A Comprehensive Bibliography of Linguistic Steganography

Richard Bergmair

University of Cambridge Computer Laboratory, 15 JJ Thomson Avenue, Cambridge CB3 0FD, UK

ABSTRACT

In this paper, we will attempt to give a comprehensive bibliographic account of the work in linguistic steganography published up to date. As the field is still in its infancy there is no widely accepted publication venue. Relevant work on the subject is scattered throughout the literature on information security, information hiding, imaging and watermarking, cryptology, and natural language processing. Bibliographic references within the field are very sparse. This makes literature research on linguistic steganography a tedious task and a comprehensive bibliography a valuable aid to the researcher.

Keywords: comprehensive bibliography, linguistic, natural language, steganography, steganalysis, information hiding, watermarking

1. MOTIVATION

Linguistic steganography is, as a research area, still in its infancy, with an average of only 4.5 publications appearing per year since 2001. Nevertheless, the field seems to be gaining attention lately as research groups around the world independently discover linguistic steganography as a relevant area of study (the numbers in Figure 1 seem to support this, although the small volume of publications naturally doesn't allow for any significant statistics).

There is no single widely accepted publication venue for this highly interdisciplinary subject, which is why relevant contributions are scattered throughout publication venues of information security, information hiding, imaging and watermarking, cryptology, and natural language processing. Another problem is that the field does not have an agreed upon name which could serve as an index into the literature. A Google search for 'linguistic steganography', for example, will only return a very small fraction of the relevant work.

This makes literature search in linguistic steganography a tedious task and consequently contributes to the cost of independent discovery and improper attribution of results. Although there are at least 8 research groups that have repeatedly published on linguistic steganography, only two of them are routinely cited by others, while the bulk of the 36 papers we have identified is never cited at all. This, of course, applies to all scientific publications, but one might argue that the situation is particularly counterproductive for a young field like linguistic steganography.

Therefore we have compiled, in this paper, what we hope to be a comprehensive account of the literature in linguistic steganography indexed as of November 2006. The bibliography resides on `http://semantilog.ucam.org/biblingsteg/` and is available in BibTeX and PDF formats. With the help of the community, we hope to keep it up to date and relevant.

Author e-mail `rbergmair@acm.org`, web `http://richard.bergmair.eu/`;
This bibliography resides at `http://semantilog.ucam.org/biblingsteg/`

1992	1	1
1995	1	1
1997	2	2
1999	1	1
2001	3	2
2002	3	2
2003	2	2
2004	6	5
2005	3	2
2006	9	5
(2007)	(4)	(3)

Figure 1. Number of publications on linguistic steganography per year. (Rightmost column disregards publications published by the most active group)

2. 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.

3. CALL FOR COMMUNITY PARTICIPATION

This bibliography will be updated and maintained on `http://semantilog.ucam.org/biblingsteg/`. The maintainer invites the participation of the whole community in keeping this resource relevant and up to date and will gratefully incorporate any suggestions brought to his attention by contacting the maintainer via e-mail at `biblingsteg@semantilog.ucam.org`.

- If you spot any omissions in this bibliography, please bring it to the maintainer's attention, preferably by emailing a BibTeX file including abstracts.

- If you are an author or editor publishing new material on linguistic steganography, please bring it to the maintainer's attention, preferably by emailing a BibTeX file including abstracts.

- If you are maintaining a personal webpage of a researcher in linguistic steganography or any other webpage related to work in linguistic steganography, please send a link to your webpage to the maintainer and add a link to `http://semantilog.ucam.org/biblingsteg/` to your own webpage. This will help search engines produce more relevant results for queries related to linguistic steganography.

PRIMARY BIBLIOGRAPHY

Peter Wayner

1. P. Wayner, "Mimic functions," *Cryptologia* **XVI**, pp. 193–214, July 1992.

2. P. Wayner, "Strong theoretical steganography," *Cryptologia* **XIX**, pp. 285–299, July 1995.

Purdue CERIAS

3. M. J. Atallah, V. Raskin, M. Crogan, C. Hempelmann, F. Kerschbaum, D. Mohamed, and S. Naik, "Natural language watermarking: Design, analysis, and a proof-of-concept implementation," in *Information Hiding: Fourth International Workshop*, I. S. Moskowitz, ed., *Lecture Notes in Computer Science* **2137**, pp. 185–199, Springer, April 2001.

4. M. J. Atallah, V. Raskin, C. F. Hempelmann, M. Karahan, R. Sion, U. Topkara, and K. E. Triezenberg, "Natural language watermarking and tamperproofing," in *Information Hiding: Fifth International Workshop*, F. A. P. Petitcolas, ed., *Lecture Notes in Computer Science* **2578**, pp. 196–212, Springer, October 2002.

5. K. Bennett, "Linguistic steganography: Survey, analysis, and robustness concerns for hiding information in text," Tech. Rep. TR 2004-13, Purdue CERIAS, May 2004.

6. M. Topkara, C. M. Taskiran, and E. J. Delp, "Natural language watermarking," in *Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents*, E. J. Delp and P. W. Wong, eds., **5681**, January 2005.

7. C. M. Taskiran, U. Topkara, M. Topkara, and E. 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.

8. M. Topkara, G. Riccardi, D. Hakkani-Tur, and M. 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.

9. U. Topkara, M. Topkara, and M. 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*, pp. 164–174, ACM Press, (New York, NY, USA), 2006.

10. M. Topkara, U. Topkara, and M. J. Atallah, "Words are not enough: Sentence level natural language watermarking," in *Proceedings of the ACM Workshop on Content Protection and Security (in conjuction with ACM Multimedia)*, October 2006.

11. M. Topkara, U. Topkara, and M. 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.

Grothoff et al.

12. C. Grothoff, K. Grothoff, L. Alkhutova, R. Stutsman, and M. Atallah, "Translation-based steganography," Tech. Rep. TR 2005-39, Purdue CERIAS, 2005.

13. C. Grothoff, K. Grothoff, L. Alkhutova, R. Stutsman, and M. Atallah, "Translation-based steganography," in *Proceedings of Information Hiding Workshop (IH 2005)*, pp. 213–233, Springer, 2005.

14. R. Stutsman, M. Atallah, C. Grothoff, and K. 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. M. T. Chapman, "Hiding the hidden: A software system for concealing ciphertext as innocuous text," Master's thesis, University of Wisconsin-Milwaukee, May 1997.

16. M. T. Chapman and G. I. Davida, "Hiding the hidden: A software system for concealing ciphertext as innocuous text," in *Information and Communications Security: First International Conference*, O. S. Q. Yongfei Han Tatsuaki, ed., *Lecture Notes in Computer Science* **1334**, Springer, August 1997.

17. M. T. Chapman, G. I. Davida, and M. Rennhard, "A practical and effective approach to large-scale automated linguistic steganography," in *Information Security: Fourth International Conference*, G. I. Davida and Y. Frankel, eds., *Lecture Notes in Computer Science* **2200**, p. 156ff, Springer, October 2001.

18. M. T. Chapman and G. I. Davida, "Plausible deniability using automated linguistic steganography," in *Infrastructure Security: International Conference*, G. I. Davida and Y. Frankel, eds., *Lecture Notes in Computer Science* **2437**, pp. 276–287, Springer, October 2002.

Richard Bergmair

19. R. 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.

20. R. Bergmair, "Natural language steganography and an "ai-complete" security primitive." talk, December 2004. talk held at the 21st Chaos Communication Congress.

Bolshakov et al.

21. I. A. Bolshakov, "A method of linguistic steganography based on collocationally-verified synonymy.," in *Information Hiding: 6th International Workshop*, J. J. Fridrich, ed., *Lecture Notes in Computer Science* **3200**, pp. 180–191, Springer, May 2004.

22. H. Calvo and I. A. Bolshakov, "Using selective preferences for extending a synonymous paraphrasing method in steganography," in *Avances en Ciencias de la Computacion e Ingenieria de Computo - CIC'2004: XIII Congreso Internacional de Computacion*, J. H. S. Azuela, ed., pp. 231–242, October 2004.

Murphy et al.

23. B. Murphy, "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.

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.

Meral et al.

25. 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*, pp. 1–4, IEEE, April 2006.

26. 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

27. K. 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.

28. H. Nakagawa, K. Sampei, T. Matsumoto, S. Kawaguchi, K. Makino, and I. Murase, "Text information hiding with preserved meaning – a case for japanese documents," *IPSJ Transaction* **42**(9), pp. 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.

29. A. 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.

30. M. Niimi, S. Minewaki, H. Noda, and E. Kawaguchi, "A framework of text-based steganography using sd-form semantics model," *IPSJ Journal* **44**, August 2003.

31. Y.-L. Chiang, L.-P. Chang, W.-T. Hsieh, and W.-C. Chen, "Natural language watermarking using semantic substitution for chinese text," in *Digital Watermarking: Second International Workshop, IWDW 2003*, T. Kalker, I. J. Cox, and Y. M. Ro, eds., *Lecture Notes in Computer Science* **2939**, pp. 129–140, Springer, October 2003.

32. X. Sun, G. Luo, and H. Huang, "Component-based digital watermarking of chinese texts," in *InfoSecu '04: Proceedings of the 3rd international conference on Information security*, pp. 76–81, ACM Press, 2004.

33. O. 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.

34. 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)*, **6**, p. 126b, IEEE, January 2006.

35. M. H. Shirali-Shahreza and M. Shirali-Shahreza, "A new approach to persian/arabic text steganography," in *Proceedings of the 5th IEEE/ACIS International Conference on Computer and Information Science*, pp. 310–315, IEEE Computer Society, (Washington, DC, USA), 2006.

36. 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.

RELATED BIBLIOGRAPHY

37. M. J. Atallah, C. J. McDonough, V. Raskin, and S. Nirenburg, "Natural language processing for information assurance and security: an overview and implementations," in *NSPW '00: Proceedings of the 2000 workshop on New security paradigms*, M. E. Zurko and S. J. Greenwald, eds., pp. 51–65, ACM Press, September 2000.

38. V. Raskin, C. F. Hempelmann, K. E. Triezenberg, and S. Nirenburg, "Ontology in information security: a useful theoretical foundation and methodological tool," in *NSPW '01: Proceedings of the 2001 workshop on New security paradigms*, V. Raskin and S. J. Greenwald, eds., pp. 53–59, ACM Press, September 2001.

39. P. Wayner, *Disappearing Cryptography – Information Hiding: Steganography & Watermarking*, Morgan Kaufmann Publishers, Los Altos, CA 94022, USA, second ed., 2002. Chapters 6 and 7 serve as good introductions to mimic functions.

40. I. A. Bolshakov and A. Gelbukh, "Synonymous paraphrasing using wordnet and internet," in *Natural Language Processing and Information Systems: 9th International Conference on Applications of Natural Language to Information Systems, NLDB 2004*, F. Meziane and E. E. Metais, eds., *Lecture Notes in Computer Science* **3136**, pp. 312–323, Springer, June 2004.

41. R. Bergmair and S. Katzenbeisser, "Towards human interactive proofs in the text-domain," in *Proceedings of the 7th Information Security Conference*, K. Zhang and Y. Zheng, eds., *Lecture Notes in Computer Science* **3225**, pp. 257–267, Springer Verlag, September 2004.

42. R. Bergmair and S. Katzenbeisser, "Content-aware steganography: About lazy prisoners and narrow-minded wardens," Tech. Rep. fki-252-05, Technische Universität München, Institut für Informatik AI/Cognition Group, Dec. 2005.

43. R. Bergmair and S. 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.

IMPLEMENTED SYSTEMS

44. J. Walker, "Steganosaurus." circulating on the web, December 1994. `http://www.fourmilab.ch/stego/stego.shar.gz`, accessed 2005-03-25.

45. K. Maher, "Texto." circulating on the web, February 1995. `http://www.ecn.org/crypto/soft/texto.zip`, accessed 2005-03-22.

46. M. T. Chapman and G. I. Davida, "Nicetext." Website, August 1997. `http://www.nicetext.com/`, accessed 2005-03-09.

47. P. Wayner, "Mimicry applet." Website, August 1997. `http://www.wayner.org/texts/mimic/`, accessed 2004-04-12.

48. K. Winstein, "Tyrannosaurus lex." Website, January 1999. `http://alumni.imsa.edu/~keithw/tlex/`, accessed 2005-03-09.

49. S. E. Hugg, "Stegparty." Website, November 1999. `http://www.fasterlight.com/hugg/projects/stegparty.html`, accessed 2005-03-25.

50. D. McKellar, "Spammimic." Website, June 2000. `http://www.spammimic.com/`, accessed 2004-04-12.

51. P. Shields, "Stegano." circulating on the web, November 2001. `http://zooid.org/~paul/crypto/natlang/stegano-1.02.tar.gz`, accessed 2005-03-25.

52. M. Zalewski, "snowdrop." freshmeat entry, September 2002. `http://freshmeat.net/projects/snowdrop/`, accessed 2005-03-20.

53. "Texthide." Website. `http://www.texthide.com/`, accessed 2005-03-20.

54. "Textsign." Website. accessed 2005-03-20.