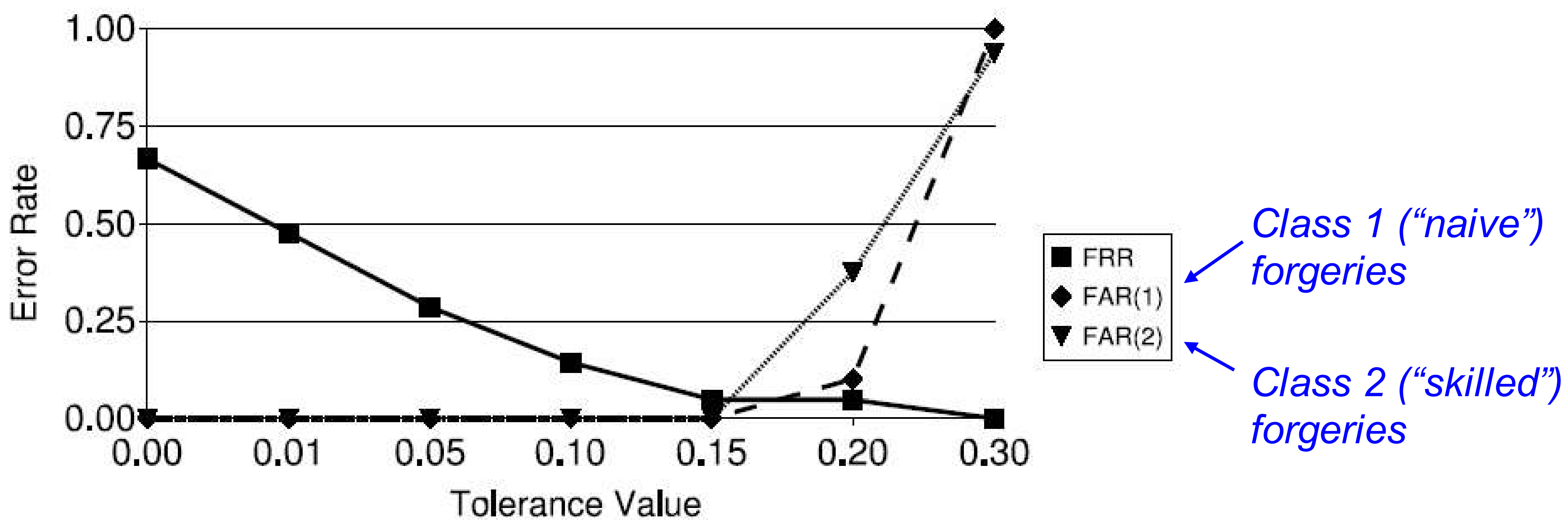
Samples of handwriting we collected:





Determining Hash Tolerance

- Training set varied from 15 to 25 samples per class.
- Cross-validation performed using 5 to 10 samples.
- Various tolerances tested, most promising was 0.15.



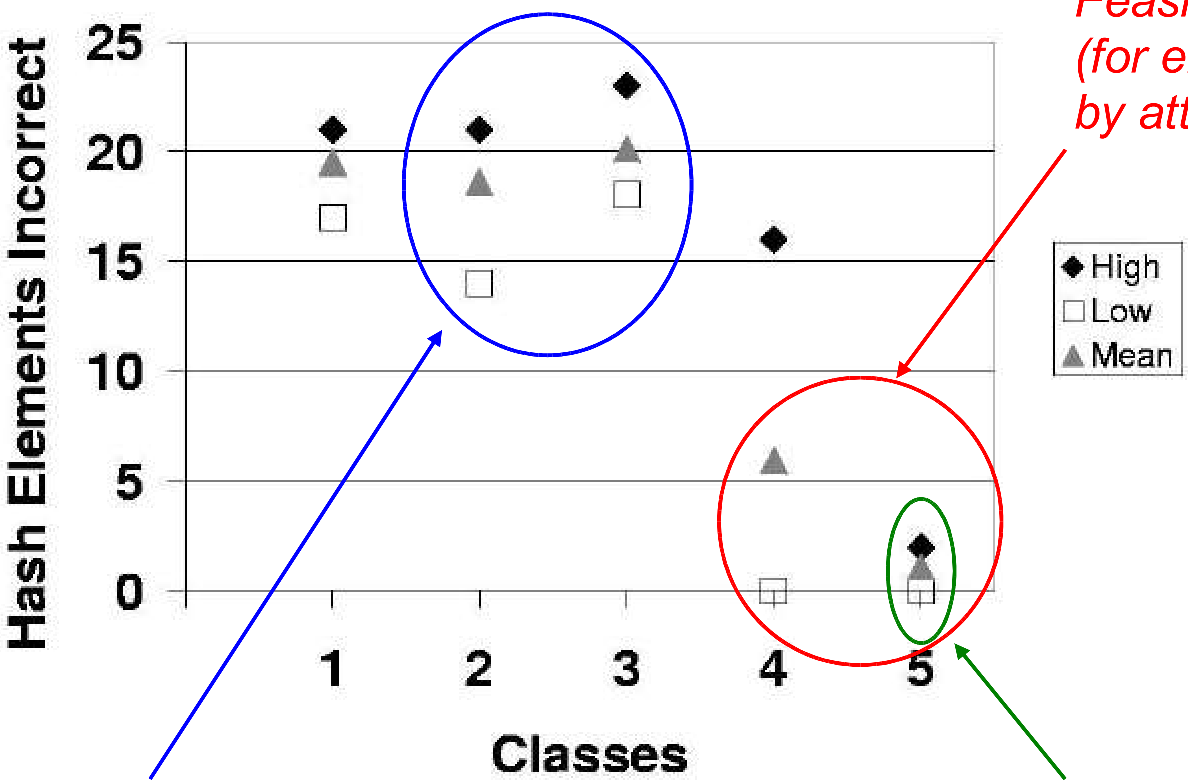
Concatenative Attack

- Separate corpus of writing samples collected and labeled on a per-character basis.
- Provides assortment of n-grams which can be selected to yield targeted password.
- Optimal concatenation can be formulated using dynamic programming, much like speech synthesis.

Original passphrase

Synthesized passphrase

Count of Incorrect Hash Elements



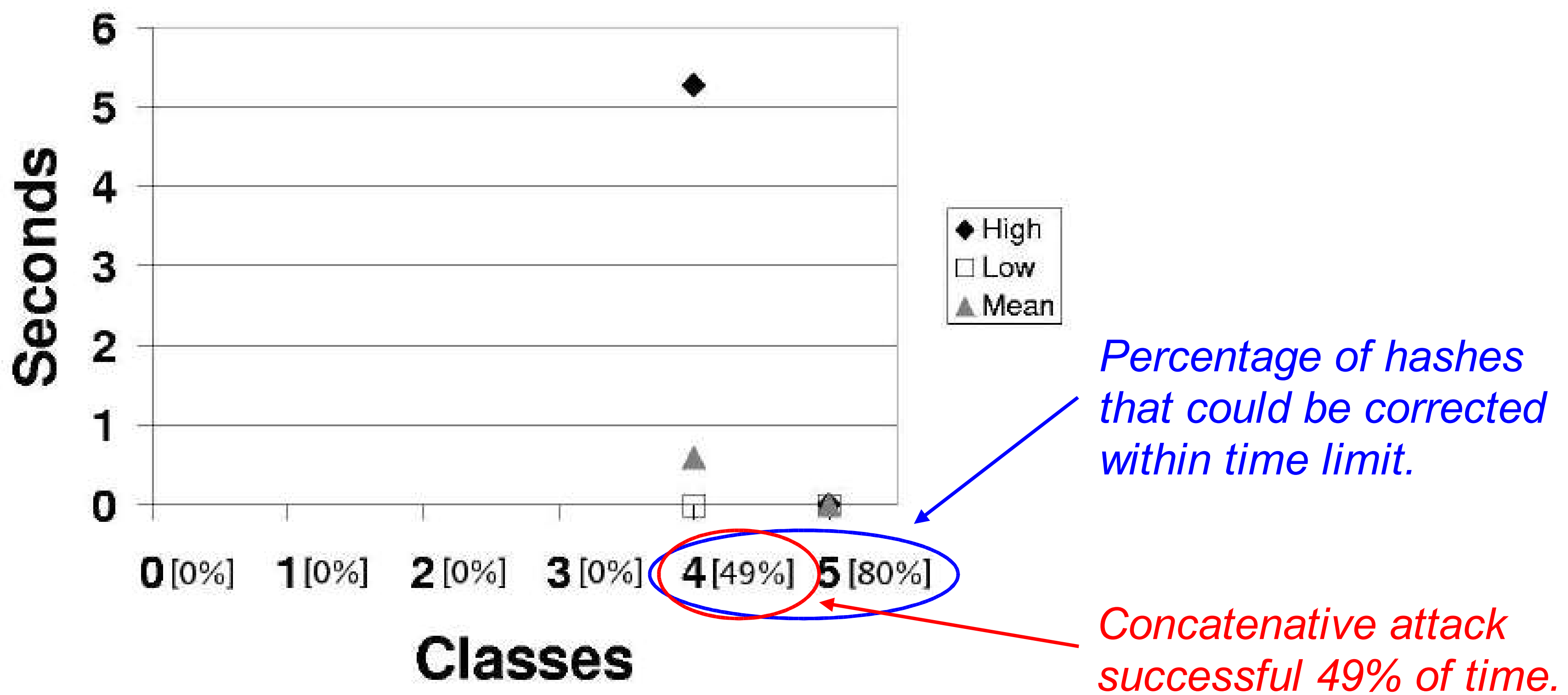
Feasible search space (for error correction, or by attacker).

Roughly same number of features sensitive to passphrase (Class 2) versus user (Class 3).

Even true user (Class 5) requires some post-error-correction.

Time to Correct Hashes

- Perform exhaustive search around hash vector.
- Timeout (failure) after 60 second time limit.
- Tests run on Pentium 4 PC, 3.2 Ghz, 1 GB RAM.



Current Data Collection

In the midst of a new, larger-scale data collection:

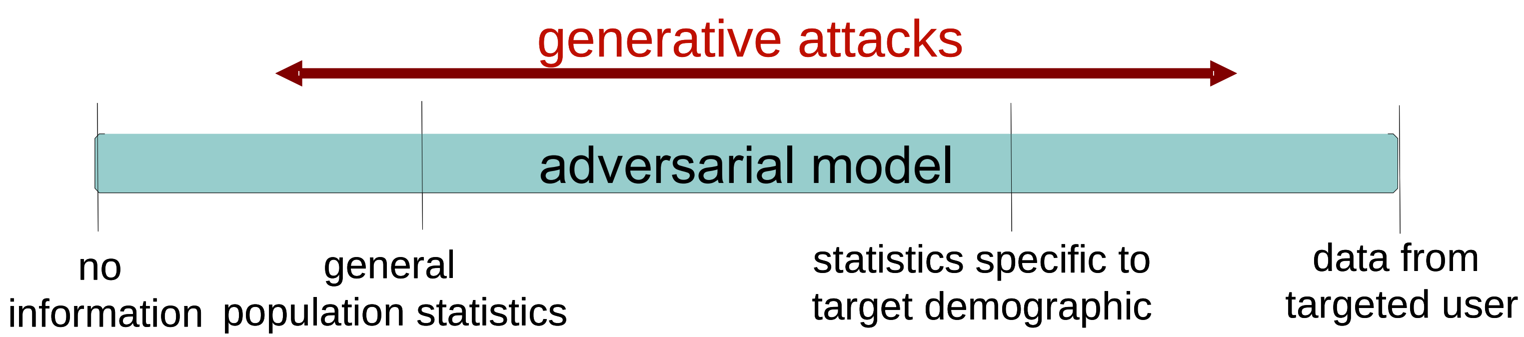
- Enlist ~100 users to write 5 passphrases 10× each on pen tablet computers (NEC, HP).
- Also have them write a general-purpose corpus to experiment with various generative attacks (guaranteed to cover all bigrams in passphrases).

Second phase (now beginning):

- Have users rewrite each passphrase 15 times.
- Ask users attempt to forge other user's writing after showing them static and/or dynamic view of target.

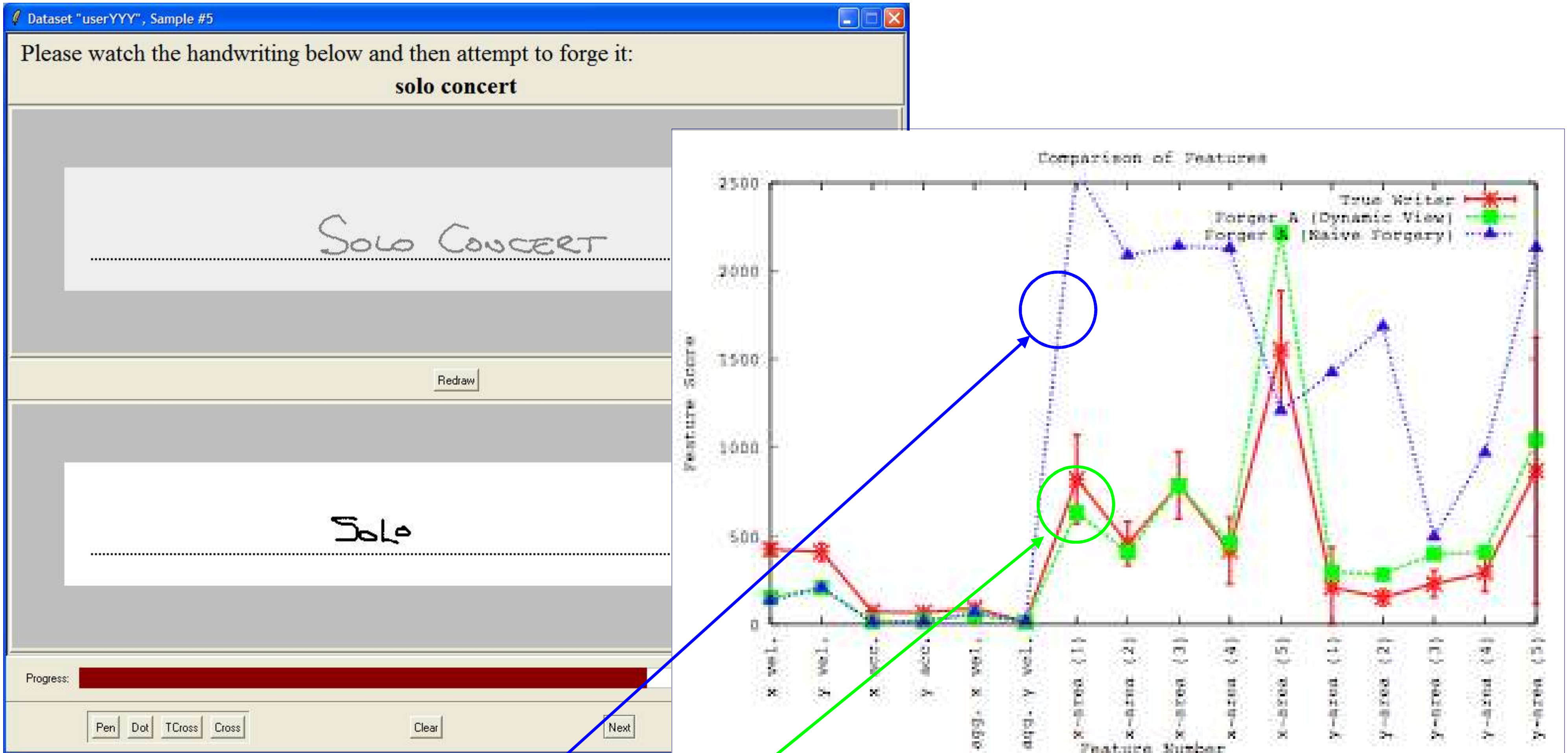
Questions We Have

- Can an average user do a credible job as a forger?
- Are some users more susceptible to attack?
- Which generative models present the greatest risk (a number have appeared in the literature)?
- What kinds of knowledge give attacker advantage?



- Can anything be done to mitigate this risk (e.g., enforcing “good” passphrase choices)?

Early Result



Naïve user

Forger with access to dynamic replay

online

offline

Conclusions

- Generative models for human behavior present a threat to security of biometric systems.
- The traditional approach to performance evaluation, i.e., human studies involving “naive” and “skilled” forgers, is inadequate for assessing this threat.
- Full extent of this threat not yet characterized: much more work needs to be done.

Acknowledgements

This work is supported in part by:

- National Science Foundation CNS CYBER TRUST 0430178, "Using generative models to evaluate and strengthen biometrically enhanced systems," Fabian Monrose, Daniel Lopresti, and Mike Reiter.

- The Keystone Alliance for Homeland Security.