

Inside Lehigh: Computer Science in the P.C. Rossin College of Engineering and Applied Science

Daniel Lopresti

Associate Professor

dal9@lehigh.edu

Department of Computer Science and Engineering

Renaissance of RCEAS

- Faculty renewal – investment in faculty hiring
- Investment in facilities and infrastructure
- Cross-disciplinary programs
- Partnering across colleges

Catalyzed through 2020 Initiative dollars.

U.S. News & World Report released its 2005 Best Graduate School rankings on April 2, 2004, and the P.C. Rossin College of Engineering and Applied Science ranked 37th, up from 41st last year.



Computer Science as a Discipline

Old view:

- Computer science as “care and feeding” of computing systems.
- People adapt work habits to limitations of computing environment, while others simply ignore it and do without.

New view:

- Computing and information technology as hidden enabler.
- Most people don't even know computer is there – it adapts quietly to their needs and work habits (“ubiquitous” computing).

Computer science pertains both to “core CS” and to integrated and/or embedded information-based environments.

Computer Science at Lehigh

- Artificial intelligence
 - Case-based reasoning
 - Machine learning
 - Intelligent agents
- Biometrics & security
- Computational biology
- Computer architecture
- Database systems
 - Text & data mining
 - Transaction & query processing
- Digital libraries & document analysis
- Embedded systems
- Enterprise information systems
- Graphics
- Human-computer interaction
 - Virtual environments
- Image processing
- Internet
 - Semantic web
 - Search
 - Peer-to-peer systems
- Machine vision
- Networking & distributed systems
- Network security
- Parallel processing
- Robotics
- Software engineering
- Ubiquitous & mobile computing

Topics today

All of these research areas are represented in our faculty.

Lehigh CSE Department

Current faculty size:

- 18 tenured / tenure-track faculty

Students:

- 244 majors (sophomore and up)
- 69 seniors graduated in academic year 2002-2003
- 37 masters students
- 15 masters degrees granted in academic year 2002-2003
- 36 Ph.D. students
- 3 Ph.D. degrees granted in academic year 2002-2003

Attract and retain top-notch undergraduate and graduate students.

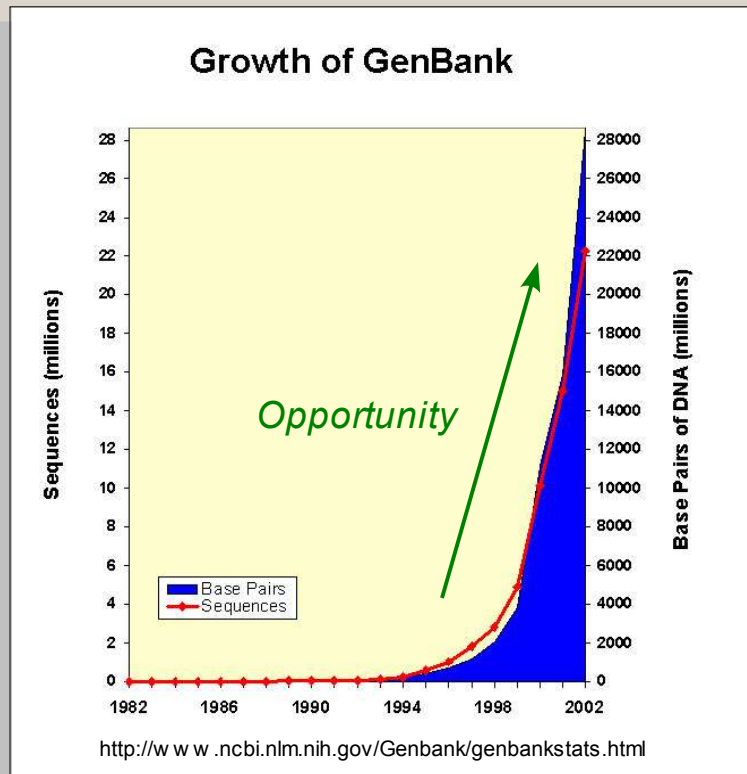
Recent CSE Faculty Hiring

- Jan 2003 • Hank Korth, Ph.D. Princeton with long distinguished career in academia and industry. Internationally known database researcher. New chair of CSE.
- Fall 2003 • John Spletzer, Ph.D. University of Pennsylvania. Sensor fusion, computer vision, robotics.
- Me* → • Dan Lopresti, Ph.D. Princeton with over 15 years in academia and industry. Pattern recognition, biometric-based security, computational biology.
- Spring 2004 • Mooi Choo Chuah, Ph.D. UCSD with over 12 years at Bell Labs. Networks, network security.
- Henry Baird, Ph.D. Princeton with over 20 years at Xerox PARC and Bell Labs. Digital libraries, document analysis, image processing.

Computational Biology

“Biology easily has 500 years of exciting problems to work on.”

Donald E. Knuth



Expertise in CSE Department:

- algorithms
- pattern recognition
- data mining
- databases
- graphics & visualization
- image processing
- robotics

New CSE Course in Comp Bio

Introduced by me in Spring 2004. We studied algorithms for:

- sequence comparison & alignment (pairwise & multiple)
- sequence assembly (shotgun sequencing)
- physical mapping of DNA
- constructing phylogenetic (evolutionary) trees
- computing genome rearrangements
- RNA and protein structure prediction
- DNA microarray analysis
- DNA computing

Materials @ <http://www.cse.lehigh.edu/~lopresti/courses.html>

Sequence Comparison

What is the problem?

- Given new DNA or protein sequence, biologist will want to search databases of known sequences to look for anything similar.

Why is it important?

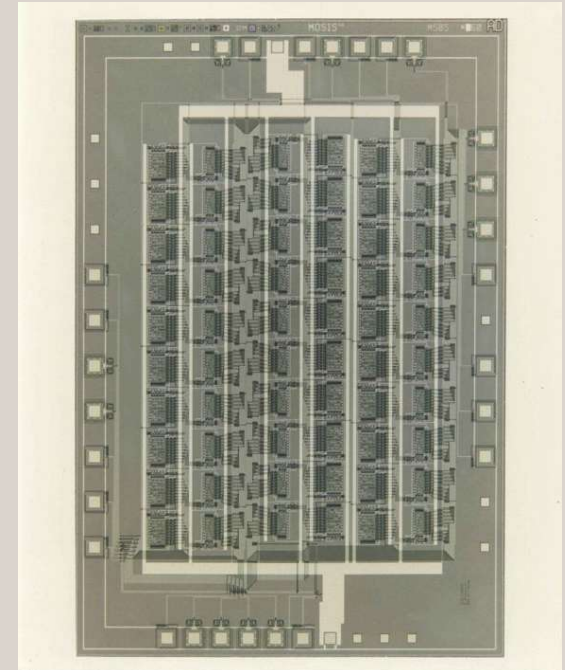
- Sequence similarity can provide clues about function and also about evolutionary relationships.
- Many other problems from computational biology incorporate some notion of sequence similarity as a basic premise.

Why is it interesting?

- Can't depend on exact matches: genomes aren't static and humans (and machines) make errors in reading sequencing gels.
- Sequences are long and databases are huge and growing rapidly.

Sequence Comparison

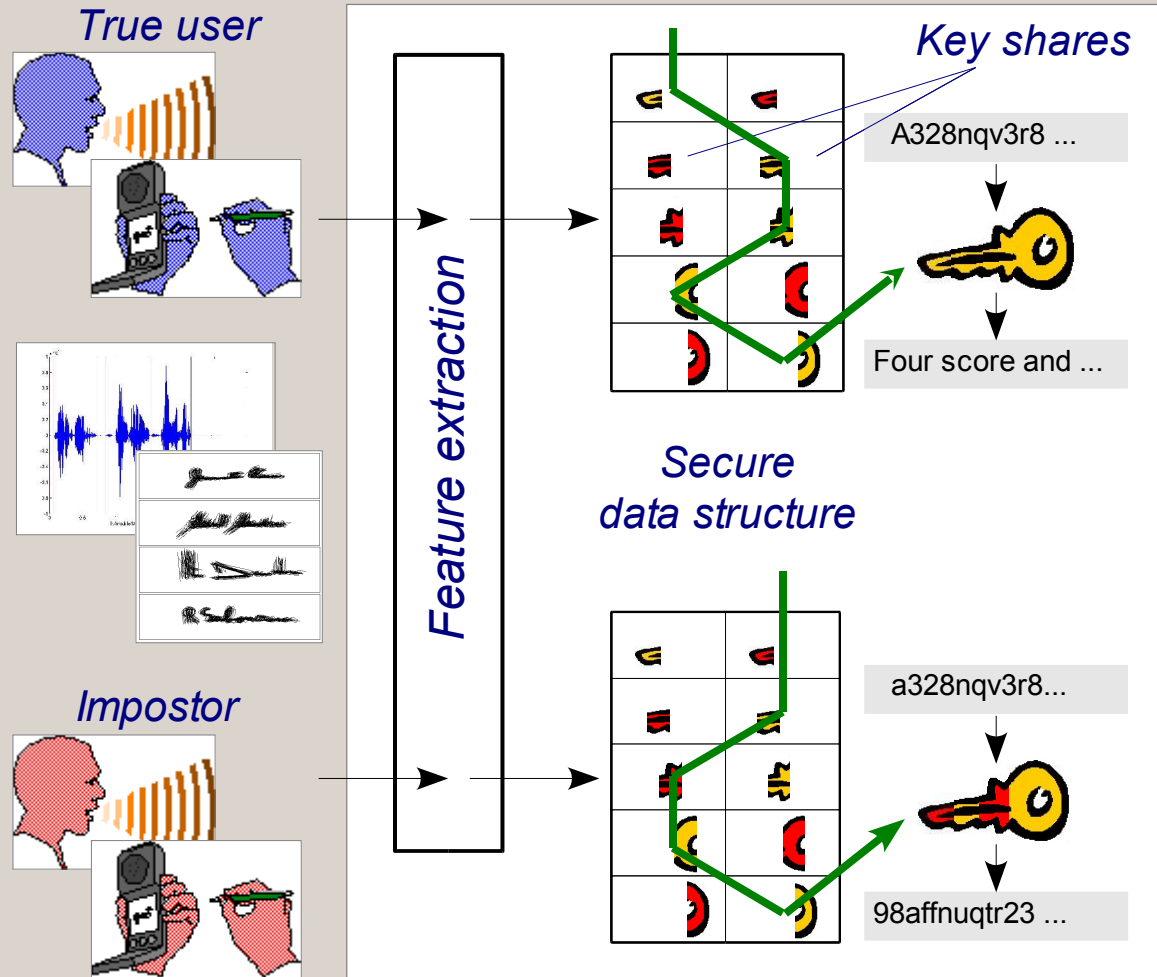
- My past work has included new algorithms for sequence comparison.
- New cross-departmental research collaborations just now developing with faculty from Biological Sciences.
- Other collaborations between CSE and faculty in new Bioengineering program are beginning as well.
- Applications outside of biology include searching noisy data (coping with errors in speech and character recognition).



*Systolic array: 30 processors
on custom VLSI chip.*

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 will select correct shares to yield proper key.



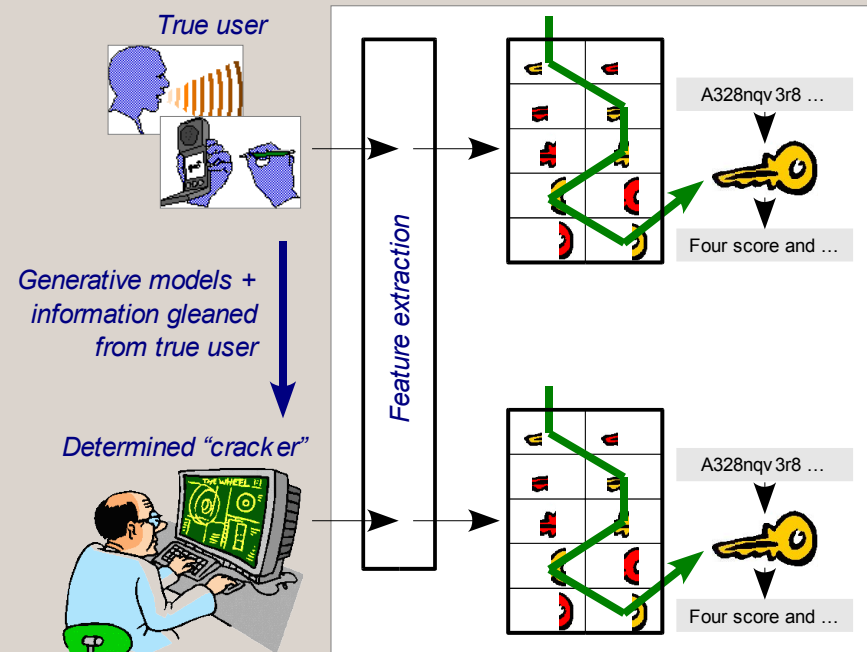
Using Biometrics to Protect Data

Current plans:

- quantify effectiveness
- increase number of bits
- identify potential attacks

Biometrics may be vulnerable:

- study generative models
- if successful, many current systems called into doubt



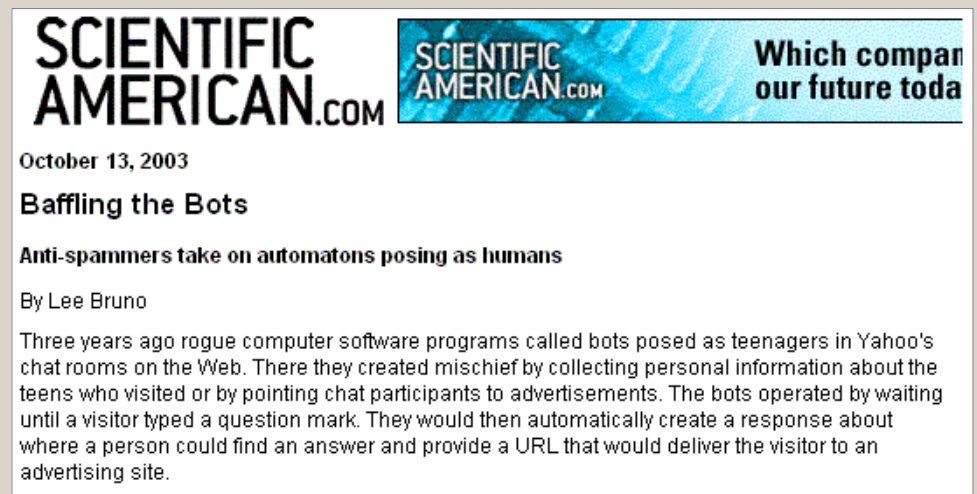
Use our experience to improve biometrics, increase security.

"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.

Protecting Online Services

The Internet has become vehicle for distributing valuable content. But malicious programs (“bots”) attempt to exploit online services intended for human users.

Idea: create a pattern recognition task that is easy for humans to solve, but hard for machines.



SCIENTIFIC AMERICAN.COM **SCIENTIFIC AMERICAN.COM** Which compares our future today

October 13, 2003

Baffling the Bots

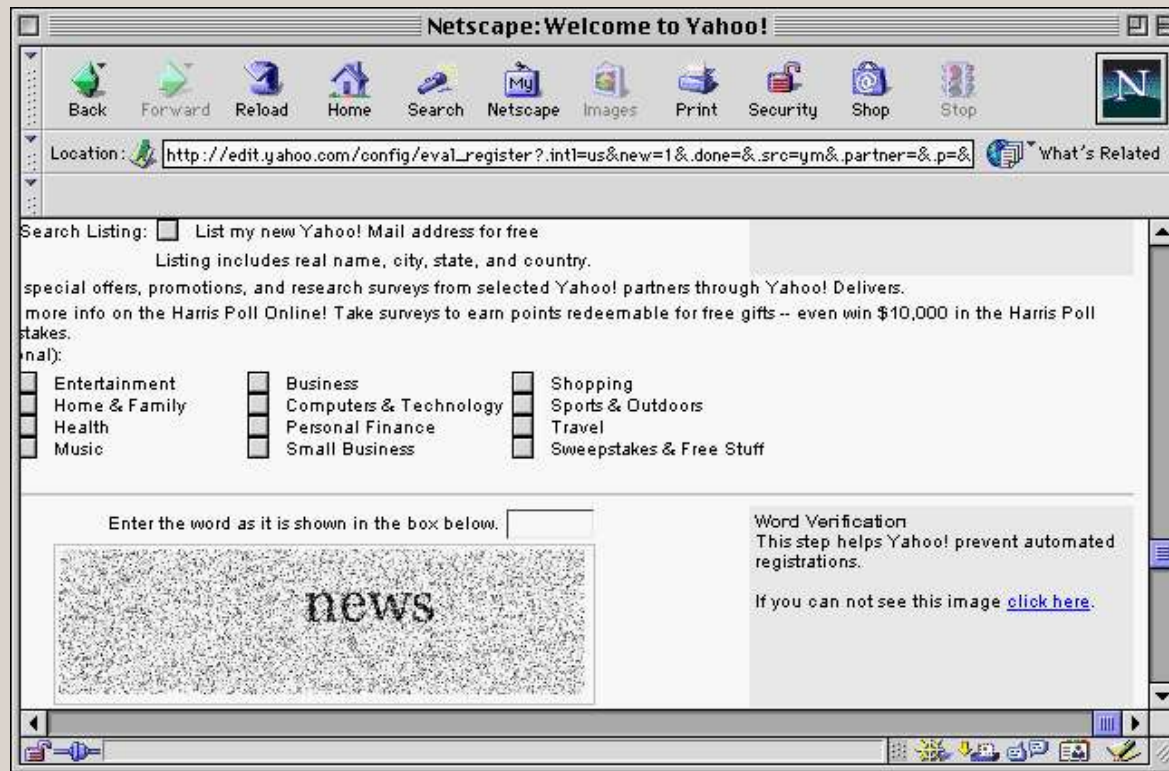
Anti-spammers take on automatons posing as humans

By Lee Bruno

Three years ago rogue computer software programs called bots posed as teenagers in Yahoo's chat rooms on the Web. There they created mischief by collecting personal information about the teens who visited or by pointing chat participants to advertisements. The bots operated by waiting until a visitor typed a question mark. They would then automatically create a response about where a person could find an answer and provide a URL that would deliver the visitor to an advertising site.

Protecting Online Services

Yahoo!'s method for protecting free email service. User must solve simple character recognition task:



Visual CAPTCHAs

Currently, most CAPTCHAs exploit gap in reading ability between humans and machines when confronted with degraded images of text. Fortunately, one of our latest hires is Henry Baird, a world expert on optical character recognition and an originator of this line of research.



EZ-Gimpy



PessimPrint



BaffleText

Second International Workshop on Human Interactive Proofs will take place at Lehigh next Spring (co-chaired by Baird and Lopresti).

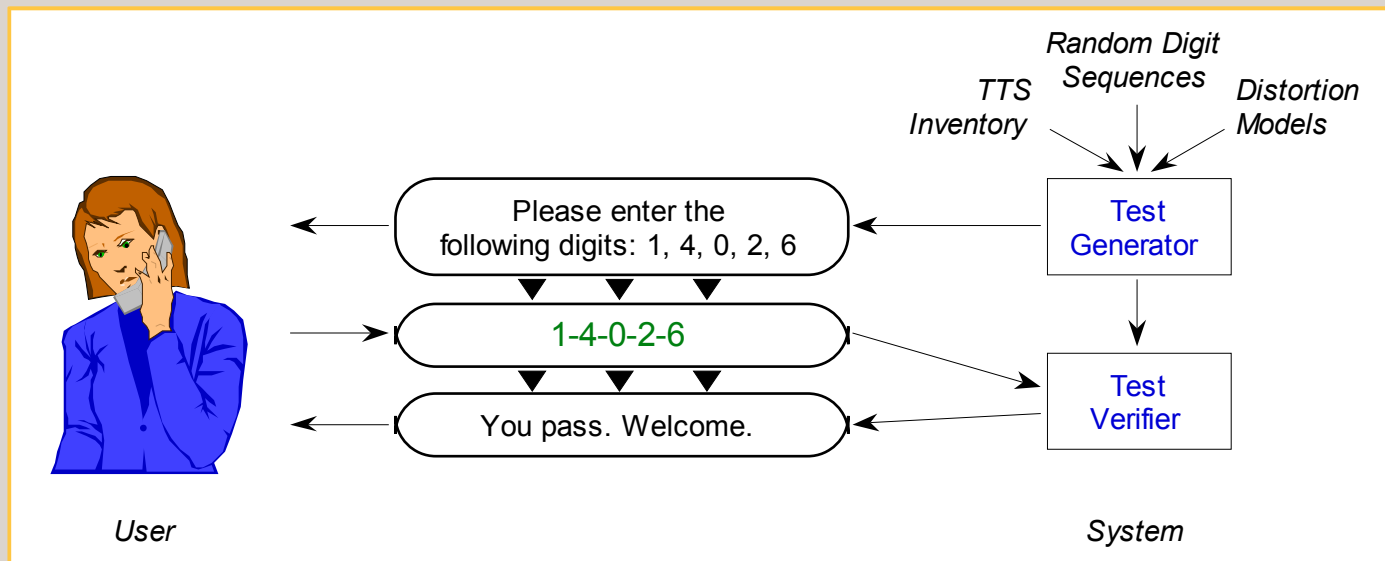
CAPTCHA = Completely Automated Public Turing test to tell Computers and Humans Apart.

Spoken Language CAPTCHAs

Speech interfaces are becoming popular, with similar security issues:

- phone access to financial services, reservations, etc.
- hands- and/or eyes-free applications (e.g., driving car)

We are continuing work I started with colleagues at Bell Labs.



"A Reverse Turing Test Using Speech," G. Kochanski, D. Lopresti, and C. Shih, Proceedings of the Seventh International Conference on Spoken Language Processing, September 2002, Denver, CO, pp. 1357-1360.

Computer Science & Engineering

- Build on the Lehigh tradition of excellence.
- Enhance research and academic agenda in:
 - core areas of computing,
 - computer applications,
 - interdisciplinary collaborations.
- Use Lehigh core strengths to reach out across disciplines.

These slides @ <http://www.cse.lehigh.edu/~lopresti/talks.html>