

CHRISTOPHER W. CLIFTON

Department of Computer Sciences
Purdue University
305 N. University St
West Lafayette, IN 47907-2107
(765)494-6005
<http://www.cs.purdue.edu/people/clifton>

72 Limberlost Lane
West Lafayette, IN 47906-9400
(765)746-3979
United States Citizen

Email: clifton@cs.purdue.edu

EDUCATION

Princeton University

Princeton, NJ

Ph.D. in Computer Science, June 1991; M.A. in Computer Science, February 1988. Awarded IBM Graduate Fellowship for academic years 1988-89, 1989-90; GTE Fellowship for Academic year 1987-88; NCR Fellowship for academic year 1986-87.

Massachusetts Institute of Technology

Cambridge, MA

M.S. and B.S. in Computer Science and Engineering, September 1986.

RESEARCH INTERESTS

Novel uses of data mining technology, including data mining of text, data mining techniques applied to interoperation of heterogeneous information sources, and security and privacy issues raised by data mining. Fundamental data mining challenges posed by these applications, such as extracting knowledge from noisy data, identifying knowledge in highly skewed data (few examples of “interesting” behavior), limits on learning. Database support for widely distributed and autonomously controlled information, particularly information administration issues.

PROFESSIONAL EXPERIENCE

Purdue University

August 2001 to Present

Associate Professor in the Department of Computer Sciences. Research on Data Mining and Database Management System Security. Taught graduate courses *Database Systems*, *Distributed Database Systems*, *Information Security*, *Advanced Topics in Distributed Systems*; graduate seminars on *Data Mining and Security* and *Data Mining*; an undergraduate seminars *Introduction to Data Mining*; and the undergraduate course *Computer Architecture*.

The MITRE Corporation

September 1995 to August 2001

Principal Scientist (promoted from Lead Scientist in 1999) in the Component Based Architectures & Data Management department of the Information Technology Center. Research on Database Management Systems for applications with loosely-structured data; in particular data mining for text and image databases. Director’s Distinguished Accomplishment Award for Data Mining on Text and Imagery, August 1997; SPOT Recognition award for Knowledge Discovery in Text, September 1998; Program Recognition Award for GeoNODÉ, June 2000; Senior Vice President/General Manager’s Award for Joint Battlespace Infosphere, February 2001.

Northwestern University

September 1991 to August 1995

Assistant Professor in the Department of Electrical Engineering and Computer Science. Research on Database Management Systems for applications with loosely-structured data; in particular software database and heterogeneous databases. Developed and taught a two quarter sequence *Software Project Management and Development*; also taught *Introduction to Computer Programming, Fundamentals of Computer System Software, Data Management and Information Processing, Introduction to Database Systems*, and a graduate seminar on *Heterogeneous Databases*.

Matsushita Information Technology Laboratory

June through August 1991

Research on data management systems for office information.

Princeton University

September 1986 through June 1991

Research on database support for multimedia. Thesis title: *HyperFile, a Database Manager for Documents* (Advisor: Professor Hector Garcia-Molina.) Teaching Assistant for *Introduction to Programming Systems and Database Systems*.

Raritan Valley Community College

January through May 1989

Adjunct Faculty. Taught *Foundations of Computer Science*.

Massachusetts Institute of Technology

September 1985 through June 1986

Thesis title: *Dynamic Load Balancing* (Advisors: Professor Nancy Lynch and Dr. Flaviu Cristian (IBM Almaden Research).) Teaching Assistant in the Department of Electrical Engineering and Computer Science: Recitation instructor for *Laboratory in Software Engineering*; Tutorial instructor for *Structure and Interpretation of Computer Programs*.

I.B.M. Almaden Research

Summers 1983 and 1984, January through August 1985

Co-op student through the MIT VI-A program. Worked with a team of computer scientists studying fault tolerant computer systems using a cluster of processors. Responsible for creation of a system to prevent multiple management of a database given a communication failure, the development of a debugging system for programs spread across multiple processors, and the design and implementation of a dynamic load balancing system.

Tektronix, Inc.

September 1979 through August 1981, Summer 1982

Software technician. Worked through a High school work-study program, as well as summers, as part of a team of 10 engineers responsible for development and maintenance of microprocessor cross-compilers on a CDC-175. Personally responsible for implementation and documentation, as well as partial development, of a microprocessor-based controller.

BOOKS AND BOOK CHAPTERS

1. Chris Clifton, Wei Jiang, Mummoorthy Murugesan, and M. Ercan Nergiz, "Is Privacy Still an Issue for Data Mining?", Chapter 18 in *Next Generation of Data Mining*, Hillol Kargupta, Jiawei Han, Philip Yu, Rajeev Motwani, and Vipin Kumar (Eds.), CRC Press, 2008.
2. Christopher W. Clifton, Deirdre K. Mulligan, and Raghu Ramakrishnan, "Data Mining and Privacy: An Overview", in *Privacy and Technologies of Identity A Cross-Disciplinary Conversation*, Katherine Strandburg and Daniela Stan Raicu (Eds.), Springer-Verlag, 2006, pp. 191-208.
3. Jaideep Vaidya, Michael Zhu, and Chris Clifton, *Privacy-Preserving Data Mining*, Volume 19 in *Advances in Information Security*, Springer, 2005, 99p.

4. Chris Clifton, Murat Kantarcioglu and Jaideep Vaidya, "Privacy-Preserving Data Mining", in *Foundations and Advances in Data Mining*, Wesley Chu and T.Y. Lin, eds., Studies in Fuzziness and Soft Computing vol. 180, Springer-Verlag, 2005, pp. 309-340.
5. Chris Clifton, Murat Kantarcioglu, and Jaideep Vaidya, "Defining Privacy for Data Mining", in *Data Mining: Next Generation Challenges and Future Directions*, AAAI Press, 2004, pp. 255-272.
6. Chris Clifton, "Security and Privacy", Chapter 18 in *Handbook of Data Mining*, Lawrence Erlbaum Associates, Mahwah, NJ, April 2003, pp. 441-452.

JOURNAL ARTICLES

1. "Efficient Privacy-Preserving Similar Document Detection", Mummoorthy Murugesan, Wei Jiang, Chris Clifton, Luo Si, and Jaideep Vaidya, *The VLDB Journal*, to appear.
2. " δ -Presence Without Complete World Knowledge", M. Ercan Nergiz and Chris Clifton, *Transactions on Knowledge and Data Engineering*, IEEE Computer Society, Los Alamitos, CA, to appear.
3. "MultiRelational k-Anonymity", Mehmet Ercan Nergiz, Chris Clifton, and Ahmet Erhan Nergiz, *Transactions on Knowledge and Data Engineering* 21(8):1104-1117, IEEE Computer Society, Los Alamitos, CA, August 2009.
4. "Privacy-Preserving Kth Element Score over Vertically Partitioned Data", with Jaideep Vaidya, *Transactions on Knowledge and Data Engineering* 21(2):253-258, IEEE Computer Society, Los Alamitos, CA, February 2009.
5. "Privacy-Preserving Decision Trees over Vertically Partitioned Data", Jaideep Vaidya, Chris Clifton, Murat Kantarcioglu, and A. Scott Patterson, *ACM Transactions on Knowledge Discovery in Data* 2(3), October 2008. 27pages.
6. "Privacy Preserving Naive Bayes Classification", with Jaideep Vaidya and Murat Kantarcioglu, *The VLDB Journal* 17(4):879-898, July 2008.
7. "Transforming Semi-Honest Protocols to Ensure Accountability", with Wei Jiang and Murat Kantarcioglu, *Data and Knowledge Engineering* 65(1):57-74 special section on "Privacy Aspects of Data Mining", Elsevier Science, Amsterdam, April 2008. Expanded version of PADM06 paper.
8. "An Approach to Identifying Beneficial Collaboration Securely in Decentralized Logistics Systems", with Ananth Iyer, Richard Cho, Wei Jiang, Murat Kantarcioglu, and Jaideep Vaidya, *Manufacturing & Service Operations Management* 10(1):108-125, INFORMS, Linthicum, Maryland, Winter 2008.
9. "Thoughts on k-Anonymization", with Mehmet Ercan Nergiz, *Data and Knowledge Engineering* 63(3):622-645, Elsevier Science, Amsterdam, December 2007. Invited article, expanded version of PDM06 paper.
10. "A Secure Distributed Framework for Achieving k-Anonymity", with Wei Jiang, *The VLDB Journal* 15(4):316-333 Special Issue on Privacy-Preserving Data Management, November 2006.
11. "Secure Set Intersection Cardinality with Application to Association Rule Mining", with Jaideep Vaidya, *Journal of Computer Security* 13(4):593-622, IOS Press, November 2005.
12. "Privacy Preserving Clustering with Distributed EM Mixture Modeling", with Xiaodong Lin and Michael Zhu, *Knowledge and Information Systems* 8(1):66-81, Springer-Verlag, London, July 2005.
13. "Privacy-Preserving Data Mining: Why, How, and What For?", with Jaideep Vaidya, *IEEE Security & Privacy* 2(6):19-27, New York, NY, November/December 2004.
14. "Privacy Preserving Data Mining of Association Rules on Horizontally Partitioned Data", with Murat Kantarcioglu, *IEEE Transactions on Knowledge and Data Engineering* 16(9):1026-1037, IEEE Computer Society Press, Los Alamitos, CA, September 2004.
15. "TopCat: Data Mining for Topic Identification in a Text Corpus", with Robert Cooley and Jason Rennie, *Transactions on Knowledge and Data Engineering* 16(8):949-964, IEEE Computer Society Press, Los

Alamitos, CA, August 2004.

16. "Change Detection in Overhead Imagery using Neural Networks", *International Journal of Applied Intelligence*, 18(2):215-234, Kluwer Academic Publishers, Dordrecht, The Netherlands, March 2003.
17. "Emerging Standards for Data Mining", with Bhavani Thuraisingham, *Computer Standards and Interfaces* 23(3):187-193, Elsevier Science, Amsterdam, July 2001.
18. "Using Sample Size to Limit Exposure to Data Mining", invited article, *Journal of Computer Security* 8(4):281-307, IOS Press, November 2000. Expanded version of *Protecting Against Data Mining through Samples*.
19. "SEMINT: A Tool for Identifying Attribute Correspondences in Heterogeneous Databases Using Neural Networks", with Wen-Syan Li, *Data and Knowledge Engineering* 33(1):49-84, Elsevier Science, Amsterdam, April 2000.
20. "Database Integration Using Neural Networks: Implementation and Experiences", with Wen-Syan Li and Shu-Yao Liu, *Knowledge and Information Systems* 2(1):73-96, Springer-Verlag, Singapore, February 2000.
21. "Multidatabase Query Processing with Uncertainty in Global Keys and Attribute Values", with Peter Scheuermann and Wen-Syan Li, *Journal of the American Society for Information Science* 49(3):283-301, John Wiley & Sons, Philadelphia, PA, March 1998.
22. "Dynamic Integration in Multidatabase Systems", with Wen-Syan Li, *Journal of Database Management* 7(1):28-40, IDEA Group Publishing, Harrisburg, PA, Winter 1996.
23. "HyperFile: A Data and Query Model for Documents", with Hector Garcia-Molina and David Bloom, *VLDB Journal* 4(1):100-141, The Boxwood Press, Pacific Grove, CA, March 1995.

REFEREED CONFERENCES AND WORKSHOPS

1. "t-Plausibility: Semantic Preserving Text Sanitization", with Wei Jiang, Mummoorthy Murugesan and Luo Si, *the 2009 IEEE International Conference on Privacy, Security, Risk and Trust (PASSAT-09)*, Vancouver, Canada, August 29-31, 2009.
2. "Providing Privacy through Plausibly Deniable Search", with Mummoorthy Murugesan, *2009 SIAM International Conference on Data Mining (SDM09)*, Sparks, Nevada, April 30-May 2, 2009.
3. "Plausibly Deniable Search", with Mummoorthy Murugesan, *Proceedings of the Workshop on Secure Knowledge Management (SKM 2008)*, Dallas, Texas, November 3-4, 2008.
4. "A Privacy Preserving Credentialing System for Health Care", with Ahmet Erhan Nergiz, *Proceedings of the Workshop on Secure Knowledge Management (SKM 2008)*, Dallas, Texas, November 3-4, 2008.
5. "Similar Document Detection with Limited Information Disclosure", with Wei Jiang, Mummoorthy Murugesan, and Luo Si, *Proceedings of the 24th International Conference on Data Engineering (ICDE 2008)*, Cancun, Mexico, April 7-12, 2008.
6. "Identifying Rare Classes with Sparse Training Data", with Mingwu Zhang, Wei Jiang, and Sunil Prabhakar, *Proceedings of the 18th International Conference on Database and Expert Systems Applications*, Regensburg, Germany, September 4-7, 2007.
7. "Hiding the Presence of Individuals from Shared Databases", with Mehmet Nergiz and Maurizio Atzori, *2007 ACM SIGMOD International Conference on Management of Data*, Beijing, China, June 11-14, 2007.
8. "AC-Framework for Privacy-Preserving Collaboration", with Wei Jiang, *2007 SIAM International Conference on Data Mining (SDM07)*, Minneapolis, Minnesota, April 26-28, 2007.
9. "MultiRelational k-Anonymity", with Mehmet Ercan Nergiz and Ahmet Erhan Nergiz, *The 23rd IEEE International Conference on Data Engineering (ICDE 2007)*, Istanbul, Turkey, April 16-20, 2007.
10. "Transforming Semi-Honest Protocols to Ensure Accountability", with Wei Jiang, *The ICDM workshop on Privacy Aspects of Data Mining (PADM06)*, Hong Kong, China, Dec. 18, 2006.

11. “Thoughts on k-Anonymization”, with Mehmet Ercan Nergiz, *The Second International Workshop on Privacy Data Management (PDM06)* held in conjunction with *The 22nd International Conference on Data Engineering*, Atlanta, Georgia, April 8, 2006.
12. “Privacy-Preserving Distributed k-Anonymity”, with Wei Jiang, *The 19th Annual IFIP WG 11.3 Working Conference on Data and Applications Security*, August 7-10, 2005, Storrs, Connecticut.
13. “Security Issues in Querying Encrypted Data”, with Murat Kantarcioglu *The 19th Annual IFIP WG 11.3 Working Conference on Data and Applications Security*, August 7-10, 2005, Storrs, Connecticut.
14. “Privacy-Preserving Decision Trees over Vertically Partitioned Data”, with Jaideep Vaidya *The 19th Annual IFIP WG 11.3 Working Conference on Data and Applications Security*, August 7-10, 2005, Storrs, Connecticut.
15. “Dependable Real-time Data Mining”, with Bhavani Thuraisingham, Latifur Khan, John Maurer, and Marion Ceruti, *8th IEEE International Symposium on Object-Oriented Real-Time Distributed Computing (ISORC 2005)*, May 18-20, 2005, Seattle, Washington.
16. “Privacy-Preserving Top-K Queries” with Jaideep Vaidya, *The 21st International Conference on Data Engineering (ICDE 2005)*, April 5-8, 2005, Tokyo, Japan.
17. “Knowledge Discovery from Transportation Network Data”, with Wei Jiang, Jaideep Vaidya, Zahir Balaoria, and Brett Banich, *The 21st International Conference on Data Engineering (ICDE 2005)*, April 5-8, 2005, Tokyo, Japan. Best paper, Industrial Track.
18. “Privacy-Preserving Outlier Detection”, with Jaideep Vaidya, *The Fourth IEEE International Conference on Data Mining*, November 1-4, 2004, Brighton, UK.
19. “Privately Computing a Distributed k-nn Classifier”, with Murat Kantarcioglu, *8th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)*, September 20-24, 2004, Pisa, Italy.
20. “When do Data Mining Results Violate Privacy?”, with Murat Kantarcioglu and Jiashun Jin, *The Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 22-25, 2004, Seattle, Washington.
21. “Privacy Preserving Data Integration and Sharing”, with AnHai Doan, Ahmed Elmagarmid, Murat Kantarcioglu, Gunther Schadow, Dan Suci, and Jaideep Vaidya, *The 9th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD’2004)*, June 13, 2004, Paris, France.
22. “Privacy Preserving Naive Bayes Classifier for Vertically Partitioned Data”, with Jaideep Vaidya, *2004 SIAM International Conference on Data Mining*, April 22-24, 2004, Lake Buena Vista, Florida.
23. “Derived Access Control Specification for XML”, with Siddhartha Goel and Arnon Rosenthal, *Workshop on XML Security* held in association with the *10th ACM Conference on Computer and Communications Security*, October 31, 2003, Fairfax, Virginia.
24. “Leveraging the “Multi” in Secure Multi-Party Computation”, with Jaideep Vaidya, *Workshop on Privacy in the Electronic Society* held in association with the *10th ACM Conference on Computer and Communications Security*, October 30, 2003, Washington, DC.
25. “Privacy-Preserving K-Means Clustering over Vertically Partitioned Data”, with Jaideep Vaidya, *The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 24 - 27, 2003, Washington, D.C. Honorable Mention, Best Paper Competition.
26. “Assuring Privacy when Big Brother is Watching”, with Murat Kantarcioglu, *The 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD’2003)*, June 13, 2003, San Diego, California.
27. “Privacy Preserving Association Rule Mining in Vertically Partitioned Data”, with Jaideep Vaidya, *The Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, July 23 - 26,

2002, Edmonton, Alberta, Canada.

28. "Privacy-preserving Distributed Mining of Association Rules on Horizontally Partitioned Data", with Murat Kantarcioglu, *The ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD'2002)*, June 2, 2002, Madison, Wisconsin.
29. "Coordinating Accessibility versus Restrictions in Distributed Object Systems", with Arnon Rosenthal, *3rd International Symposium on Distributed Objects & Applications*, 17-20 September, 2001 Rome, Italy.
30. "Directions for Web and E-Commerce Applications Security", with Bhavani Thuraisingham, Amar Gupta, Elisa Bertino, and Elena Ferrari, *Enabling Technologies: Infrastructure for Collaborative Enterprises, WET ICE 2001*, Cambridge, Massachusetts, June 20-22, 2001.
31. "Real-Time Data Mining of Multimedia Objects", with Bhavani Thuraisingham, Marion G. Ceruti, and John Maurer, *Fourth IEEE International Symposium on Object-Oriented Real-Time Distributed Computing (ISORC 2001)*, Magdeburg, Germany, May 2-4, 2001.
32. "Developing Custom Intrusion Detection Filters Using Data Mining", with Gary Gengo, *2000 Military Communications International Symposium (MILCOM2000)*, Los Angeles, California, October 22-25, 2000.
33. "TopCat: Data Mining for Topic Identification in a Text Corpus", with Robert Cooley, *3rd European Conference on Principles and Practice of Knowledge Discovery in Databases*, Prague, Czech Republic, September 15-18, 1999.
34. "Protecting Against Data Mining through Samples", *Thirteenth Annual IFIP WG 11.3 Working Conference on Database Security*, July 26-28, 1999, Seattle, WA. Expanded version invited for submission to *Journal of Computer Security*, IOS Press (see above).
35. "Data Mining on Text", with Rick Steinheiser, in *Proceedings of the Twenty-Second Annual International Computer Software and Applications Conference*, Vienna, Austria, August 19-21, 1998.
36. "Query Flocks: A Generalization of Association Rule Mining", with Dick Tsur, Jeffrey D. Ullman, Serge Abiteboul, Rajeev Motwani, Svetlozar Nestorov, and Arnon Rosenthal, in *Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data*, Seattle, WA, June 1-4, 1998.
37. "Knowledge Discovery in Text", with Arnon Rosenthal, Rick Steinheiser, and Jeffrey D. Ullman, *First Federal Data Mining Symposium*, AFCEA, December 16-17, 1997, Washington, D.C. Revised version appeared in the *European Conference on Machine Learning Workshop on Text Mining*, April 24, 1998, Chemnitz, Germany.
38. "Experience with a Combined Approach to Attribute-Matching Across Heterogeneous Databases", with Ed Housman and Arnon Rosenthal, *7th IFIP 2.6 Working Conference on Database Semantics*, October 7-10, 1997, Leysin, Switzerland.
39. "Dynamic Integration and Query Processing with Ranked Role Sets", with Peter Scheuermann and Wen-Syan Li, *Proceedings of the First IFCIS International Conference on Cooperative Information Systems*, June 19-21, 1996, Brussels, Belgium.
40. "Security and Privacy Implications of Data Mining", with Don Marks, *Proceedings of the ACM SIGMOD Workshop on Data Mining and Knowledge Discovery*, Montreal, Canada, June 2, 1996.
41. "Classifying Software Components Using Design Characteristics", with Wen-Syan Li, *Proceedings of the Tenth Knowledge-Based Software Engineering Conference*, November 12-15, 1995, Boston, Massachusetts.
42. "Experiences with SEMINT: Automated Semantic Integration", with Wen-Syan Li *Proceedings of the International Conference on Advanced Science and Technology*, March 25, 1995, Chicago, IL.
43. "Semantic Integration in Heterogeneous Databases using Neural Networks", with Wen-Syan Li, *20th International Conference on Very Large Data Bases*, Santiago, Chile, September 12-15 1994.

44. “Compiling Local Test Formulas for Global Constraints in Distributed Databases”, with Herts Chen, *International Conference on Intelligent Information Management Systems*, Washington, D.C., June 1-3 1994.
45. “Information Brokers: Sharing Knowledge in a Heterogeneous Distributed System”, with Daniel Barbará, *Proceedings of the 4th International Conference on Database and Expert Systems Applications*, Prague, Czechoslovakia, September 6-8, 1993.
46. “Using Field Specifications to Determine Attribute Equivalence in Heterogeneous Databases”, with Wen-Syan Li, *Proceedings of RIDE-IMS’93: Third International Workshop on Research Issues in Data Engineering*, Vienna, Austria, April 18-20, 1993.
47. “The Gold Mailer”, with Daniel Barbará, Hector Garcia-Molina, Ben Kao, Stephen Johnson, Sharad Mehrotra, Jens Tellefsen, and Rosemary Walsh, *Proceedings of the Ninth International Conference on Data Engineering*, Vienna, Austria, April 19-23, 1993.
48. “Distributed Processing of Filtering Queries in HyperFile”, with Hector Garcia-Molina, *Proceedings of the 11th International Conference on Distributed Computing Systems*, Arlington, TX, May 20-24 1991.
49. “Data Management with Massive Memory: A Summary”, with H. Garcia-Molina, R. Abbott, C. Staelin, and K. Salem, in *Parallel Database Systems*, Pierre America, Editor, PRISMA Workshop Proceedings, Noordwijk, The Netherlands, September 1990.
50. “Indexing in a Hypertext Database”, with Hector Garcia-Molina *Proceedings of the 1990 conference on Very Large Data Bases*, Brisbane, Australia, August 13-16 1990.
51. “The Design of a Document Database”, with Hector Garcia-Molina and Robert Hagmann, *ACM Conference on Document Processing Systems*, Santa Fe, NM, December 5-9, 1988.

Refereed System Demonstration Descriptions:

52. “GeoNODE: An End-to-End System from Research Components”, with John Griffith and Rod Holland, *The 17th International Conference on Data Engineering*, Heidelberg, Germany, April 2-6, 2001.
53. “Natural Language Processing for News Analysis”, with Tom Barteel, Stanley Boykin, Thad Cooper, Shereif El-Sheikh, Jerry Fitzgerald, Joe Francoeur, John Gibson, John Griffith, Steve Hansen, Lynette Hirschman, Rod Holland, Rob Hyland, Steve Janiak, Mark Maybury, Andy Merlino, Carsten Oertel, Marc Reichman, Marc Richards, Paul Silvey and Shane Steward, *Language Technology Joint Conference (ANLP-NAACL2000)*, April 29-May 40, 2000, Seattle, WA.
54. “Semint: A System Prototype for Semantic Integration in Heterogeneous Databases”, with Wen-Syan Li, *Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data*, May 23-25, 1995, San Jose, CA.

INVITED ARTICLES (editor-refereed)

1. Thorben Burghardt, Erik Buchmann, Klemens Böhm, and Chris Clifton, “Collaborative Search and User Privacy - How Can They Be Reconciled?”, invited full paper at *The 4th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom 2008)*, Orlando, Florida, November 13-16, 2008.
2. Chris Clifton, “What is Privacy? Critical Steps for Privacy-Preserving Data Mining”, invited article at the workshop on *Privacy and Security Aspects of Data Mining* in conjunction with *The Fifth IEEE International Conference on Data Mining*, Houston, Texas, November 27, 2005.
3. Chris Clifton, Murat Kantarcioglu, Jaideep Vaidya, Xiaodong Lin, and Michael Zhu, “Tools for Privacy Preserving Distributed Data Mining”, invited article in *ACM SIGKDD Explorations* 4(2):28-34, Johannes Gehrke, editor, December 2002.
4. Yücel Saygin, Vassilios S. Verykios, and Chris Clifton, “Using Unknowns to Prevent Discovery of Association Rules”, invited article in *ACM SIGMOD Record* 30(4):45-54 Special Section on Data Mining for Intrusion Detection and Threat Analysis, Daniel Barbará, editor, December 2001.

SELECTED PRESENTATIONS

1. “Plausibly Deniable Search”, invited talk at the *Symposium ‘In search of privacy’*, Radboud University, Nijmegen, The Netherlands, June 25, 2009.
2. “What an Information Technology Vendor Can (and Can’t) Do For Your Client”, invited presentation at the *Indiana Continuing Legal Education Forum on Privacy Law*, Indianapolis, IN, May 14, 2009.
3. “Measuring Privacy: A Risk-Based Approach”, talk at SAP Research CEC Karlsruhe, Germany, December 8, 2008.
4. “Opportunities for Private and Secure Machine Learning”, keynote at the *2008 AISec Workshop* held in conjunction with the *15th ACM Conference on Computer and Communications Security (CCS)*, Alexandria, VA, October 27, 2008. Revised as “Saleable Opportunities for Private and Secure Computation” for the Informatik-Kolloquium at Universit and FENS seminars at Sabanci University, December 5, 2008.
5. Panelist on “Technologies for Privacy-Protective Data Mining” at the Department of Homeland Security Privacy Office workshop on *Implementing Privacy Protections in Government Data Mining*, Washington, DC, July 24-25, 2008.
6. “Information metrics for privacy and anonymity in spatiotemporal data”, invited talk at the *First Interdisciplinary Workshop on Mobility, Data Mining and Privacy: Preserving anonymity in geographically referenced data*, Rome, Italy, February 14 2008.
7. “Is Privacy Still an Issue for Data Mining?”, invited talk at the *National Science Foundation Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation*, Baltimore, Maryland, October 10-12, 2007.
8. “Privacy and Anonymity in Text”, Max-Planck-Institut for Informatik, Saarbrucken, Germany, August 31, 2007.
9. “Will Privacy Kill Ubiquitous Knowledge Discovery?”, invited talk at the *KDUbiq Working Group on Privacy and Security* meeting, Berlin, Germany, September 19, 2006.
10. “Privacy Conflicts Inherent in Data Analysis”, Institute for Software Research International Seminar, Carnegie-Mellon University, March 29, 2006.
11. “What is Privacy? Critical Steps for Privacy-Preserving Data Mining”, Invited talk at the *Workshop on Privacy and Security Aspects of Data Mining* held in conjunction with the *IEEE International Conference on Data Mining*, November 27, 2005, New Orleans, Louisiana.
12. “Open Issues in Privacy-Preserving Data Mining”, Maebashi Institute of Technology, April 4, 2005.
13. Panel: “Data Mining and Data Aggregation” at the Symposium on *Privacy and Identity: The Promise and Perils of a Technological Age*, DePaul University, October 14-15, 2004.
14. Panel: “Data Aggregation, Mining, and Privacy” at the *NSF Cybertrust PI Meeting*, Pittsburgh, Pennsylvania August 18-20, 2004.
15. “Knowledge Projection for Tele-Maintenance”, Crane Naval Surface Warfare Center Distinguished Lecture Series, February 3, 2004.
16. “Turning Data into Knowledge Without Violating Privacy”, Joint Biostatistics & Medical Informatics / Statistics / Computer Sciences Seminar, University of Wisconsin, January 29, 2004.
17. “Privacy-Enhanced Data Management for Next-Generation e-Commerce”, with Iriini Fundulaki, Richard Hull, Bharat Kumar, Daniel Lieuwen, and Arnaud Sahuguet, tutorial given at the *29th International Conference on Very Large Data Bases*, September 9-12, 2003, Berlin, Germany.
18. “Privacy Preserving Data Mining”, tutorial given at *The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 24-27, 2003, Washington, D.C.
19. “Privacy Preservation to Enable Data Mining Across Enterprises”, invited talk at the *IBM Almaden Institute Symposium on Privacy*, April 9-11, 2003, San Jose, California.

20. "Privacy Preserving Distributed Data Mining", data mining seminar at the University of Illinois, March 19, 2003.
21. "Defining Privacy for Data Mining", invited talk at the *National Science Foundation Workshop on Next Generation Data Mining*, November 1-3, 2002, Baltimore, MD.
22. "Privacy, Security, and Data Mining", tutorial given at the combined conference *13th European Conference on Machine Learning (ECML'02)* and *6th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'02)*, Helsinki, Finland, August 19, 2002.
23. "Developing Custom Intrusion Detection Filters Using Data Mining", Purdue Center for Education and Research in Information Assurance, December 10, 1999.
24. "Data Mining at MITRE" *Third Applied Database Technology Conference*, The MITRE Corporation, Bedford, MA, June 5 1997.
25. "Data Mining in Text", *Symposium on Advanced Information Processing and Analysis*, Tysons Corner, Virginia, March 26, 1997.
26. "Neural Networks for Semantic Integration" *Second Applied Database Technology Conference*, The MITRE Corporation, Bedford, MA, December 7, 1995.
27. "Component Based Software: How Database Technology Can Help", AT&T Bell Labs, Naperville, IL, June 3, 1994.
28. "Query Processing in HyperFile", New Jersey Institute of Technology, May 11, 1992.

SOFTWARE DEVELOPED

GeoNODE (Geographic News On Demand Environment), developed while at the MITRE Corporation, combines research results from natural language, data mining, multimedia, and visualization into a system for analyzing news. News, in the context of GeoNODE, is anything from television broadcasts to focused email and newsgroups. GeoNODE is in use as a concept evaluation prototype at several sites within the Department of Defense. Responsible for database development, as well as the text mining research component. See the publications list for more details.

RESEARCH FUNDING

1. Tim Finin (PI, UMBC) et al. (19 co-PIs, 4 Purdue), *A Framework for Managing the Assured Information Sharing Lifecycle*, DoD Multidisciplinary University Research Initiative, 5/1/2008-4/30-/2013, \$4.35M base + 3M option years (\$1.5M Purdue).
2. Elisa Bertino (PI) and Chris Clifton, *Identity Management and Trust Negotiation Services*, Department of Homeland Security / Institute for Information Infrastructure Protection, 4/1/2007-3/31/2009, \$300,000.
3. David Ebert (Director), Alok Chaturvedi, William Cleveland, Chris Clifton, Ahmed Elmagarmid, and Marc Overhage, *Purdue University Regional Visualization and Analytics Center*, Department of Homeland Security, January 2006-December 2008, \$517,000.
4. Ann Hendrich (Ascension Health), Marilyn Chow (Kaiser Permanente), Nelson Lee (Rapid Modeling), William Cleveland, Chris Clifton, Jason Abrevaya, *A Multi-Site Study of How Medical Surgical Nurses Spend Their Time: A Baseline Study in Preparation for an Electronic Health Record and an Evidenced-based Nursing Unit Design*, Robert Wood Johnson Foundation, April 2005-August 2006. (Purdue CS portion \$91,269).
5. Chris Clifton, AnHai Doan (University of Illinois), Ahmed Elmagarmid, Gunther Schadow (Regenstrief Institute for Healthcare), and Dan Suci (University of Washington), *Privacy-Preserving Data Integration and Sharing*, NSF ITR, September 2004-August 2008, \$1,012,000.
6. Clement J. McDonald (PI) et al., Chris Clifton (Purdue PI) *A Center of Excellence in Medical Informatics to Provide an Advanced Infrastructure for Human Research: A Catalyst for Indiana Research*, Indiana 21st Century Fund, August 2004-August 2006, \$3,832,196 (Lead: Regenstrief Institute for Healthcare; Purdue

portion \$50,000.)

7. Chris Clifton and Mikhail Atallah, *Collaborative Research: Distributed Data Mining to Protect Information Privacy*, NSF Small ITR, August 2003 - August 2006, \$276,274. (Collaborative with Wenliang Du, Syracuse University - funded under separate award.)
8. Chris Clifton, Walid Aref, Ahmed Elmagarmid, Arif Ghafoor, and Sunil Prabhakar, *Knowledge Projection for TeleMaintenance*, EG&G / Naval Surface Warfare Center, Crane Division, October 21, 2002 - September 30, 2005, \$1,746,813.
9. Chris Clifton, Ananth Iyer, and Reha Uzsoy, *A Prototype Integrated Transaction Data Analysis and Visualization Environment for the Transportation, Distribution and Logistics Sector*, e-Enterprise Center at Discovery Park, August 19, 2002 - May 10, 2003, \$34,278.
10. Chris Clifton, Pascal Meunier, and Gustavo Rodriguez-Rivera, *Secure Programming: Add-on Course Module and Short Course*, e-Enterprise Center at Discovery Park, August 19, 2002 - June 30, 2003, \$19,344.
11. *Privacy Preserving Data Mining*, Purdue Research Foundation, August 2002 through July 2004, \$26,403.
12. *Data Mining for Intrusion Detection*, U.S. Army CECOM Space and Terrestrial Communications Directorate, June 1999 through September 2001, ~\$360,000.
13. *Data Mining in Text*, Co-PI with Arnon Rosenthal and Jeffrey Ullman (Stanford University), Community Management Staff (Intelligence Community), September 1996 through June 2000, \$589,913 (MITRE's portion \$442,174).
14. *Data Mining in Text and Images*, MITRE Sponsored Research, January 1996 through September 1999, \$521,421.
15. *Data Mining Security*, National Security Agency, October 1995 through September 1997, ~\$120,000.
16. *Shared Software Database*, NSF Research Initiation Award, September 1992 through August 1995, \$90,000.

PhD AND MS STUDENTS SUPERVISED

Current Students

1. Mummoorthy Murugesan, privacy and anonymity issues in text data, funded on Ascension Healthcare / RCHE grant from Summer 2005, NSF Spring/Summer 2007, DHS/I3P Fall 2007, AF MURI 2008-2009, Ph.D. expected 2009.

Completed

2. Mehmet Ercan Nergiz, Dissertation *Anonymization-Based Privacy Protection*, Ph.D. December 2008. Currently Postdoctoral researcher at Sabanci University.
3. Wei Jiang, Dissertation *Incentive Driven and Privacy-Preserving Collaborative Computing*, Ph.D. August 2008. Currently at the Missouri University of Science & Technology.
4. Yang Wang, correlation of high-dimensional data with time and motion study data, MS 2007. Currently at Oracle Corp.
5. Murat Kantarcioglu, Dissertation *Privacy Preserving Data Mining in Horizontally Partitioned Data*, Ph.D. August 2005. Currently at University of Texas at Dallas.
6. Jaideep S. Vaidya, Dissertation *Privacy Preserving Data Mining for Vertically Partitioned Data*, Ph.D. August 2004. Honorable Mention, Best Paper Competition at *The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* for "Privacy-Preserving K-Means Clustering over Vertically Partitioned Data". Currently at Rutgers University.
7. Siddhartha K. Goel, thesis *Derived Access Control Specification for XML*, MS August 2003. Currently pursuing Masters of Management at Indian Institute of Management, Calcutta.
8. Wen-Syan Li, Dissertation *Semantic Integration in Heterogeneous Databases*, PhD December 1995. *Using*

Field Specifications to Determine Attribute Equivalence in Heterogeneous Databases, MS December 1992. Student paper award for presentation of “Experiences with SEMINT: Automated Semantic Integration” (see publications). Currently at IBM Almaden Research.

9. Herts T. Chen, Dissertation *Concurrency Control Using Semantic Knowledge: A Proof Theoretic Basis*, PhD December 1994.
10. Robert Haddleton, *An Analysis of an Active Database as Applied to a Real-Time Application*, MS June 1994.
11. Shashidar Jasthi, *Application of Software Quality Models to Individual Evaluation*, MS December 1993.
12. Christopher Fernandes, *Determining Semantic Equivalence Using a Nucleotide Base Genetic Sequencing Algorithm*, MS December 1993.
13. Robert Milfajt, *A Comparison of Two Heterogeneous Database Schema Integration Methods*, MS June 1992.
14. Stephen Gene Jacobson, *Integrating DB Design and CASE Tools*, MS June 1992.

TEACHING HIGHLIGHTS

Purdue University

August 2001 to Present

CS62600	Advanced Information Assurance	Fall 2009
CS54200	Distributed Database Systems	Spring 2009
CS250	Computer Architecture	Fall 2005, Spring 2007
CS590D	Data Mining	Spring 2005, 2006
CS526	Information Security	Spring & Fall 2003, Fall 2004
CS490D	Introduction to Data Mining	Spring 2004
CSx97	Honors Program	2007-2008
CS390S	Secure Programming (mini-course with Drs. Meunier, Rodriguez-Rivera)	Fall 2002
CS541	Database Systems	Fall 2002, Fall 2007
CS603	Advanced Topics in Distributed Systems	Spring 2002
CS590M	Security Issues in Data Mining	Fall 2001

Independent Study Projects Supervised

Ryan Adams	CS497: Secure Similar Document Detection using KLDivergence	Spring 2008
John Horst	CS497: Secure Content Validation	Spring 2007
Brett Banich	CS490: Data Mining of Graphs	Summer 2003
Y.B. Lee	CS590: Information Security Certifications	Summer 2003
Hong Liang	CS590: Advanced Data Mining	Summer 2003
Kris Reyes	CS490: Exploit Analysis (Co-advised with Dr. Meunier)	Fall 2002
Wonjun Lee	CS590: Data Mining for Intrusion Detection	Summer 2002
Hong Liang	CS590: Multivariate Data Mining using PCA	Summer 2002
Mashrur Nabi and Halim Sandikin	CS490: Exploit Database (Co-advised with Dr. Meunier)	Spring 2002

Northwestern University

September 1991 to August 1995

Developed and taught a two quarter sequence *Software Project Management and Development*. Course included innovative collaboration with local industry as source of projects.

Taught *Fundamentals of Computer System Software, Introduction to Database Systems, Seminar: Heterogeneous Databases.*

Raritan Valley Community College

January through May 1989

Taught *Foundations of Computer Science.*

COMMITTEE MEMBERSHIPS

Purdue Departmental Committees

Undergraduate Committee	2003-2004, Chair 2007-Present
Hiring Committee	2006-2007
COALESCE Search	2005-2007
Graduate Committee	2001-2003, Chair 2004-2006
Graduate Admissions Committee	2003-2004
Strategic Planning Committee	2001-2002

Northwestern University

School of Engineering Academic Standing Committee	1992-1994
Departmental Computer Committee	1992-1994
Departmental Graduate Committee	1991-1992
Departmental Graduate Admissions Committee	1991-1992

PROFESSIONAL ACTIVITIES

Editorial Boards

Editorial board, *Journal of Data Management Research* (the journal of the VLDB conference), VLDB Endowment, 2008 - present.

Editorial board, *IEEE Transactions on Knowledge and Data Engineering*, IEEE Computer Society, Los Alamitos, CA, 2003 - 2007. Survey editor, 2007-present.

Editorial board, *Knowledge and Information Systems*, Springer-Verlag, London, 2000 - present.

Editorial Board, *Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies*, Kluwer Academic Publishers, The Netherlands, 2008-present (Review Board 2000 - 2007).

Editorial Board, *Transactions on Data Privacy*, Barcelona, Spain, 2008 - present.

Editorial Board, *Journal of Privacy Technology*, Pittsburgh, PA, 2004 - present.

Guest editor, *The Edge Data Mining Issue*, The MITRE Corporation, April 2000.

Funding Review Panels

Proposal review panel, National Science Foundation Division of Information and Intelligent Systems, 1999, 2003, 2004, 2007, 2008.

Proposal review panel, Texas Higher Education Board, 1993.

Conference and Workshop Leadership

Senior program committee, *The 16th ACM SIGKDD International Conference on Knowledge Discovery*

and Data Mining, July 25 - 28, Washington, D.C.

Senior program committee, *The 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, June 28 - July 1, 2009, Paris, France.

Area chair for Privacy and Social Factors, *8th SIAM International Conference on Data Mining (SDM)*, April 24-26, 2008, Atlanta, GA.

General Co-Chair, *The Seventh IEEE International Conference on Data Mining*, October 28-31, 2007, Omaha, NE.

Senior program committee, *The 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 12-15, 2007, San Jose, CA.

Program Committee Co-Chair, *The Sixth IEEE International Conference on Data Mining*, December 18 - 22, 2006, Hong Kong.

Treasurer, *The Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 21-24, 2005, Chicago, Illinois.

Program Committee Vice Chair, *The Third IEEE International Conference on Data Mining*, November 19 - 22, 2003, Melbourne, Florida, USA. Also co-organized associated workshop on *Privacy-Preserving Data Mining*.

Organizer and chair, workshop on “Privacy, Security, and Data Mining” held in conjunction with *The 2002 IEEE International Conference on Data Mining (ICDM '02)*.

Tutorials chair and member of program committee, *The 2001 IEEE International Conference on Data Mining (ICDM '01)*.

Local Arrangements Chair, Fourth International Conference on Foundations of Data Organization and Algorithms, October 13-15, 1993.

Program Committees

The 36th International Conference on Very Large Data Bases (VLDB 2010), September 13 - 17, 2010, Singapore.

The 26th International Conference on Data Engineering (ICDE 2010), March 1-6, 2010, Long Beach, California.

IEEE International Conference on Data Mining series (ICDM 2009), December 6-9, 2009, Miami, Florida.

The 2009 International Workshop on Privacy Aspects of Data Mining held in conjunction with ICDM'09, December 6, 2009, Miami, Florida.

The 2009 IEEE International Conference on Information Privacy, Security, Risk and Trust, August 29-13, Vancouver, Canada.

The 23rd Annual IFIP WG 11.3 Working Conference on Data and Applications Security, July 12-15, 2009, Montreal, Canada.

International Workshop on Privacy-Preserving Data Analysis (PPDA), April 20, 2009, Brisbane, Australia.

25th International Conference on Data Engineering, March 29-April 4, 2009, Shanghai, China.

The 2008 IEEE International Conference on Data Mining, December 15-19, 2008, Pisa, Italy.

3rd Workshop on Secure Knowledge Management, November 3-4, 2008, Dallas, Texas.

European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2008), September 15-19, 2008, Antwerp, Belgium.

34th International Conference on Very Large Databases (VLDB'08), August 25-28, 2008, Auckland, New Zealand.

2nd ACM SIGKDD International Workshop on Privacy, Security, and Trust in KDD (PinKDD'08), August

24-27, 2008, Las Vegas, Nevada.

The 27th ACM SIGMOD International Conference on Management of Data, June 9-12, 2008, Vancouver, Canada.

The 24th International Conference on Data Engineering (ICDE 2008), April 7-12, 2008, Cancun, Mexico.

18th International Conference on Database and Expert Systems Applications (DEXA '07), September 3-7, Regensburg, Germany.

The 20th Annual IFIP WG 11.3 Working Conference on Data and Applications Security, July 31-August 2, 2006, Sophia Antipolis, France.

The 25th ACM SIGMOD International Conference on Management of Data, June 26-29, 2006, Chicago, Illinois.

The 22nd International Conference on Data Engineering (ICDE 2006), April 3-7, 2006, Atlanta, GA.

The Fifth IEEE International Conference on Data Mining, November 27-30, 2005, New Orleans, Louisiana.

2005 SIAM International Conference on Data Mining, April 21-23, 2005, Newport Beach, California, USA.

International Workshop on Privacy Data Management held in conjunction with *The 21st International Conference on Data Engineering (ICDE 2005)*, April 8-9, 2005, Tokyo, Japan.

The Fourth IEEE International Conference on Data Mining, November 1-4, 2004, Brighton, United Kingdom. Also served on program committee of the associated *Workshop on Privacy and Security Aspects of Data Mining* and *Workshop on Data Mining and the Grid*.

The 8th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD), September 20-24, 2004, Pisa, Italy. Also served on program committee of the associated workshop *International Workshop on Privacy and Security Issues in Data Mining*.

The Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 22-25, 2004, Seattle, Washington.

2004 SIAM International Conference on Data Mining, April 22-24, 2004, Lake Buena Vista, Florida, USA.

The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 24-27, 2003, Washington, D.C. Also served on program committee for associated workshop *Data Cleaning, Record Linkage, and Object Consolidation*, and on panel *Privacy and Data Mining: Friends or Foes?*

Workshop on Mining Data Across Multiple Customer Touchpoints for CRM (in conjunction with PAKDD-02), May 6, 2002.

Workshop on Data Mining for Security Applications held in conjunction with the 8th ACM Conference on Computer Security, November 8, 2001.

The 2001 Symposium on Applications and the Internet (SAINT-2001).

Fifth IFCIS International Conference on Cooperative Information Systems (CoopIS'2000).

1999 IEEE Knowledge and Data Engineering Exchange Workshop (KDEX-99). Served on panel *Convergence of Technologies on the Web: the Case of Distributed Artificial Intelligence, Databases and Information Retrieval*, chaired session on *Data Mining*.

1999 IEEE Computer Systems and Applications Conference (COMPSAC-99).

ACM Digital Libraries Conference Workshop on Organizing Web Space (WOWS), August 14, 1999.

Planning committee, 1999 DARPA-sponsored workshop on novelty detection in documents.

1998 IEEE Knowledge and Data Engineering Exchange Workshop (KDEX-98).

Sixth International Workshop on Research Issues on Data Engineering: Interoperability of Nontraditional

Database Systems. Chaired session on “Applications”.

1995 International Conference on Distributed Computing Systems. Organized and chaired panel “Fee-for-Service in Distributed Systems: Research and Policy Issues”.

Invited attendance, NSF Workshop on Hyperbase Systems, October 15-16, 1992, Washington DC.

Senior member, ACM and IEEE.