PORTIA: Privacy, Obligations, and Rights in Technologies of Information Assessment

Dan Boneh
Computer Science Department
Stanford University
Stanford, CA 94350 USA
dabo@cs.stanford.edu

Joan Feigenbaum and Avi Silberschatz
Computer Science Department
Yale University
New Haven, CT 06520 USA
{jf,avi}@cs.yale.edu

Rebecca N. Wright
Computer Science Department
Stevens Institute of Technology
Hoboken, NJ 07030 USA
rwright@cs.stevens-tech.edu

Abstract

Increasing use of computers and networks in business, government, recreation, and almost all aspects of daily life has led to a proliferation of sensitive data about people and organizations. Without proper precautions, these sensitive data can be misused, misinterpreted, or mismanaged. The PORTIA project aims to develop a comprehensive, end-to-end technological infrastructure for handling sensitive data over the entire course of their lifetime.

1 Introduction

Increasing use of computers and networks in business, government, recreation, and almost all aspects of daily life has led to a proliferation of sensitive data about people and organizations. By *sensitive data*, we mean data that, if used improperly, can harm data subjects, data owners, data users, or other relevant parties. These data are stored by a multiplicity of entities, ranging from individuals to small businesses to large government agencies, and concern about the ownership, control, privacy, and accuracy of these data has become a top priority in technical, academic, business, and political circles. Social trends ranging from a fluid and unpredictable business climate (that leads to unanticipated uses and exchanges of sensitive data) to a homeland-security-focused political climate (that leads to increased use of electronic surveillance technologies) make it likely that the creation and collection of massive amounts of sensitive data and the attendant nervousness about whether they are being handled properly will both increase as time goes on. Already, large population data banks store information that many are uncomfortable with [18]. The proposed "Total Information Awareness" initiative [32] caused widespread concern about potential erosion of data subjects' rights, and thus the high-tech sector anticipates

Bulletin of the IEEE Computer Society Technical Committee on Data Engineering

growing demand for surveillance technology with some built-in privacy protections [24]. Clearly, there is an urgent need for appropriate technology and policy for sensitive-data handling.

One often hears that individuals must give up some privacy or convenience so that American society as a whole can benefit from greater security. This vague and unproven assertion is used to justify the growing number of venues in which one is asked to "provide a government-issued photo ID," the thriving background-check industry, the fingerprinting of foreign visitors to the US, and the ubiquity of surveillance cameras in public places. Yet, there is no official national-ID system with provable security properties, publicly available data feeds used in low-cost background checks are error-prone, some foreign visitors are exempt from the fingerprinting requirement, and there are neither efficient procedures for mining massive sets of images nor commonly agreed-upon social norms for the appropriate use of surveillance data. Without attention to these issues, one can imagine a world in which the misuse of sensitive data continues to grow more prevalent, individuals' civil rights are routinely violated for the sake of "security measures" that may or may not actually provide the desired security, and poor data quality results in inconvenience and worse.

Some basic technological tools exist for finding critical information and patterns in a sea of disparate data feeds, for storing huge amounts of data and accessing it efficiently for certain mission-critical purposes, and for hiding or safeguarding some private information while still accomplishing tasks such as web-shopping on a large-scale public network. Yet, there is no comprehensive, end-to-end technological and public policy infrastructure for handling sensitive data over the entire course of their lifetime.

The PORTIA project [26] is exploring the design and development of such an infrastructure. Rejecting the overly simplistic "security vs. privacy" tradeoff for sensitive data, PORTIA addresses the need for deeper understanding of both the rights and the responsibilities of a number of parties, including the data subjects, the data owners, and the data users. The difficult, but necessary, goal is to balance apparently conflicting objectives such as privacy, anonymity, authenticity and integrity, appropriate-use policies, and auditability.

In this paper, we provide brief overviews of three major themes of the PORTIA project: privacy-preserving data mining, database-policy enforcement tools, and identity theft and identity protection.

Privacy-preserving data mining seeks to satisfy the desires to disclose or discover some information while protecting the privacy of other information. For example, is it possible to satisfy law-enforcement requirements for extensive mining of diverse databases in a way that does not compromise individuals' rights? We describe our vision for privacy-preserving data mining in Section 2.

Database-policy enforcement tools can allow specification and enforcement of distributed access-control policies and integrate them with existing database systems. An important goal is to control the exposure of *derived sensitive data* that can be deduced by aggregating information from multiple agents. We describe our goals for database-policy enforcement tools in Section 3.

Identity theft is the fastest growing crime in the US [30]. Often identity theft is done by impersonating a user's digital identity on a web service such as e-Bay. Can one develop techniques that make such impersonation harder to accomplish? Challenges include preventing web-spoofing attacks designed for stealing identities, defending against fraudulent spam email that targets naive users, and protecting user identities by keeping them private. We describe our work on preventing identity theft and protecting identity privacy in Section 4.

2 Privacy-preserving Data Mining

The lives of people and organizations typically involve contacts with multiple entities, including individuals, businesses, government bodies, and law-enforcement agencies. Traces of these contacts are left in the form of electronic data records, collected for *e.g.*, billing and service in the case of commercial parties or census and taxation in case of the government. Consistent advances in databases and data mining, storage, and networking technologies, as well as a rapid growth in online data collection, has led to increasing concerns about privacy protection for individual and institutional data.

Government and commercial organizations already store, manage, and mine their information. This is the basis for the database industry that has matured over the past three decades, supported by a formal theory of data management and robust commercial systems that implement it. As concern for homeland security has grown, data-mining needs have become more distributed and more urgent. Privacy-preserving data mining can act as an enabling technology by respecting the needs of individuals, corporations, and government agencies to protect certain information, while allowing the disclosure or discovery of other information. Adding privacy protection to the list of requirements for data-management systems is desirable for at least three reasons:

- **Protection of innocent bystanders**: One should not sacrifice the privacy of individuals beyond what is necessary. For example, answering the security-relevant question of whether a *targeted* person has been at a particular location should not require surveillance systems to reveal the identities of *everyone* who has been at that location.

- **Protection of sensitive information**: Pairs of organizations, *e.g.*, airlines and law-enforcement agencies or hospitals and the Centers for Disease Control, should be able to find common elements of their databases without revealing the entire databases to each other or even revealing the specific potential matches.

- **Collaboration among disparate agencies**: As has been well documented, different federal and local agencies do not always cooperate to the degree necessary to provide the highest security. In principle, such agencies should be able to use cryptographic protocols to determine security-relevant outcomes based on their joint data, without requiring any agency to reveal its data to the others or to a trusted third party. For example, photographic databases owned by two different agencies should be matchable in such a way that neither agency needs to trust the other with all of its images.

At this point, it is appropriate to ask whether such strong requirements could ever be satisfied. Isn't the phrase "privacy-preserving data mining" (or, *a fortiori*, "privacy-preserving surveillance") an oxymoron? In fact, the cryptographic-research community has, over almost three decades, developed tools that are extremely (almost paradoxically) powerful. Computing exactly one relevant fact about a distributed data set while concealing everything else about it is precisely what cryptographic theory enables *in principle* [35, 36, 4, 8]. Government agencies should be able to use these ideas to find security-critical matches in their databases without revealing the databases to each other. Similarly, medical practioners and researchers should be able to conduct their necessary activities while protecting the privacy of the individuals involved and complying with relevant legislation such as HIPAA [17].

More generally, researchers have developed vast bodies of cryptographic techniques (*e.g.*, [23, 20, 7, 14, 6, 19]), data-perturbation and data-sanitization techniques (*e.g.*, [3, 2, 28, 31, 34, 19, 11, 12]), and policy-specification and policy-evaluation techniques (*e.g.*, [5, 13, 21, 22, 27]) that should be useful in balancing homeland-security and other goals with individual rights.

A central focus of PORTIA is the further development of techniques that are simultaneously privacy-preserving, computationally efficient, and practical. Specific agenda items in this area include:

- Develop algorithms and protocols for privacy-protecting distributed data mining, in particular for security and surveillance applications. Consider both structured and unstructured data, simple database queries, data mining, and integration of disparate public and private databases. In many cases, this will require both advances in the state of the art of data mining and new privacy-preserving data transformations. One wide-open technical area is the algorithmics of *structural* data mining, *i.e.*, the problem of discovering ordering structures from unordered data.

- Formulate new technical definitions of privacy that are weaker than those in the existing cryptographic literature (and hence may be achievable by more efficient algorithms) but that still provide meaningful and quantifiable protection.

- Fully explore the principled use of approximation as a resource. Traditionally, approximation algorithms have been viewed as a necessary evil—substitutes for exact computations that are simply too expensive. However, recent work has shown that approximation can be used to achieve privacy preservation as well as computational efficiency [14]. PORTIA goals include the study approximation in privacy-preserving use of surveillance data, census data, *etc.*

- Develop techniques for privacy-preserving data cleaning. Most databases and data warehouses suffer from the problem that data received from external sources is likely to contain errors, *e.g.*, those introduced by noisy data collection, data-entry mistakes, missing fields, and inconsistent conventions across data sources. Thus, a lot of effort goes into *data cleaning*, the task of detecting and correcting errors in data. The PORTIA project will develop and rigorously analyze techniques for validating and correcting input data before loading, with the goal of using this cleaning stage of database construction and update as an opportunity for privacy-preserving transformations.

- Integrate the technical approaches of privacy policies and privacy-preserving data mining. Ideally, each data holder would have a machine-readable privacy policy, each data-mining system would have a machine-readable specification of what it does and does not reveal about a data set, and a policy-compliance-checking component could give data holders the information they need in order to decide whether they are willing to feed data to or receive data from the system. Detailed evaluation of mismatches between privacy policies and ostensibly "privacy-preserving" data-mining algorithms could inform the development of better algorithmic tradeoffs between privacy and information-discovery.

- Explore the social implications of proposed technological solutions. Consider both the goals of upholding current social standards and enabling new standards, particularly if the new standards might have been desirable in the face-to-face world but weren't technologically possible in that world.

- Explore the usefulness of "trusted-computing platforms," such as those under development by the TCPA [33] and by Microsoft's Next-Generation Secure-Computing Base program (formerly Palladium) [25], in privacy-preserving data transformations. Trusted platforms enable a machine to attest to a remote peer that it is running specific executable code. Thus, trusted platforms can potentially be used to implement "trusted third parties" to whom an organization can outsource data mining and data cleaning. The organization is assured that its data will never become available in the clear, because the data mining machine attested to the fact that its software deletes the data after mining it [16].

Recent progress by PORTIA participants in this area includes new protocols for computing the k^{th}-largest element of the union of confidential data sets in the multiparty case [1] and for computing the intersection in the two-party case [15].

3 Database-Policy Enforcement Tools

Much of the ongoing research in databases focuses on the challenges of making them more efficient, functional, and reliable. Database security tends to focus on access-control mechanisms that allow one to state explicitly what types of operations the various users can invoke on the relations and views constituting the database. PORTIA's goal is to complement these efforts by taking a user-centric and data-centric approach to database security. Access policies and queries tend to be quite complex in database systems, and methods for enforcing them cannot interfere with stringent performance requirements.

- **Distributed Access Control via Complex Policy.** An important part of developing a comprehensive, end-to-end technological infrastructure for handling sensitive data involves defining, managing, and enforcing information-access policies. Any organization dealing with sensitive data must develop a *published* policy governing release of information that may compromise an individual's privacy. The success of companies such as Oblix, Securant, and Netegrity illustrates the growing trend of formulating complex policies and making decisions based on them. The development of XrML, an XML-based language for expressing rights policies, also illustrates a growing interest in complex policies and their use. PORTIA goals in this area include:

  - Investigate the use of trust management [5] and other authorization frameworks in the specification of enterprise-wide information-disclosure policies. Enable the specification of policies that depend on the enterprise, the individual and organizational data users, and the data subject(s). Investigate the tradeoffs that may arise between support for these policies and the very fast access to data that today's database systems are designed to provide.
  - Develop infrastructure that supports distributed use of the policy framework by multiple, heterogeneous organizations. This infrastructure may include software supporting key management and policy exchange.
  - Develop methods and tools for policy development, testing, and maintenance. For example, the author of a policy that depends on other organizations may wish to test the consequences of her policy and examine ways in which extrinsic changes affect information flow.

- **Policy enforcement.** Can one build a *trusted data-management service* (TDMS) to enforce policies specified as above? One of the main challenges is to control the exposure of *derived data*. Derived data are obtained by transforming, copying, or aggregating data obtained from multiple agents. If one ignores the issue of derived data, as most commercial DBMSs do today, then providing a trusted data service is much simpler. The PORTIA project will address this issue explicitly and pursue two implementation approaches:

  - Wrapper-based: One can start with a conventional DBMS that provides data robustness and some fixed security and privacy mechanisms and then add a wrapper that intercepts all interactions between the DBMS and the outside world (including other wrappers). Each wrapper must be able to describe its policies to its peer wrappers, so that they can share some of the data. Initially, all programs can be required to access data through the wrapper, so that policies that are not implemented by the underlying DBMS can be enforced by the wrapper. Eventually, the wrapper will allow trusted and confined programs to run directly on the DBMS.
  - Native TDMS: Starting with an open-source DBMS (like Postgress), one can extend it to provide TDMS functionality natively. This approach is more efficient, because security and privacy rules can be enforced by the DBMS directly. Also, if the DBMS has exclusive control over the storage disks, then it is possible to track how data are modified and to confine derived database values. One important question to be addressed is whether enforcement is at the lowest level (*e.g.*, the record interface), or at a higher level (*e.g.*, the SQL level), or both.

4 Identity Theft and Identity Protection

Identity theft is the fastest growing crime in the US [30]. Identity thieves use a number of techniques to steal sensitive information in the digital world:

- Using network-based attacks, criminals break into databases and extract sufficient customer information to impersonate those customers. Online merchants and credit-card processors frequently come under such attacks, resulting in the theft of personal data about millions of customers.

- By creating replicas (a.k.a. "spoofs") of legitimate web sites, criminals trick honest customers into revealing their passwords and other identifying information. Recent spoofs of the e-Bay web site resulted in thousands of user passwords being exposed. Spoofers can then masquerade as those users on e-Bay and deceive honest e-Bay users and merchants.

- Criminals are well aware that users tend to use the same password at many different sites. Password thieves often break into the user database of a low-security site (say, a high-school reunion site) and then try all exposed username/password pairs at e-commerce sites such as e-Bay. As a result, security at an e-commerce site can be nullified by poor password management at another site, despite good information-security practices at the e-commerce site.

- Scam artists send out millions of fraudulent spam-email messages promising to provide a service, when all they actually want is to extract bank account numbers from naive users. Nigerian email scams are known to have caused at least five million dollars in damages, as well as at least one murder.

In general, improvements in privacy and authenticity of data and data access are directly relevant to preventing these attacks. Data-mining techniques have been shown to be very effective in exposing identity theft at online financial services such as PayPal and e-Bay. Using privacy-preserving data mining, one could potentially mine databases at multiple organizations without compromising user privacy or organizational policies. Mining of multiple databases could result in faster and more accurate identification of stolen identities.

How can one defend against web-site spoofing? Consider how these attacks work. A web criminal (usually outside the US) copies web pages from a high volume site such as e-Bay. He specifically takes care to make user-authentication pages look familiar and therefore apparently authentic, with all the usual banners and logos. He then sends legitimate-looking email to lots of users inviting them to take advantage of some deal on the target web site. The email contains links to the spoofed sites. Most users simply click on the link in the email without checking the link. They *enter their identifying information on the spoofed site* and are redirected to the real site. Note that SSL server authentication provides little protection against this attack; spoofed sites either turn off SSL or have a valid certificate for their own domain.

There has been recent progress by PORTIA researchers on the web-spoofing problem. Techniques used are similar in nature to those that have proved useful in spam-email filtering. One needs an automated tool that examines three factors—the contents of a web page, user actions at the page, and how the user arrived at the page—and decides, based on these factors, whether the page is a spoof or a valid page. Many heuristics can be used here. For example, if the page contains an image resembling the e-Bay logo, the page contains a password field, but the domain is not e-Bay and, furthermore, the user reached the page by clicking on an email link, then most likely the web page is a spoof. PORTIA researchers have built a browser plug-in that implements many such rules [9]. This tool will be made available to the public and will be updated as more sample spoofed pages are collected and studied.

Questions now under investigation include: What is an acceptable false-positive rate? Can collaborative spoof detection help? Should the tool automatically alert the site being spoofed? Can a trusted online service help in determining the legitimacy of a site requesting sensitive user data? Note that, if the attacker cannot get users to visit the spoofed site, the threat is greatly reduced. Building on work on spam-email reduction using puzzles [10] and trusted platforms [16], one might be able to provide additional defense against web-spoofing attacks.

A related question is whether one can build and deploy practical user-identification mechanisms that preserve privacy. For example, are there practical mechanisms by which a user can prove membership in a group without identifying himself? Similarly, are there mechanisms that reveal sensitive data only to parties that need to use it (*e.g.*, reveal a credit-card number to a credit-card processor, but not to merchants).

The problems caused by use of one password at multiple sites appear to have a simple technical solution. Suppose a user enters his password on the `www.yyy.com` login page. Just prior to sending the password to the site, the user's browser can compute a hash of the password and the domain-name `yyy.com` (technically, the browser would compute $\text{HMAC}_{pwd}(yyy.com)$) and send the resulting hash to the site instead of sending the user's password. In this manner, one ensures that a break-in at the low-security site, which now only sees a site-specific hash, will not reveal the password for the high-security site. As one might expect, making this simple idea work on the web is not that easy. There are many hurdles to overcome [29]. PORTIA researchers are currently developing a browser plug-in for this task. The plug-in would enable each user to have just one password that he can safely use at all sites. Note that this mechanism also provides a defense against web-spoofing, because spoofed sites now only obtain a worthless hash of the user's password.

5 Conclusions

The overarching goal of the PORTIA project is to produce both a next generation of technology for handling sensitive information that is qualitatively better than the current generation's and an effective conceptual framework for policy making and philosophical inquiry into the rights and responsibilities of data subjects, data owners, and data users. Along the way, project participants hope to focus public attention on the need to think more deeply about these issues and to reject simplistic hypotheses, including the assumptions that increased security requires decreased privacy or that restricting access to personal information is necessarily more important or effective than controlling use of that information. In this paper, we have given a few examples of the recent and ongoing work of project participants.

Acknowledgements

The PORTIA project is funded by the National Science Foundation, under the Information Technology Research program. Since September 2003, it has been carried out by 10 academic PIs and co-PIs (including the four authors of this paper), 12 research partners (from the computer industry, key user communities, Washington DC-based advocacy organizations, and the law and public-policy communities), and numerous students. Detailed information can be found on the project web site [26].

We thank all project participants for their crucial input to the PORTIA vision and the description of it that we have provided here.

References

[1] G. Aggarwal, N. Mishra, and B. Pinkas, " Secure computation of the k^{th}-ranked element," to appear in *Proceedings of Eurocrypt 2004*.

[2] D. Agrawal and C. C. Aggarwal, "On the design and quantification of privacy preserving data mining algorithms," in *Proceedings of the 20th Symposium on Principles of Database Systems*, ACM Press, New York, 2001, pp. 247–255.

[3] R. Agrawal and R. Srikant, "Privacy-preserving data mining," in *Proceedings of the 19th International Conference on Management of Data*, ACM Press, New York, 2000, pp. 439–450.

[4] M. Ben-Or, S. Goldwasser, and A. Wigderson, "Completeness theorems for non-cryptographic fault-tolerant distributed computing," *Proceedings of 20th Symposium on the Theory of Computing*, ACM Press, New York, 1988, pp. 1–10.

[5] M. Blaze, J. Feigenbaum, J. Ioannidis, and A. Keromytis, "The Role of Trust Management in Distributed Systems Security," in Secure Internet Programming: Security Issues for Distributed and Mobile Objects, LNCS volume 1603, Springer, Berlin, 1999, pp. 183–210.

[6] R. Canetti, Y. Ishai, R. Kumar, M. Reiter, R. Rubinfeld, and R. Wright, "Selective private function evaluation with applications to private statistics," in *Proceedings of the 20th Symposium on Principles of Distributed Computing*, ACM Press, New York, 2001, pp. 293–304.

[7] C. Cachin, S. Micali, and M. Stadler, "Computationally private information retrieval with polylogarithmic communication," in *Advances in Cryptology: EUROCRYPT '99*, LNCS volume 1592, Springer, Berlin, 1999, pp. 402–414.

[8] D. Chaum, C. Crépeau, and I. Damgård, "Multiparty unconditionally secure protocols," in *Proceedings of 20th Symposium on the Theory of Computing*, ACM Press, New York, 1988, pp. 11–19.

[9] N. Chou, R. Ledesma, Y. Teraguchi, and J. Mitchell, "Client-Side Defense Against Web-Based Identity Theft," to appear in *Proceedings of the 2004 Network and Distributed System Security Symposium*.

[10] C. Dwork and M. Naor, "Pricing via Processing, Or, Combatting Junk Mail," in *Advances in Cryptology – CRYPTO '92*, LNCS volume 740, Springer, Berlin, 1993, pp. 139–147.

[11] A. Evfimievski, J. Gehrke, and R. Srikant, "Limiting Privacy Breaches in Privacy Preserving Data Mining," in *Proceedings of the 22nd Symposium on Principles of Database Systems*, ACM Press, New York, 2003, pp. 211–222.

[12] A. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, "Privacy Preserving Mining of Association Rules," in *Proceedings of the 8th SIGKDD International Conference on Knowledge Discovery in Databases and Data Mining*, ACM Press, New York, 2002, pp. 217–228.

[13] C. Ellison, B. Frantz, B. Lampson, R. Rivest, B. Thomas, and T. Ylonen, "SPKI Certificate Theory," IETF RFC 2693, September 1999.

[14] J. Feigenbaum, Y. Ishai, T. Malkin, K. Nissim, M. Strauss, and R. Wright, "Secure multiparty computation of approximations," in *Proceedings of 28th International Colloquium on Automata, Languages and Programming*, LNCS volume 2076, Springer, Berlin, 2001, pp. 927–938.

[15] M. Freedman, K. Nissim, and B. Pinkas, "Efficient Private Matching and Set Intersection," to appear in *Proceedings of Eurocrypt 2004*.

[16] T. Garfinkel, M. Rosenblum, D. Boneh, "Flexible OS Support and Applications for Trusted Computing," to appear in *Proceedings of the 2003 USENIX Workshop on Hot Topics in Operating Systems*.

[17] Health Insurance Portability and Accountability Act
`http://www.cms.hhs.gov/hipaa/`

[18] J. Kaiser, "Population Databases Boom, from Iceland to the US," *Science*, November 8, 2002, pp. 1158–1161.

[19] M. Kantarcioglu and C. Clifton, "Privacy-preserving distributed mining of association rules on horizontally partitioned data," in *Proceedings of the SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery*, ACM Press, New York, 2002, pp. 24–31.

[20] E. Kushilevitz and R. Ostrovsky, "Replication is not needed: Single database, computationally-private information retrieval," in *Proceedings of 38th Symposium on Foundations of Computer Science*, IEEE Computer Society Press, Los Alamitos, 1997, pp. 364–373.

[21] N. Li, B. Grosof, and J. Feigenbaum, "Delegation Logic: A Logic-based Approach to Distributed Authorization," *ACM Transactions on Information and System Security*, 6:128-171, 2003.

[22] N. Li and J. Mitchell, "RT: A Role-based Trust-management Framework," in *Proceedings of the 3rd DARPA Information-Survivability Conference and Exposition*, IEEE Computer Society Press, Los Alamitos, 2003, pp. 201–212.

[23] Y. Lindell and B. Pinkas, "Privacy preserving data mining," *J. of Cryptology*, 15:177-206, 2002.

[24] S. Mollman, "Betting on Private Data Search," *Wired News*, March 5, 2003.
`http://www.wired.com/news/technology/0,1282,57903,00.html`

[25] Microsoft Next-Generation Secure-Computing Base – Technical FAQ
`http://www.microsoft.com/Technet/security/news/NGSCG.asp`

[26] PORTIA: Privacy, Obligations, and Rights in Technologies of Information Assessment
`http://crypto.stanford.edu/portia/`

[27] R. Rivest and B. Lampson, "SDSI: A simple, distributed security infrastructure,"
`http://theory.lcs.mit.edu/~rivest/sdsi11.html`

[28] S. Rizvi and J. Haritsa, "Maintaining Data Privacy in Association Rule Mining," in *Proceedings of the 28th International Conference on Very Large Data Bases*, Morgan Kaufmann, San Francisco, 2002, pp. 682–693.

[29] B. Ross and D. Boneh, "Simple password management for the web,"
`http://crypto.stanford.edu/WebPwd`

[30] R. Stana, Director Justice Issues, "Identity theft," Statement to the Subcommittee on Technology, Terrorism and Government Information, Committee on the Judiciary, U.S. Senate, Feb. 2002.

[31] L. Sweeney, "Replacing Personally-Identifying Information in Medical Records, the Scrub System," *J. American Medical Informatics Association*, Washington, DC: Hanley & Belfus, Inc., 1996, pp. 333–337.

[32] Total Information Awareness, `http://www.darpa.mil/iao/TIASystems.htm`

[33] Trusted Computing Platform Alliance, `http://www.trustedpc.org`

[34] J. S. Vaidya and C. Clifton, "Privacy preserving association rule mining in vertically partitioned data," in *Proceedings of the 8th SIGMOD International Conference on Knowledge Discovery and Data Mining*, ACM Press, New York, 2002, pp. 639–644.

[35] A. C. Yao, "Protocols for secure computation," in *Proceedings of the 23rd Symposium on Foundations of Computer Science*, IEEE Computer Society Press, Los Alamitos, 1982, pp. 160–164.

[36] A. C. Yao, "How to generate and exchange secrets, in *Proceedings of the 27th Symposium on Foundations of Computer Science*, IEEE Computer Society Press, Los Alamitos, 1986, pp. 162–167.