

A Comprehensive Bibliography of Linguistic Steganography

Richard Bergmair

NOV-04 2006

This bibliography is available from <http://semantilog.ucam.org/biblingsteg/>.

Methodological Remarks

Under our interpretation of the term, linguistic steganography excludes techniques that directly operate on typeset or written text as a graphic, or speech as a waveform. It also excludes techniques that rely on specific file formats such as ASCII or HTML.

This bibliography is divided into three sections. The primary bibliography gives pointers into the literature directly dealing with linguistic steganography, the related bibliography contains literature about more remotely connected subjects, and the implemented systems section lists software prototypes.

The primary bibliography is grouped by research collaborations. The order in which they are listed has no particular significance. Within each group, references are sorted in chronological order.

In order to bootstrap a process of community collaboration on keeping this bibliography relevant and up to date, the following material was collected in an extensive and systematic literature search carried out in March 2005 and reiterated in November 2006. The sources that were searched included the English and German web as perceived by Yahoo, MSN Search, Google, Google Scholar, CiteSeer, and the meta search engine Copernic Agent Pro, as well as the computing literature as indexed in the ACM Guide, the ACM Digital Library and IEEExplore.

PRIMARY BIBLIOGRAPHY

Peter Wayner

1. Peter Wayner. Mimic functions. *Cryptologia*, XVI(3):193–214, July 1992.
2. Peter Wayner. Strong theoretical steganography. *Cryptologia*, XIX(3):285–299, July 1995.

Purdue CERIAS

3. Mikhail J. Atallah, Victor Raskin, Michael Crogan, Christian Hempelmann, Florian Kerschbaum, Dina Mohamed, and Sanket Naik. Natural language watermarking: Design, analysis, and a proof-of-concept implementation. In Ira S. Moskowitz, editor, *Information Hiding: Fourth International Workshop*, volume 2137 of *Lecture Notes in Computer Science*, pages 185–199. Springer, April 2001.

4. Mikhail J. Atallah, Victor Raskin, Christian F. Hempelmann, Mercan Karahan, Radu Sion, Umut Topkara, and Katrina E. Triezenberg. Natural language watermarking and tamperproofing. In Fabien A. P. Petitcolas, editor, *Information Hiding: Fifth International Workshop*, volume 2578 of *Lecture Notes in Computer Science*, pages 196–212. Springer, October 2002.
5. Krista Bennett. Linguistic steganography: Survey, analysis, and robustness concerns for hiding information in text. Technical Report TR 2004-13, Purdue CERIAS, May 2004.
6. Cuneyt M. Taskiran, Umut Topkara, Mercan Topkara, and Edward J. Delp. Attacks on lexical natural language steganography systems. In *Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents*, January 2006.
7. Mercan Topkara, Guiseppa Riccardi, Dilek Hakkani-Tur, and Mikhail J. Atallah. Natural language watermarking: Challenges in building a practical system. In *Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents*, January 2006.
8. Mercan Topkara, Cuneyt M. Taskiran, and Edward J. Delp. Natural language watermarking. In Edward J. Delp and Ping W. Wong, editors, *Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents*, volume 5681, January 2005.
9. Mercan Topkara, Umut Topkara, and Mikhail J. Atallah. Words are not enough: Sentence level natural language watermarking. In *Proceedings of the ACM Workshop on Content Protection and Security (in conjunction with ACM Multimedia)*, October 2006.
10. Mercan Topkara, Umut Topkara, and Mikhail J. Atallah. Information hiding through errors: A confusing approach. In *Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents*, January 2007.
11. Umut Topkara, Mercan Topkara, and Mikhail J. Atallah. The hiding virtues of ambiguity: quantifiably resilient watermarking of natural language text through synonym substitutions. In *MM&Sec '06: Proceeding of the 8th workshop on Multimedia and security*, pages 164–174, New York, NY, USA, 2006. ACM Press.

Grothoff et al.

12. Christian Grothoff, Krista Grothoff, Ludmila Alkhutova, Ryan Stutsman, and Mikhail Atallah. Translation-based steganography. Technical Report TR 2005-39, Purdue CERIAS, 2005.
13. Christian Grothoff, Krista Grothoff, Ludmila Alkhutova, Ryan Stutsman, and Mikhail Atallah. Translation-based steganography. In *Proceedings of Information Hiding Workshop (IH 2005)*, pages 213–233. Springer, 2005.
14. Ryan Stutsman, Mikhail Atallah, Christian Grothoff, and Krista Grothoff. Lost in just the translation. In *Proceedings of the 21st Annual ACM Symposium on Applied Computing (SAC 2006)*, April 2006.

Chapman et al.

15. Mark T. Chapman. Hiding the hidden: A software system for concealing ciphertext as innocuous text. Master's thesis, University of Wisconsin-Milwaukee, May 1997.
16. Mark T. Chapman and George I. Davida. Hiding the hidden: A software system for concealing ciphertext as innocuous text. In Okamoto Sihan Qing Yongfei Han Tatsuaki, editor, *Information and Communications Security: First International Conference*, volume 1334 of *Lecture Notes in Computer Science*. Springer, August 1997.
17. Mark T. Chapman and George I. Davida. Plausible deniability using automated linguistic steganography. In George I. Davida and Yair Frankel, editors, *Infrastructure Security: International Conference*, volume 2437 of *Lecture Notes in Computer Science*, pages 276–287. Springer, October 2002.
18. Mark T. Chapman, George I. Davida, and Marc Rennhard. A practical and effective approach to large-scale automated linguistic steganography. In George I. Davida and Yair Frankel, editors, *Information Security: Fourth International Conference*, volume 2200 of *Lecture Notes in Computer Science*, page 156ff. Springer, October 2001.

Richard Bergmair

19. Richard Bergmair. Natural language steganography and an “ai-complete” security primitive. talk, December 2004. talk held at the 21st Chaos Communication Congress.
20. Richard Bergmair. Towards linguistic steganography: A systematic investigation of approaches, systems, and issues. final year thesis, April 2004. handed in in partial fulfillment of the degree requirements for the degree “B.Sc. (Hons.) in Computer Studies” to the University of Derby.

Bolshakov et al.

21. Igor A. Bolshakov. A method of linguistic steganography based on collocationally-verified synonymy. In Jessica J. Fridrich, editor, *Information Hiding: 6th International Workshop*, volume 3200 of *Lecture Notes in Computer Science*, pages 180–191. Springer, May 2004.
22. Hiram Calvo and Igor A. Bolshakov. Using selectional preferences for extending a synonymous paraphrasing method in steganography. In J. H. Sossa Azuela, editor, *Avances en Ciencias de la Computacion e Ingenieria de Computo - CIC'2004: XIII Congreso Internacional de Computacion*, pages 231–242, October 2004.

Murphy et al.

23. B. Murphy. Syntactic information hiding in plain text. Master's thesis, Department of Computer Science, Trinity College Dublin, 2001.
24. B. Murphy and C. Vogel. Statistically constrained shallow text marking: techniques, evaluation paradigm, and results. In *Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents*, January 2007.

25. B. Murphy and C. Vogel. The syntax of concealment: reliable methods for plain text information hiding. In *Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents*, January 2007.

Meral et al.

26. H. M. Meral, B. Sankur, and A. S. Ozsoy. Watermarking tools for turkish texts. In *Proceedings of the 14th IEEE Conference on Signal Processing and Communications Applications*, pages 1–4. IEEE, April 2006.
27. H. M. Meral, B. Sankur, and S. Ozsoy. Syntactic tools for natural language watermarking. In *Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents*, January 2007.

Others

28. V. Chand and C. O. Orgun. Exploiting linguistic features in lexical steganography: Design and proof-of-concept implementation. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS '06)*, volume 6, page 126b. IEEE, January 2006.
29. Yuei-Lin Chiang, Lu-Ping Chang, Wen-Tai Hsieh, and Wen-Chih Chen. Natural language watermarking using semantic substitution for chinese text. In Ton Kalker, Ingemar J. Cox, and Yong Man Ro, editors, *Digital Watermarking: Second International Workshop, IWDW 2003*, volume 2939 of *Lecture Notes in Computer Science*, pages 129–140. Springer, October 2003.
30. B. Macq and O. Vybornova. A method of text watermarking using presuppositions. In *Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents*, January 2007.
31. Hiroshi Nakagawa, Kouji Sampei, Tsutomu Matsumoto, Shuji Kawaguchi, Kyoto Makino, and Ichiro Murase. Text information hiding with preserved meaning – a case for japanese documents. *IPSJ Transaction*, 42(9):2339 – 2350, 2001. originally published in Japanese. A similar paper was disseminated by the first author in English and is kept available for download from <http://www.r.dl.itc.u-tokyo.ac.jp/nakagawa/academic-res/finpri02.pdf>.
32. Michiharu Niimi, Sayaka Minewaki, Hideki Noda, and Eiji Kawaguchi. A framework of text-based steganography using sd-form semantics model. *IPSJ Journal*, 44(8), August 2003.
33. M. Hassan Shirali-Shahreza and Mohammad Shirali-Shahreza. A new approach to persian/arabic text steganography. In *Proceedings of the 5th IEEE/ACIS International Conference on Computer and Information Science*, pages 310–315, Washington, DC, USA, 2006. IEEE Computer Society.
34. Xingming Sun, Gang Luo, and Huajun Huang. Component-based digital watermarking of chinese texts. In *InfoSecu '04: Proceedings of the 3rd international conference on Information security*, pages 76–81. ACM Press, 2004.
35. Adam J. Tenenbaum. Linguistic steganography: Passing covert data using text-based mimicry. final year thesis, April 2002. submitted in partial fulfillment of the requirements for the degree of “Bachelor of Applied Science” to the University of Toronto.

36. Ozlem Uzuner. Natural language processing with linguistic information for digital fingerprinting and watermarking. In *Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents*, January 2006.
37. Keith Winstein. Lexical steganography through adaptive modulation of the word choice hash, January 1999. Was disseminated during secondary education at the Illinois Mathematics and Science Academy. The paper won the third prize in the 2000 Intel Science Talent Search.

RELATED BIBLIOGRAPHY

38. Mikhail J. Atallah, Craig J. McDonough, Victor Raskin, and Sergei Nirenburg. Natural language processing for information assurance and security: an overview and implementations. In Mary Ellen Zurko and Steven J. Greenwald, editors, *NSPW '00: Proceedings of the 2000 workshop on New security paradigms*, pages 51–65. ACM Press, September 2000.
39. Richard Bergmair and Stefan Katzenbeisser. Towards human interactive proofs in the text-domain. In Kan Zhang and Yuliang Zheng, editors, *Proceedings of the 7th Information Security Conference*, volume 3225 of *Lecture Notes in Computer Science*, pages 257–267. Springer Verlag, September 2004.
40. Richard Bergmair and Stefan Katzenbeisser. Content-aware steganography: About lazy prisoners and narrow-minded wardens. Technical Report fki-252-05, Technische Universität München, Institut für Informatik AI/Cognition Group, December 2005.
41. Richard Bergmair and Stefan Katzenbeisser. Content-aware steganography: About lazy prisoners and narrow-minded wardens. In *Proceedings of the 8th Information Hiding Workshop*, Lecture Notes in Computer Science. Springer Verlag, 2006. in print.
42. Igor A. Bolshakov and Alexander Gelbukh. Synonymous paraphrasing using wordnet and internet. In Farid Meziane and Elisabeth Elisabeth Metais, editors, *Natural Language Processing and Information Systems: 9th International Conference on Applications of Natural Language to Information Systems, NLDB 2004*, volume 3136 of *Lecture Notes in Computer Science*, pages 312–323. Springer, June 2004.
43. Victor Raskin, Christian F. Hempelmann, Katrina E. Triezenberg, and Sergei Nirenburg. Ontology in information security: a useful theoretical foundation and methodological tool. In Victor Raskin and Steven J. Greenwald, editors, *NSPW '01: Proceedings of the 2001 workshop on New security paradigms*, pages 53–59. ACM Press, September 2001.
44. Peter Wayner. *Disappearing Cryptography – Information Hiding: Steganography & Watermarking*. Morgan Kaufmann Publishers, Los Altos, CA 94022, USA, second edition, 2002. Chapters 6 and 7 serve as good introductions to mimic functions.

IMPLEMENTED SYSTEMS

45. Mark T. Chapman and George I. Davida. Nicetext. Website, August 1997. <http://www.nicetext.com/>, accessed 2005-03-09.

46. Texthide. Website. <http://www.texthide.com/>, accessed 2005-03-20.
47. Textsign. Website. accessed 2005-03-20.
48. Steven E. Hugg. Stegparty. Website, November 1999. <http://www.fasterlight.com/hugg/projects/stegparty.html>, accessed 2005-03-25.
49. Kevin Maher. Texto. circulating on the web, February 1995. <http://www.ecn.org/crypto/soft/texto.zip>, accessed 2005-03-22.
50. David McKellar. Spammimic. Website, June 2000. <http://www.spammimic.com/>, accessed 2004-04-12.
51. Paul Shields. Stegano. circulating on the web, November 2001. <http://zooid.org/~paul/crypto/natlang/stegano-1.02.tar.gz>, accessed 2005-03-25.
52. John Walker. Steganosaurus. circulating on the web, December 1994. <http://www.fourmilab.ch/stego/stego.shar.gz>, accessed 2005-03-25.
53. Peter Wayner. Mimicry applet. Website, August 1997. <http://www.wayner.org/texts/mimic/>, accessed 2004-04-12.
54. Keith Winstein. Tyrannosaurus lex. Website, January 1999. <http://alumni.imsa.edu/~keithw/tlex/>, accessed 2005-03-09.
55. Michal Zalewski. snowdrop. freshmeat entry, September 2002. <http://freshmeat.net/projects/snowdrop/>, accessed 2005-03-20.