

# A Framework for Categorizing and Applying Privacy-Preservation Techniques in Big Data Mining

**Lei Xu and Chunxiao Jiang**, Tsinghua University

**Yan Chen**, University of Electronic Science and Technology of China

**Jian Wang and Yong Ren**, Tsinghua University

**D**ata mining, or knowledge discovery from data (KDD), aims to discover interesting patterns and knowledge from big data. Although its application has been largely successful, data mining can set up compromising situations for sensitive information, which is a serious privacy threat. The response to this problem has been significant enough to lead to *privacy-preserving data mining* (PPDM),<sup>1</sup> the goal of which is to safeguard information from unsolicited or unsanctioned disclosure while preserving the data's utility. PPDM approaches aim to avoid the direct use of sensitive raw data, such as an individual's ID and cell phone numbers and attempt to exclude sensitive patterns in mining results, such as clues to undisclosed personal information derived from a consumer's shopping behavior.

PPDM models and algorithms focus on the prevention of information disclosure during specific mining operations. As such, they can sometimes fail to consider privacy issues in other KDD stages, such as data preparation and the use of extracted patterns. For example, data preparation could expose the original data owners' identities and create vulnerabilities that lead to deliberate

*To protect sensitive information in mined data, researchers need a way to organize a variety of ongoing work. The Rampart framework categorizes protection approaches and encourages interdisciplinary solutions to the growing variety of privacy problems associated with knowledge discovery from data.*

abuse of data patterns or unintentional inappropriate use—both of which can compromise individual privacy and even national security.

Because it is critical to view privacy issues from a wider perspective than data mining, we investigated a range of approaches to privacy preservation across the entire KDD process. To organize our work, we created the Rampart framework, so named for its original meaning: a wall that defends against outside intrusion. Rampart clusters approaches into seven categories, each of which reflects a particular stage in the KDD process:

- › *anonymization* approaches help eliminate the privacy risk in data preparation;
- › *reconstruction* and *modification* approaches help protect sensitive information from unwanted discovery by mining algorithms;
- › *provenance* approaches, which evaluate information credibility, help deal with security issues in delivering mining results;
- › *agreement* and *trade* approaches, which use economic methods to establish an agreement or a trade between parties involved in data mining activities, help balance the interests of multiple parties involved in the KDD process; and
- › *restriction* approaches use non-technical methods to help regulate the use of sensitive information.

For each of the seven categories, we clarified research problems, reviewed related studies, and evaluated areas for future efforts.

## ANONYMIZATION

The KDD process starts with data preparation. Occasionally, the data collector and data miner are different entities, with the data collector processing data from its original owners and then making it available to the data miner. Processing must be done in a way that makes it impossible for the data miner to identify the data owners' identities and other sensitive information, yet still produce data that the data miner finds useful. To meet these two requirements, researchers have developed a number of anonymization approaches to realize *privacy-preserving data publishing* (PPDP).<sup>2</sup>

In PPDP's most basic form, the data collector has a table of multiple records, each of which corresponds to an individual and consists of four attribute types:<sup>2</sup> identifiers, quasi-identifiers (QIDs), sensitive attributes, and non-sensitive attributes. Before the table is published, identifiers are removed and QIDs are generalized or suppressed. As a result, individuals' identities and sensitive attribute values cannot be inferred from the published table. In that sense, the published table has anonymized data. A key issue of data anonymization is to choose an appropriate privacy model to quantify how much privacy can

Most existing *k*-anonymity techniques apply the same privacy preservation for all individuals, ignoring the individual's privacy preferences. Some proposed *k*-anonymity methods aim to support personalized privacy preservation,<sup>2</sup> either formulating a privacy preference as a specific parameter value (the value of *k* in *k*-anonymity) or having a specific node denote the preference in a domain-generalization hierarchy.

Although these methods have a worthy goal, it is somewhat unrealistic to expect individuals to declare their privacy preferences in such a formal way. Researchers need to find practical ways

## ASSOCIATION-RULE MINING OF RETAIL DATA COULD REVEAL CORRELATIONS THAT MIGHT COMPROMISE CUSTOMERS' PRIVACY.

be preserved. Among the many privacy models available—notably *k*-anonymity, *l*-diversity, and  $\epsilon$ -differential privacy—*k*-anonymity is the most widely used.<sup>2</sup> The idea of *k*-anonymity is to modify the values of QIDs so that every tuple in the anonymized table is indistinguishable from at least *k* - 1 other tuples.

Tables 1 and 2 show an original table and the same table after applying *k*-anonymity, where *k* = 2. After anonymization, the data utility decreases, so to balance privacy and utility, current methods generally take a greedy approach. The methods first generate multiple anonymous tables that satisfy the specified privacy model and then select the one with the least utility degradation.

to obtain personalized privacy preferences in *k*-anonymity techniques, as well as in other PPDP algorithms.

## MODIFICATION

After being anonymized, raw data that contains sensitive information about individuals is not directly mined, but mining results could still reveal some of that information. For example, a retailer might want to mine transaction data to analyze customers' behavior patterns. During data preparation, sensitive transaction data, such as credit card information, must be removed or anonymized. However, if a retailer performs association-rule mining on the data, it could find a strong correlation

**TABLE 1.** Data table without anonymization.

Age	Gender	Zip code	Disease
5	Female	12000	HIV
9	Male	14000	Dyspepsia
6	Male	18000	Dyspepsia
8	Male	19000	Bronchitis
12	Female	21000	HIV
15	Female	22000	Cancer
17	Female	26000	Pneumonia
19	Male	27000	Gastritis
21	Female	33000	Flu
24	Female	37000	Pneumonia

**TABLE 2.** Data in Table 1 with  $k$ -anonymity.

Age	Gender	Zip code	Disease
[1,10]	People	1****	HIV
[1,10]	People	1****	Dyspepsia
[1,10]	People	1****	Dyspepsia
[1,10]	People	1****	Bronchitis
[11,20]	People	2****	HIV
[11,20]	People	2****	Cancer
[11,20]	People	2****	Pneumonia
[11,20]	People	2****	Gastritis
[21,60]	People	3****	Flu
[21,60]	People	3****	Pneumonia

between customers' vitamin purchases and health state, which could compromise customers' privacy.

To prevent the privacy disclosure caused by data mining, the data miner needs to modify the data according to specific algorithms, which typically involves repeatedly adjusting modifications on the basis of mining results. The algorithm selected usually depends on the mining task type, but typical approaches include association-rule hiding and privacy-preserving classification and clustering.

### Rule hiding

Association-rule mining tries to find interesting relationships among data items. To prevent the discovery of sensitive rules, the data miner can hide association rules. Hiding strategies are based on blocking (replacing certain data attributes with symbols such as a question mark), and on probabilistic distortion (using randomly generated numbers to make original data unrecognizable).<sup>1</sup>

### Classification and clustering

Data modification methods proposed

for privacy-preserving classification vary with the adopted classification models. One model considers the privacy threat of classification based on a support vector machine (SVM), which stems from the support vectors in the learned classifier.<sup>3</sup> To destroy the sensitive information in support vectors, the model transforms the original decision function, determined by the support vectors, to an infinite series of linear monomial-feature combinations.

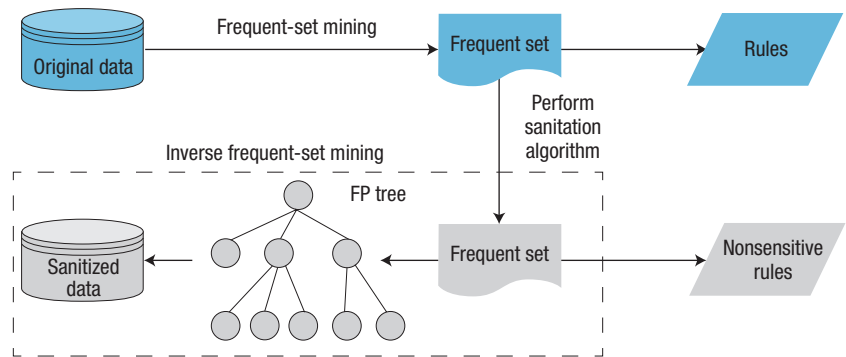
Geometric transformations, such as translation, scaling, and rotation, are often applied to establish privacy-preserving clusters.<sup>1</sup>

### Privacy versus data utility

Modified data is often less useful, so data mining must balance privacy and data utility, thereby ensuring that non-sensitive information is still available. Because data types are becoming more complex and new applications continue to emerge, finding appropriate ways to quantify privacy and utility is still a high priority in future PPDM research.

If the data miner must release the model learned from data (for example, an SVM classifier) to others, attackers might be able to infer sensitive information from the released model. Future work should look at what sensitive information can be inferred from the model's parameters, what background knowledge the attacker can use, and how to modify the learned model to prevent the sensitive-information disclosure.

Individuals, not just data miners, can modify data to protect their privacy. Apple's 2007 patent for techniques to pollute electronic profiling is an example.<sup>4</sup> The patented method essentially taints the information gathered by "network eavesdroppers" by creating a false



**FIGURE 1.** Rule hiding in data reconstruction. Rule hiding has three steps: find all frequent item sets from original dataset, use a sanitation algorithm to remove sets that correlate to sensitive association rules, and apply inverse frequent-set mining based on a frequent-pattern (FP) tree to generate a new transactional dataset.

online identity—a clone identity—for a major user, such as a service subscriber. The false identity automatically carries out numerous online actions that are quite different from the user's true activities. With this polluting technique, the user can modify true behavior data by mixing it with the massive false data created by the clone identity. In that way, an eavesdropper cannot easily discern the user's private behavior pattern.

Apple's techniques suggest the need to concentrate on data modification techniques that can help individuals protect privacy.

## RECONSTRUCTION

To prevent sensitive information from appearing in the mining results, the data miner can also reconstruct a dataset that is compatible with a given set of nonsensitive mining results.

As in data modification, a typical data-reconstruction approach is to hide sensitive rules,<sup>5</sup> and use inverse frequent-set mining (IFM) to generate a new transactional dataset.<sup>6</sup> Figure 1 shows a flowchart of this rule-hiding application, which is based on earlier work.<sup>6</sup> IFM essentially reconstructs a sanitized dataset in the following manner: given a collection of frequent-item sets and their support values (proportion of transactions that contain a frequent-item set), it finds a transactional dataset that precisely agrees with the sets' support values. Applying IFM to reconstruct a sanitized dataset seems appealing, but the problem of deciding whether a dataset exists that is compatible with the given frequent-item sets has proven to be NP-complete.<sup>6</sup>

Despite its difficulty, IFM has implications for the study of privacy-preserving approaches in data recon-

struction because it suggests that researchers could define the inverse of a range of mining algorithms. Indeed, a general inverse mining problem might be something like this: given a collection of mining results and a data mining algorithm along with its parameters, find a dataset such that the given algorithm can be used to mine the results from that dataset. If algorithms to solve the inverse problem are feasible in practice, then inverse mining can be applied to data reconstruction to meet specific privacy-preservation requirements. Because of this potential, we believe the inverse mining problem is worth exploring.

## PROVENANCE

Data mining supports decision makers who must use the results to meet certain objectives. Thus, from the decision maker's perspective, the results are sensitive information. If the results do not come directly from the data miner, the decision maker must know details about any modifications and the delivery method to determine whether the results are credible.

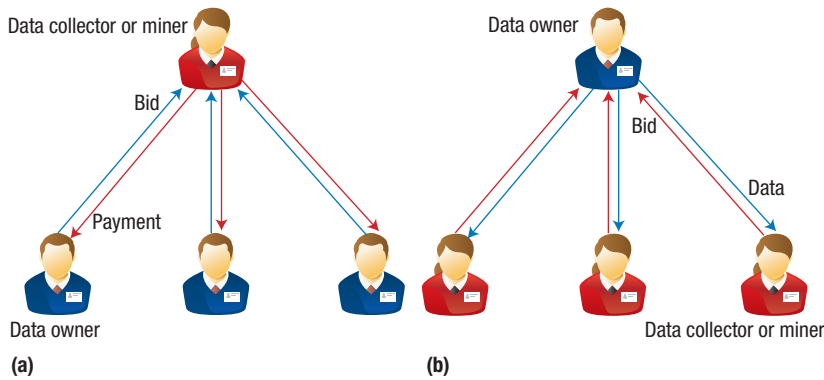
Provenance—the chronology of an object's ownership, custody, or location—enables this credibility evaluation by providing the current data's derivation history.<sup>7</sup> Provenance reveals both ancestral data (from what source the data has evolved) and transformations applied to ancestral data.

Several proposed approaches support data provenance in the context of databases and workflows.<sup>7</sup> As the Internet becomes a major platform for information sharing, there is an increasing need to know Internet information's provenance. One approach, which formally defines the data provenance problem in the context of social networks, leverages special social network features, including user profiles and interaction type and time, to obtain information provenance.<sup>8</sup>

Despite provenance's potential value in evaluating information credibility, in practice, mined-data provenance is scarce. Protocols must be set up to explicitly demand data miners and anyone else who has access to the mining result to append provenance information to the result. Moreover, if the mining result is delivered through the Internet—an open and anonymous delivery medium—it will be more difficult to get complete and accurate provenance. Current proposals to support provenance through the Internet are more theoretical than empirical. For the most part, practical approaches for obtaining mined-data provenance remain an open problem.

## TRADE

The most direct way for a data owner to protect privacy is to prevent others from accessing that data. However, in so doing, the data owner could lose the full



**FIGURE 2.** Two privacy-auction models. (a) The data owner sells data to the data collector or miner through a bid (blue arrows) and receives payment (red arrows). (b) The data owner puts data up for sale (blue arrows) to receive bids from multiple data collectors or miners (red arrows).

benefit of having that data mined. Trade is the idea that the data owner is willing to exchange some sensitive information for money or other incentives, such as discounting merchandise when the individual supplies contact information. Another incentive is personalized services. For example, a potential shopper signs up for a special membership program, essentially giving the online shopping site personal demographic information and browsing history, in exchange for tailored product recommendations and special-sales access.

As this example illustrates, in a privacy-trade scenario, the buyer, or data collector, compensates the seller, or data owner, for the owner's privacy loss. Different individuals assign different values to their privacy, so the buyer can organize an auction to get valuable information at a reasonable cost.

In one study, researchers examined a privacy auction in which multiple individuals sell their privacy to a data analyst.<sup>9</sup> Each individual possesses a private bit,  $b_i$ , which represents sensitive information, such as a medical condition that might be publicly embarrassing. The individual reports his privacy valuation, also as a bit,  $v_i$ , to the data analyst, who estimates the sum of private bits—for example, how many individuals have the condition—and determines payment to each individual.

To maximize his payment, an individual might inflate his privacy

valuation, so the data analyst needs some mechanism that encourages individuals to report a valuation truthfully. The privacy-auction study also revealed that it is possible to derive a truth mechanism by ignoring the correlation between  $v_i$  and  $b_i$ . The mechanism could help the data analyst achieve a desired tradeoff between the valuation accuracy and the payment.<sup>9</sup>

Another research group takes a different perspective from the bidding model in Figure 2a. In this privacy-auction model, shown in Figure 2b, online users put personal information up for sale, and multiple information aggregators bid to gain access.<sup>10</sup> A transactional privacy (TP) mechanism helps users decide what kind of information and how much each aggregator can obtain. TP is based on a mechanism that can bring the user near-optimal revenue.<sup>10</sup>

Although most current studies view the privacy-auction problem from the perspective of data collectors or miners and provide incentives for data owners to report truthful privacy valuations, data owners are more likely to benefit from mechanisms such as TP, which helps them gain a realistic view of their privacy information's value. Thus, we believe that research should focus on these kinds of solutions.

## AGREEMENT

Data mining typically aims to serve the interests of multiple parties. For

example, a company can profit from the results of mining customers' data, but it might need to compensate the customer to obtain high-quality data. Similarly, a customer can get better service by providing personal data to the company, but at the cost of a privacy loss.

Each party in data mining acts in a way that is most beneficial, but the parties' interests are interdependent. These characteristics are similar to those of game players, which suggests that game theory might be useful in reaching an agreement about how to use the data and allocate the profits from mining it.

Game theory formally models the situation in which agents must choose optimum actions, taking into account the effects of other agents' decisions. In PPDM research, game theory is often applied to secure multiparty computation (SMC),<sup>1</sup> which is widely used in privacy-preserving data mining. In an SMC scenario, a set of mutually distrustful parties, each with a private input, jointly compute a function over their inputs. An established protocol ensures that each party can get only the computation result, which ensures that each party's private input stays private. However, to get more benefits, a party may deviate from the protocol or collude with several other parties to expose the private input of another party.

One research group, which formalized the SMC problem as a static game, found that not penalizing anyone for dishonest behavior resulted in the parties colluding.<sup>11</sup> The group also proposed a protocol based on *cheap talk*—a preplay communication concept from game theory<sup>11</sup>—to implement a punishment mechanism that



could lead to an equilibrium (no collusion of parties).

Other researchers built a sequential game model to analyze a collector gathering private data from providers and then selling it to a user.<sup>12</sup> To protect data providers' privacy, the data collector applies some anonymization technique, which decreases the user's profit. The task then becomes how to find the game's *Nash equilibrium*—the point at which data collector and user reach consensus about the privacy-protection level. The Nash equilibrium represents the solution in a noncooperative game involving two or more players, in which each player is assumed to know the equilibrium strategies of the other players, and players have nothing to gain by changing only their own strategies.

Most approaches based on game theory adopt the following paradigm:

- › define the game's ingredients—namely, the players, actions, and payoffs;
- › determine whether the game is static or dynamic and whether the information is complete or incomplete;
- › find the game's equilibrium; and
- › analyze the equilibrium to find some practical application for it.

Although this paradigm seems straightforward, real-world applications are highly complex and require at least a few assumptions to enable any problem analysis. Unreasonable assumptions can compromise the game model's application. To ensure that analytic results have practical use, future study of approaches based on game theory should focus more on assumptions and game formulations.

## RESTRICTION

In Rampart, the first six categories concern technical solutions to privacy issues in data mining. However, non-technical solutions—laws, regulations, and industrial conventions—are also needed to protect sensitive information. Privacy-protection legislation has always been a prime concern, and many countries have established laws to regulate the disclosure of personal information. However, an individual's privacy right and what constitutes personal data remain only vaguely defined. In 2013, for example, the exposure of PRISM, the US surveillance

## COMBINING CATEGORIES: AN EXAMPLE

Rampart highlights the myriad aspects of KDD, which can provide useful insight into how different approaches can be applied together to solve the privacy issues. Even a simple example—using game theory to analyze *k*-anonymity—combines techniques in multiple categories, including anonymization, modification, and agreement. The data collector uses anonymization to preserve privacy when publishing the data, whereas the data miner uses modification and reconstruction to preserve privacy in data mining.

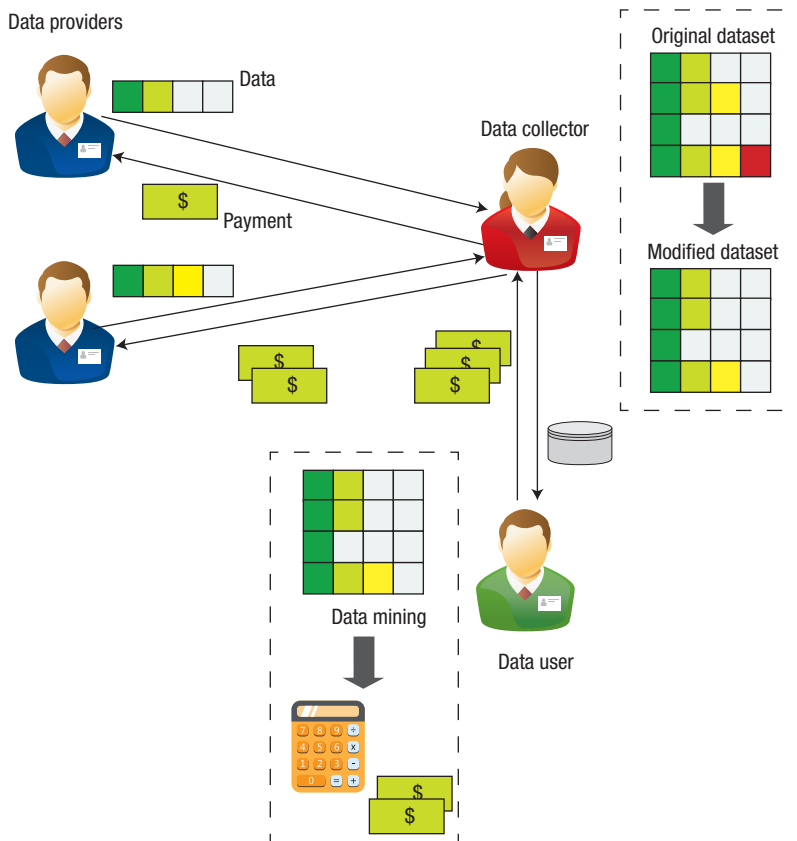
### PRIVACY-ENHANCED TECHNOLOGIES, IMPROVED LEGISLATION, AND PUBLIC AWARENESS ARE REQUIRED TO BUILD A PRIVACY-SECURE ENVIRONMENT.

program, triggered extensive debates about the extent of individuals' right to privacy ([http://en.wikipedia.org/wiki/PRISM\\_surveillance\\_program](http://en.wikipedia.org/wiki/PRISM_surveillance_program)).

US legislation must do a better job of reconciling the conflict between an individual's right to privacy and the government's need to access personal information for national security. Industry conventions on the use of personal information are also essential in building a privacy-protected environment for data mining applications. Increased public awareness of information security is an important driver; if enough people lobby for privacy protection, industry and government are more likely to respond with change.

Consider the data collection, publishing, and mining scenario depicted in Figure 3. To protect their privacy to some degree, the collector uses anonymization techniques to modify or reconstruct the data.

On the basis of the incentives and privacy-protection level the collector offers, each data provider can decide whether or not to provide private data. The collector can use an auction mechanism to manage the trade with the data providers. The data user receives modified data from the collector; if the user does not get the data from the collector directly, provenance techniques can help the user evaluate the received data's quality. The user's profit depends



**FIGURE 3.** A typical scenario involving data collection, publishing, and mining. A data collector gathers data from multiple providers and then sells it to a user, who performs certain data mining tasks. To compensate for the providers' privacy loss, the data collector offers them incentives. The tiled boxes represent a data record that consists of multiple attributes. Color tiles correspond to different sensitivity levels. Here, dark and light green indicate degrees of low sensitivity, yellow indicates modest sensitivity, and red indicates high sensitivity.

largely on the degree to which the collector modifies the data, and the user's offered price affects the data collector's modification choices. Because the data collector's decision is affected by both the data user and the providers, an agreement must be established to regulate all parties' behaviors.

To find that agreement, we developed a game model, based heavily on earlier work.<sup>12</sup> We assumed that the data collector applied some  $k$ -anonymity algorithm to the gathered data and modeled interactions among parties as a sequential game. Specifically, the data user first makes an offer to the data collector,

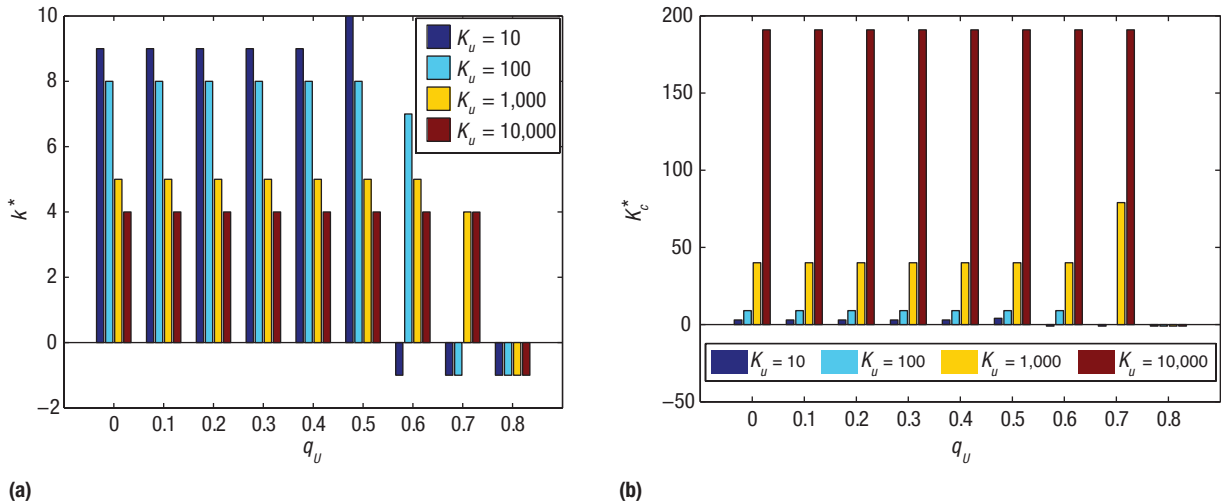
denoted as  $(K_u, q_u)$ , where  $K_u$  denotes the price the collector will pay for data of unit quality, and  $q_u$  denotes a minimum quality requirement. If the collector rejects the user's offer, the game terminates.

If the collector accepts the offer, the game continues, and the collector then announces his offer to the data providers  $(K_c, k)$ , where  $K_c$  denotes the price that the collector will pay for data of unit quality, and  $k$  is the parameter of the  $k$ -anonymity model. The larger the  $k$  value, the greater providers' privacy preservation will be. Given the collector's offer, each data provider decides whether or not to furnish data.

As in earlier work,<sup>12</sup> we used backward induction to find the subgame perfect Nash equilibriums of the game,  $[(K_u^*, q_u^*), (K_c^*, k^*)]$ . A strategy profile is a subgame perfect equilibrium if it represents a Nash equilibrium of every subgame of the original game. Because deriving analytical forms of the equilibriums is complex, we used Matlab to find the approximate optimal offer from the data collector  $(K_c^*, k^*)$  for a given pair  $(K_u, q_u)$ .

Figure 4 shows our simulation results, where  $k^* = -1$  or  $K_c^* = -1$  means that the data collector cannot find a suitable offer in response to data user's offer; in other words, the negotiation between user and collector fails. Figure 4a shows that as the data user's price increases, the value of  $k^*$  (optimal value of  $k$ -anonymity) decreases and the number of failed negotiations decreases.

For the user, a decreased  $k^*$  value means that the released data will have a higher utility; for the data collector, it means less effort to protect the data providers' privacy, which in turn means that the collector must increase the




**FIGURE 4.** Game analysis of privacy-preserving data publishing, where  $(K_c, q_u)$  is the data user's offer to the data collector. Simulation results with changing values of  $K_u$ , which denotes the price of data with some unit quality, and  $q_u$ , which denotes the minimum data-quality requirement. (a) Results with the optimal parameter of  $k$ -anonymity ( $k^*$ ), and (b) results with optimal incentive to attract providers, ( $K_c^*$ ). Game analysis helps find the point at which the needs of both collector and user are equally served.

incentive to attract providers ( $K_c$ ) so that it is still possible to collect a dataset of desired quantity and quality. Figure 4b shows this pattern.

The simulation results basically coincide with intuitions, which is evidence that the game-theoretical analysis is valid. Nevertheless, a more reasonable game formulation is worth additional research.

Exploring ways to protect sensitive information against various threats related to data mining will continue to gain importance and interest and will result in novel approaches in many of Rampart's categories. Some categories, such as anonymization and modification, have been studied extensively; others are relatively nascent, such as provenance techniques and analysis based on

game theory. We expect to continually expand Rampart's scope to keep pace with the broadening view of privacy issues related to data mining. 

#### ACKNOWLEDGMENTS

This work was partly funded by projects 61371079, 61273214, and 61271267, which are supported by the National Natural Science Foundation of China, the National Basic Research Program of China (973 Program) under grant 2013CB329105, and the Postdoctoral Science Foundation.

#### REFERENCES

1. C.C. Aggarwal and P.S. Yu, "A General Survey of Privacy-Preserving Data Mining Models and Algorithms," *Privacy-Preserving Data Mining*, C.C.

- Aggarwal and P.S. Yu, eds., Springer, 2008, pp. 11–52.
2. B.C.M. Fung et al., "Privacy-Preserving Data Publishing: A Survey of Recent Developments," *ACM Computing Surveys*, vol. 42, June 2010, pp. 14:1–14:53.
3. K.-P. Lin and M.-S. Chen, "On the Design and Analysis of the Privacy-Preserving SVM Classifier," *IEEE Trans. Knowledge and Data Eng.*, vol. 23, no. 11, 2011, pp. 1704–1717.
4. S. Carter, "Techniques to Pollute Electronic Profiling," 26 Apr. 2007, US Patent App. 11/257,614; [www.google.com/patents/US20070094738](http://www.google.com/patents/US20070094738).
5. Y. Guo, "Reconstruction-Based Association Rule Hiding," *Proc. SIGMOD2007 PhD Workshop Innovative Database Research (IDAR 07)*, 2007, pp. 51–56.
6. T. Mielikäinen, "On Inverse Frequent Set Mining," *Proc. Workshop*



## ABOUT THE AUTHORS

**LEI XU** is a postdoctoral researcher in the Department of Computer Science and Technology at Tsinghua University (THU), Beijing. Her research interests include privacy issues in data mining, text mining, and game theory. Xu received a PhD in electronic engineering from THU. Contact her at [xul0815@mail.tsinghua.edu.cn](mailto:xul0815@mail.tsinghua.edu.cn).

**CHUNXIAO JIANG** is a research associate in the Department of Electronic Engineering at THU. His research interests include the application of game theory and queueing theory in wireless communications and networking and social networks. Jiang received a PhD in electronic engineering from THU. He is a Senior Member of IEEE, the IEEE Communications Society, and the IEEE Signal Processing Society. Contact him at [jchx@tsinghua.edu.cn](mailto:jchx@tsinghua.edu.cn).

**YAN CHEN** is a professor in the School of Electronic Engineering at the University of Electronic Science and Technology of China. His research interests include data science, network science, game theory, social learning and networking, signal processing, and wireless communications. Chen received a PhD in electrical and computer engineering from the University of Maryland, College Park. He is a Senior Member of IEEE and the IEEE Signal Processing Society. Contact him at [eeecyan@uestc.edu.cn](mailto:eeecyan@uestc.edu.cn).

**JIAN WANG** is an associate professor in the Department of Electronic Engineering at THU. His research interests include information security, signal processing in the encrypted domain, and cognitive networks. Wang received a PhD in electronic engineering from THU. He is a member of IEEE and the IEEE Communications Society. Contact him at [jian-wang@tsinghua.edu.cn](mailto:jian-wang@tsinghua.edu.cn).

**YONG REN** is a professor in the Department of Electronics Engineering and director of the Complexity Engineered Systems Laboratory (CESL) at THU. His research interests include complex systems theory and its application to the optimization of information sharing on the Internet and Internet of Things, ubiquitous networks, cognitive networks, and cyber-physical systems. Ren received a PhD in electronic engineering from the Harbin Institute of Technology, China. He is a member of IEEE and the IEEE Signal Processing Society. Contact him at [reny@tsinghua.edu.cn](mailto:reny@tsinghua.edu.cn).

- Privacy-Preserving Data Mining*, 2003, pp. 18–23.
7. Y.L. Simmhan, B. Plale, and D. Gannon, *A Survey of Data Provenance Techniques*, tech. report 47405, CS Dept., Indiana Univ., 2005.
  8. G. Barbier et al., “Provenance Data in Social Media,” *Synthesis Lectures on Data Mining and Knowledge Discovery*, vol. 4, Jan. 2013, pp. 1–84.
  9. A. Ghosh and A. Roth, “Selling Privacy at Auction,” *Proc. 12th ACM Conf. Electronic Commerce (EC 11)*, 2011, pp. 199–208.
  10. C. Riederer et al., “For Sale: Your Data by You,” *Proc. 10th ACM Workshop Hot Topics in Networks (HotNets 11)*, 2011, p. 13.
  11. H. Kargupta, K. Das, and K. Liu, “Multiparty, Privacy-Preserving Distributed Data Mining Using a Game Theoretic Framework,” *Proc. Knowledge Discovery in Databases (PKDD 07)*, 2007, pp. 523–531.
  12. R.K. Adl et al., “Privacy Consensus in Anonymization Systems via Game Theory,” *Proc. Data and Applications Security and Privacy XXVI (DBSec 12)*, 2012, pp. 74–89.

# IEEE Intelligent Systems

THE #1 ARTIFICIAL INTELLIGENCE MAGAZINE!

*IEEE Intelligent Systems* delivers the latest peer-reviewed research on all aspects of artificial intelligence, focusing on practical, fielded applications. Contributors include leading experts in

- Intelligent Agents • The Semantic Web
- Natural Language Processing
- Robotics • Machine Learning

Visit us on the Web at  
[www.computer.org/intelligent](http://www.computer.org/intelligent)



Selected CS articles and columns are also available for free at <http://ComputingNow.computer.org>.