Workshop on Privacy, Security, and Data Mining
How do we mine data when we aren't allowed to see it?

To be held in conjunction with *The 2002 IEEE International Conference on Data Mining.*

Maebashi TERRSA, Maebashi City, Japan, December 9, 2002

http://www.cs.purdue.edu/~clifton/psdm.html

In the light of developments in technology to analyze personal data, public concerns regarding privacy are rising. While some believe that statistical and Knowledge Discovery and Data Mining (KDDM) research is detached from this issue, we can certainly see that the debate is gaining momentum as KDDM and statistical tools are more widely adopted by public and private organizations hosting large databases of personal records. One of the key requirements of a data mining project is *access* to the relevant data. Privacy and Security concerns can constrain such access, threatening to derail data mining projects. The purpose of this workshop is to discuss these issues and promote achievements of researchers in the area. We want to bring together experts, including both researchers and practitioners, in privacy, data mining and its applications, and statistical database security.

Background

There are many data mining situations where these privacy and security issues arise. A few examples are:

- Identifying public health problem outbreaks (e.g., epidemics, biological warfare instances). There are many data collectors (insurance companies, HMOs, public health agencies). Individual privacy concerns limit the willingness of the data custodians to share data, even with government agencies such as the U.S. Centers for Disease Control. Can we accomplish the desired results while still preserving privacy of individual entities?

- Collaborative corporations or entities. Ford and Firestone shared a problem with a jointly produced product: Ford Explorers with Firestone tires. Ford and Firestone may have been able to use association rule techniques to detect problems earlier. This would have required extensive data sharing. Factors such as trade secrets and agreements with other manufacturers stand in the way of the necessary sharing. Could we obtain the same results, while still preserving the secrecy of each side's data?

  Government entities face similar problems, such as limitations on sharing between law enforcement, intelligence agencies, and tax collection.

- Multi-national corporations. An individual country's legal system may prevent sharing of customer data between a subsidiary and its parent.

Workshop Content and Format

This will be a full day workshop, opening with a presentation by an invited speaker to set the stage. The rest of the day will consist of paper sessions with ample time for questions and breaks for discussion. The goal is to bring participants up to speed on the issues and solutions in this area, outline key research problems, and encourage collaborations to address these problems.

To encourage quality submissions, the proceedings will be published as Volume 14 of the *Conferences in Research and Practice in Information Technology* series. In addition, the best paper from the workshop will compete with regular papers from the International Conference on Data Mining to have extended versions considered for possible publication in the *Journal of Knowledge and Information Systems.*

Topics of Interest

Papers are solicited that identify and propose technical solutions to such problems. Sample topics (by no means an exhaustive list) include:

- Privacy and security policies and their implications on data mining, including issues of data collection and ownership.
- Learning from perturbed / obscured data.
- Techniques for protecting confidentiality of sensitive information, including work on statistical databases, and obscuring or restricting data access to prevent violation of privacy and security policies.
- Learning from distributed data sets with limits on sharing of information.
- Algorithms for balancing privacy and knowledge discovery in data mining.
- Use of data mining results to reconstruct private information, and corporate security in the face of analysis by KDDM and statistical tools of public data by competitors.
- Case studies of security and privacy policies and their impact on data mining, e.g., privacy issues in medical databases or analysis of personal records for customer relationship management.
- Controversial applications of Knowledge Discovery and Data Mining, including secondary use of personal data, fraud detection, credit record checking, knowledge discovery of competitors' (suppliers') strengths by transaction analysis.

Attendance

Attendance is not limited to the paper authors. We strongly encourage other interested parties to attend the workshop. One of the objectives of the workshop is to promote the interaction among researchers and those who have experienced security and privacy constraints on data mining.

Submission Guidelines

Papers should be at most 12 pages long in single-column format, 12-point font, with at least 1-inch margins on all sides. Please submit electronically (PDF or PostScript preferred, ask the chairs for help with other formats) to clifton@cs.purdue.edu on or before September 23, 2002.

Important Dates

Intent to submit	*September 1, 2002*
(appreciated, but not required)	
Submission Deadline:	September 23, 2002
Acceptance Notification:	October 7, 2002
Camera-ready Copies:	October 21, 2002
Workshop date:	December 9, 2002

Program Committee

Chairs

Ljiljana Brankovic, University of Newcastle
Department of Computer Science and Software Engineering
Callaghan, New South Wales, 2308 Australia
+61-4921-6054, Fax: +61-4921-6929
lbrankov@cs.newcastle.edu.au
http://www.cs.newcastle.edu.au/~lbrankov/

Chris Clifton, Purdue University
Department of Computer Sciences
West Lafayette, Indiana 47907-1398 USA
+1 765-494-6005, Fax: +1 765 494-0739
clifton@cs.purdue.edu
http://www.cs.purdue.edu/people/clifton

Vladimir Estivill-Castro, Griffith University
School of Computing and Information Technology
Nathan and Logan Campuses
Brisbane 4111, Queensland, Australia
+61-7-3875-5402, Fac: +61-7-3875-5051
v.estivill-castro@cit.gu.edu.au
http://www.cit.gu.edu.au/~s2130677

Daniel Barbará, George Mason University
Stephen Fienberg, Carnegie-Mellon University
Johannes Gehrke, Cornell University
Franco Malvestuto, University of Rome
Benny Pinkas, DIMACS
John Roddick, Flinders University
Jozef Siran, Slovak University of Technology
Ramakrishnan Srikant, IBM Research