Umut Topkara

School of Computer Science
Carnegie-Mellon University
5000 Forbes Ave.
Pittsburgh, PA, 15213
topkara@cs.cmu.edu
http://www.cs.cmu.edu/∼topkara/

675 Sherman Ave. Apt:2
Thornwood, NY 10594
(765) 532-5741

Education

◇ **Doctor of Philosophy** January 2002- August 2007
Department of Computer Science
Purdue University, West Lafayette, IN

· Thesis: *Information security applications of natural language processing techniques.*

· Adviser: Prof. Mikhail J. Atallah

◇ **Masters in Language Technologies** August 2001 – December 2001
Degree not earned due to two-body problem
Language Technologies Institute
Carnegie Mellon University, Pittsburgh, PA

◇ **Master of Science** August 2001
Department of Computer Engineering
Bilkent University, Ankara, Turkey

· Thesis: *Prefix-Suffix Based Statistical Models of Turkish.*

· Advisers: Kemal Oflazer, Ilyas Cicekli

◇ **Bachelor of Science** August 1999
Computer Engineering and Information Science
Bilkent University, Ankara, Turkey

◇ **Scholarships**

· Turkish Scientific Research Foundation, for Ph.D. education, (2001 – present)

· Bilkent University, for M.Sc. education, (1999 – 2001).

◇ **Fellowships**

· Bilkent University, for B.Sc. education, (1995 – 1999).

Research interests

Human computer interaction aspects of information security. Information security, usable authentication, information hiding. Machine learning, natural language processing. Information retrieval, collaborative multimedia content creation. Grid computing, reliability assessment of evolving software systems.

Research projects

◇ **Usable authentication** 2006 – 2007

· Designed and implemented Shesu, an authentication system that can co-exist with current text password systems installed in UNIX operating systems.

· Shesu incorporates a flexible password mechanism that allows usage of input devices with a wide range of input bandwidth.

· We enable authentication in environments where keyboards and keypads are not practical (as in small devices), not secure (as in airplanes), or not usable (as for users with temporary or permanent disabilities).

◇ **Password Mnemonics** 2005 – 2007

- Used natural language processing methods to generate reminder sentences (mnemonics) for secure passwords. Such methods increase the usability of security policies, thereby improving the overall security of the systems they operate.
- Multiple Password Mnemonics Generation: We encoded more than one password within single mnemonic sentence without sacrificing from the security even if one of the passwords are compromised.
- EMPATHE implements a system that enables users to remember more than one password with the help of a single mnemonic sentence.
- Password Mnemonics Generation for English: Template-based mnemonic sentence generation achieves usability for password authentication mechanisms without sacrificing from their security.

◇ **Self Erasing Covert Communication** 2006 – 2007

- Designed a covert communication system in which the message life times can be tied to random updates of publicly available content.
- WANEMARK implements such a design, and can be used in blogs, emails and forum posts.

◇ **Rights Protection in Natural Language Text** 2002 – 2007

- Designed robust text watermarking methods that leverage on information asymmetry of the author and the adversary.
- Such methods are called Computationally Asymmetric Transformations, and require a disproportional amount of computation at the adversary to remove the injected message.
- EQUIMARK is a Computationally Asymmetric Transformation based information hiding system and implements a watermarking algorithm using robust synonym substitutions.
- MARKERR implements steganography and robust watermarking algorithms based on mimicking of human typographical errors.
- Implemented a text steganoghraphy detection method using statistical language model features.

◇ **Grid Software Systems** 2006 – 2007

- Service oriented software systems grow and evolve by additions of newly developed components and by improvements of existing components.
- Maintenance and management of these software systems become overwhelming due to the complexity the system reaches during this evolution.
- Designed SALSA, an architecture that achieves a sustainable growth of grid software systems by monitoring and emphasizing their emergent features.
- Designed software reliability assessment tools that help the software system architects visualize the critical points of grid software systems as their components are updated.

◇ **Adaptive information hiding** 2002 – 2004

- Designed a method for reducing the damage of information embedding process on the cover document.
- We provide an efficient way to control the stealthiness of the embedded message during the embedding process, thereby making the overall embedding process computationally more efficient.

◇ **Multilingual statistical language models** 2001

- Designed a novel method to create bilingual statistical language models from monolingual models for using in multilingual speech recognition.
- Used JANUS speech recognition toolkit to implement and test this new method.
- This method was later used in another comparative study for multilingual statistical language models.

◇ **Prefix-suffix based statistical models of Turkish** 1999 – 2001

- Studied statistical modeling of agglutinative languages using sub-word tokens in order to relieve the speech recognition systems of the large vocabulary space in these languages.
- Designed and tested several statistical sub-word language models on Turkish hypothesis lists from JANUS speech recognizer.
- The resulting language models can efficiently serve at initial hypothesis filtering.

TEACHING PREFERENCES

◇ Undergraduate Level

- Algorithms and Data Structures
- Cryptography
- Operating Systems
- Artificial Intelligence
- Computer Networks
- Information Retrieval
- I would be comfortable in teaching any systems courses in undergraduate level due to the exceptional systems education in Purdue University.

◇ Graduate Level

- Information Security
- Cryptography
- Algorithms
- Machine Learning
- Natural Language Processing

TEACHING EXPERIENCE

◇ **Purdue University, Computer Science Department** Teaching Assistant

- Compilers Principles and Practice (CS352): Office hours and laboratory instruction. Fall 2003.
- Parallel Computing (CS525): Office hours and grading. Spring 2004.
- Data Structures (CS251): Office hours and laboratory instruction. Spring 2004.
- Compiling and Programming Systems (CS502): Office hours and laboratory instruction. Fall 2004.

◇ **Bilkent University, Computer Engineering Department** Teaching Assistant (English Tuition)

- Algorithms & Programming (CS101): Engineering and science freshmen students. Laboratory course instruction, 8 lectures/week. Fall 1999.
- Introduction to Programming (MAN254): Management sophomore students. Laboratory course instruction, 8 lectures/week. Spring 2000.
- Fundamentals of Computer Science (CS201): Computer science sophomore students. Office hours, homework preparation and grading. Fall 2000.

· Fundamentals of Computer Science (CS202): Computer science sophomore students. Office hours, homework preparation and grading. Spring 2001.

Refereed Publications

◇ **Passwords for Everyone: Secure Mnemonic-based Accessible Authentication**
U. Topkara, M. Topkara, M. J. Atallah
Proceedings of the 2007 USENIX Annual Technical Conference, Santa Clara, California, June 17 – 22, 2007.

◇ **Passwords Decay, Words Endure: Secure and Re-usable Multiple Password Mnemonics**
U. Topkara, M. J. Atallah, M. Topkara
Proceedings of the 22nd Annual ACM Symposium on Applied Computing (SAC 07), Seoul, Korea, March 2007.

◇ **Information Hiding through Errors: A Confusing Approach**
M. Topkara, U. Topkara, M. J. Atallah
Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents, San Jose, CA, January 29 – February 1, 2007.

◇ **Connected in a Small World: Rapid Integration of Heterogeneous Biological Resources**
U. Topkara, C. X. Song, J. Woo, S. P. Park
Grid Computing Environments Workshop, Supercomputing (SC06), Tampa, FL, November 5, 2006.

◇ **Words Are Not Enough: Sentence Level Natural Language Watermarking**
M. Topkara, U. Topkara, M. J. Atallah
Proceedings of the ACM Workshop on Content Protection and Security (in conjunction with ACM Multimedia), Santa Barbara, CA, October 27, 2006.

◇ **The Hiding Virtues of Ambiguity: Quantifiably Resilient Watermarking of Natural Language Text through Synonym Substitutions**
U. Topkara, M. Topkara, M. J. Atallah
Proceedings of the ACM Multimedia and Security Workshop (MMSEC06), Geneva, Switzerland, September 26 – 27, 2006.

◇ **Have the cake eat it too: Infusing Usability into Password Authentication Systems**
S. Jeyaraman, U. Topkara
Proceedings of the 21st Annual Computer Security Applications Conference (ACSAC05), Tucson, AZ, December 5 – 9, 2005.

◇ **Pairwise Alignment of Protein Interaction Networks**
M. Koyuturk, Y. Kim, U. Topkara, S. Subramaniam, W. Szpankowski, and A. Grama
Journal of Computational Biology, 13(2), 182 – 199, 2006.

◇ **Inferring Functional Information from Domain Co-evolution**
Y. Kim, M. Koyuturk, U. Topkara, A. Grama, S. Subramaniam
Bioinformatics, 22(1), 40 – 49, 2006.

◇ **A Hierarchical Protocol for Increasing the Stealthiness of Steganographic Methods**
M. Topkara, U. Topkara, C. Taskiran, E. Lin, M. J. Atallah, E. J. Delp
Proceedings of the ACM Multimedia and Security Workshop (MMSEC04), Magdeburg, Germany, September 20 – 21, 2004.

◇ **Natural Language Watermarking and Tamperproofing**
M. J. Atallah, V. Raskin, C. F. Hempelmann, M. Karahan, R. Sion, U. Topkara, K. E. Triezenberg
Proceedings of the Information Hiding Workshop (IHW 2002), Noordwijkerhout, The Netherlands, October 7 – 9, 2002.

◇ **Towards Universal Speech Recognition**
Z. Wang, U. Topkara, T. Schultz, A. Waibel
Proceedings of the IEEE International Conference on Multimodal Interfaces (ICMI02), Pittsburgh, PA, October 14 – 16, 2002.

◇ **Networks with Dynamic Topologies: A Greedy Algorithm for Satellite Placement Problem**
U. Topkara, I. Uysal, I. Yoncaci
Proceedings of the 5th Symposium on Computer Networks (BAS2000), Ankara, June 15 – 16, 2000.

UNDER SUBMISSION

◇ **The Message will Self-destruct: Using Collaborative Web Publishing for Self-erasing Covert Communication**
M. J. Atallah, M. Topkara, U. Topkara

OTHER PUBLICATIONS

◇ **Assessing Reliability of Grid Software Systems Using Emergent Features**
C. X. Song, U. Topkara, J. Woo, and S.P. Park
The 2nd Workshop on Reliability and Robustness in Grid Computing Systems, the 19th Open Grid Forum (OGF19), Chapel Hill, NC, January 31, 2007.

◇ **Enabling Advanced Bioinformatics Research through SALSA: A Scalable, Simple Architecture**
C. X. Song, U. Topkara, J. Woo, S. P. Park, and M. Bina
TeraGrid Conference (TG06), Indianapolis, IN, June 12 – 15, 2006.

◇ **Attacks on Lexical Natural Language Steganography Systems (Invited Paper)**
C. Taskiran, U.Topkara, M. Topkara, E. J. Delp
Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents, San Jose, CA, January 15 – 19, 2006.

◇ **RFID Technology in Healthcare Industry**
A. Kumcu, U. Topkara, A. Unal
Krannert Technology Forecast 2004 – 2006, Krannert School of Management, Purdue University, 2004.

◇ **D-TRACK, Providing Intelligent Drug Administration Technologies for Patient Safety**
A. Kumcu, U. Topkara, A. Unal
Purdue Life Sciences Business Plan Competition, March 2004.

◇ **Proposal for Natural Language Watermarking**
M.J. Atallah, S. Bangalore, D. Hakkani-Tur, M. Karahan, S. Katzenbeisser, O. Rambow, G. Riccardi, and U. Topkara.
Summer Workshops at the Center Language and Speech Processing Proposal Evaluations, 2003.

◇ **Prefix-Suffix Based Statistical Language Models of Turkish**
U. Topkara
MS Thesis, Bilkent University Computer Engineering Department, Ankara, 2001.

◇ **Viewing Station: A Software for Viewing and Analysis of DICOM images**
E. Cetin, E. Ozkural, U. Topkara
Conference of Internet in Turkey, Ankara, 1999.

CONFERENCE PRESENTATIONS

◇ USENIX Technical Conference
Santa Clara, CA, June 2007.

◇ TERAGRID Conference
Madison, WI, June 2007.

◇ ACM Symposium on Applied Computing
Seoul, Korea, March 2007.

- SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents
  San Jose, CA, January 2007.
- Grid Computing Environments Workshop
  Tampa, FL, November 2006.
- ACM Multimedia and Security Workshop
  Geneva, Switzerland, September 2006.
- Annual Computer Security Applications Conference
  Tucson, AZ, December 2005.
- Computer Networks Symposium
  Ankara, Turkey, June 2000.

Awards and Honors

- **Conference scholarship**, 21st Annual Computer Security Applications Conference, 2005
- **Interpro Award for Nation's Best Academic IT Projects of the Year Finalist**, 1999
- **Turkish Government Foreign Education Fellowship** for ranking in University Entrance Examination, 1995

Patents

- **Robust information hiding in natural language text by increasing ambiguity**
  Provisional application, October 2006.
- **Natural language watermarking**
  Application, April 2006.

Certificates

- **Advanced Management Principles Program**, Purdue University, West Lafayette, IN. 2006
- **Foreign Trading Program**, Academy International, Ankara, Turkey. 2000

Activities and Services

- **Mentor of Summer Undergraduate Research Fellowship Student**, 2006.
- **Computer Science Graduate Student Board**, Purdue University, 2004 – 2005.
- **Turkish Folk Dance Club**, Purdue University, 2003 – 2005.
- **Turkish Student Association**, Purdue University, 2003 – 2004.
- **NATO Advanced Study Institute on Lesser Studied Languages**, Organization Committee, Summer 2000.
- **Bilkent Programming Contest**, Organization Committee, 1997, and 1999.
- **AIESEC student**, 1998, and 1999.
- **IAESTE exchange student**, Dresden Germany, 1998.
- **Bilkent University Computer Club**, President, 1997.

Work Experience

- **Post-doctoral Fellow** September 2007 – Now
  School of Computer Science, Carnegie-Mellon University, Pittsburgh, PA
- **Research Assistant** January 2006 – May 2007
  Rosen Center for Advanced Computing, Purdue University, West Lafayette, IN
- **Research Assistant** January 2005 – January 2006
  Computer Science Department, Purdue University, West Lafayette, IN
- **Teaching Assistant** August 2003 – January 2005
  Computer Science Department, Purdue University, West Lafayette, IN
- **Research Assistant** January 2002 – August 2003
  Computer Science Department, Purdue University, West Lafayette, IN

- **Research Assistant** August 2001 – December 2001
  Interactive Systems Laboratory, Carnegie-Mellon University, Pittsburgh, PA
- **Teaching Assistant** August 1999 – August 2001
  Bilkent University, Computer Engineering Department, Ankara, Turkey
- **Software Engineer** October 1998 – January 2000
  Bilkent University, Electrical and Electronics Engineering Department RAGARIS Project, Ankara, Turkey
- **Intern Software Engineer** June 1998 – August 1998
  Technische Universitaet Dresden, Institut fur Leichtbau und Kunststofftechnik, Dresden, Germany
- **Intern Software Engineer** July 1997 – August 1997
  GURIS Construction Co., Ankara, Turkey

References

- Mikhail J. Atallah
  Distinguished Professor
  Department of Computer Science, Purdue University, West Lafayette, IN, 47907.
  E-mail: mja@cs.purdue.edu
- Carol X. Song
  Senior Research Scientist and Principal Investigator for TeraGrid at Purdue
  Rosen Center for Advanced Computing, Purdue University, West Lafayette, IN, 47907.
  E-mail: carolxsong@purdue.edu
- Cristina Nita-Rotaru
  Assistant Professor
  Department of Computer Science, Purdue University, West Lafayette, IN, 47907.
  E-mail: crisn@cs.purdue.edu
- Cuneyt Taskiran
  Senior Research Scientist
  Motorola Labs, Multimedia Research Lab, Schaumburg, IL, 60196
  Email: cuneyt.taskiran@motorola.com