A Review of Privacy Protection in E-commerce

Li Chen and Hong-wei Liu
Guangdong University of Technology, School of Management, Guangdong Province, China
Email: 173404946@qq.com, hwliu@gdut.edu.cn

***Abstract*—Online-purchase changed the life of customers. E-business enterprises offer various personalized products to their clients through collecting private information. Customers enjoy the convenience that the personalized products bring to while suffering the risk that their privacy may be second used by firms. The conflict between consumers' privacy concern and personalization offered by firm causes many problems deserved to be studied. Many researchers study about different kinds of privacy protection. This paper analyzes the literatures and summarizes the methods of privacy protection, which are classified into two categories. The review found that, however, all these methods do not consider about consumers' profit and cannot be understood and accepted by consumers. It is raised that further study about privacy should base on consumers' behaviors and profit of firms.**

***Index Terms*—privacy protection, privacy concern, consumers' behaviors**

I. Introduction

When buy books on the Amazon.cn, the website will automatically recommend other similar books to the customer; when search for restaurant nearby, the cellphone application will offer some you may like. All these personalized services are based on customers' privacy, such as cookies, trace of browse, GPS and so forth. Amazon.cn analyze the transaction history of customers and calculate the similarity between customers to provide personalize recommendations [1]. A recent study carried out by Gomez et al, analyzing the organizational privacy practices of the top 50 most visited websites, shows that even though some large and reputable firms like Google, Microsoft, Yahoo and facebook would use customers' privacy information without authorization [2]. Companies collecting and using data and information of their client without permission may cause an associated risk that customers feel more concerns about their privacy and eventually affect their decision whether buy personalized product or not. This would cause personalization-privacy tradeoff [3], [4] Since the conflict between privacy and personalization are severe, scientists have done amount of works.

II. Literature Review

It is generally to divide current studies about privacy protection into two categories. One is protecting privacy via enacting protocols on the internet. The other is using algorithms to technologically protect private data. We obtained a listing of 30 papers in Table I in Appendix A to show two aspect of privacy protection.

A. Protocols about Protecting Privacy

The main protocols of privacy protection are Fair Information Practices (FIPs) and The Platform for Privacy Preferences (P3P). Early recognition of potential dark sides of the new technologies [5], formulation of the FIPs framework and establishing government regulatory mechanisms established such as the Privacy Act of 1974. FIPs, a set of standards governing the collection and use of personal information, are best recognized as liability rules embedded in compulsory licensing system. They are based on five core principles: notice, choice, access, security and enforcement [6]. Customers will trust a firm who implements the FIPs and willing to provide privacy information to firms [7]. P3P framework, a privacy protocol that standardizes privacy policy information to allow user to gain a better understanding of how websites' privacy policies match their action involved users' privacy [8], [9]. A privacy enhancing technology named Privacy Bird uses a notification process to inform a user browsing the Internet about how privacy friendly a website is [8], [10]; a P3P- based privacy preference generator [11]; a software named iWatch to protect individual privacy [12].

B. Privacy Protection Algorithm

In recent studies, scientists mainly focused on various algorithms of data mining, especially the association rules algorithm applied on privacy protection. According to the data storage, algorithm of privacy protection could be divided into two broad categories: Privacy protection technology for centralized data set and Privacy protection algorithm for distributed data. The main technologies of centralized data set of data mining are attributes changing, blocking and random response. For example, Agrawal proposed ID3 decision tree of privacy protection based on interference [13]. This method adds random value to original data. Then, it calculates the density function of

Manuscript received October 14, 2013; revised May 2, 2014.
The work was supported by Key Research Institute of Humanities and Social Science (NO. 12ZS0112).
Corresponding author: Li Chen, e-mail:173404946@qq.com

doi: 10.12720/joams.3.1.50-53

original data via Bayes formula so that it can rebuild the decision tree. Weiping Ge *et al.* based on the transition probability matrix to translate the attributes of data. Thus generate the decision tree by restoring property values from the data translated before [14]. The reconstruction technology of association rule mining technique means that counts support of item set based on formula to figure out the association rules after randomly translating the original data. Alexandre showed using a random operator called “select-a-size” to translate the primary data. Then randomly and independently transformed each record and used these data translated to calculate the support of item set. At last figure out the frequent item set and finish the association rules mining [15]. Distributed data mining is a popular method at present, and its privacy protection algorithm is mainly based on secure multi-party computation. Secure multi-party computation is that multiple computers are inputted data and complete the joint problem solving. This method can ensure that each computer just product specified output but not getting other information. Clifton provided four algorithms of secure multi-party computation: secure sum, secure set union, secure size of set intersection and scalar product [16].

Fig. 1 shows the quantity of research about these two kinds of privacy protection.

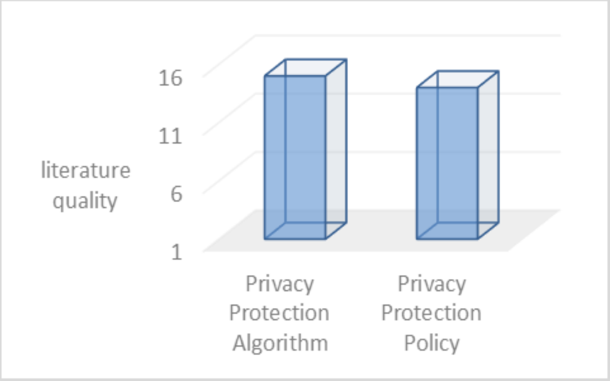


Figure 1. Category of current study

III. Problems Existed

Although a lot of works have been done about privacy protect algorithms, a large number of empirical studies confirmed that most of this algorithms were not accepted by consumers. For consumers, these algorithms are too difficult to understand. In addition, most of these algorithms are only conceptual frameworks and it is difficult to convert to actual tools. However, privacy protection policy, such as FIPs, often lack of legal authority. Companies still do not provide privacy protection for consumers and they often do not comply with the FIPs standards [17]-[21]. The primary cause of these problems is that they cannot prove how these algorithms and policies relate to the interests of consumers, and how to comply with the profit maximization principle of firms. For these reasons, there should be a way to measure the various kinds of privacy protection and methods of privacy protection should consider consumers’ profit. At present main methods of measuring privacy protections are CFIP (Concern for information privacy) and IUIPC (Internet users’ information privacy concerns). CFIP contains 4 dimensions: collection of data, unauthorized secondary use of data, improper access to data and errors [22]. More recently, Malhotra proposed IUIPC based on CFIP. IUIPC extend scale of measurement of privacy in internet. In addition to measure the protection, study should be done about addressing the concept of privacy calculus by assuming that a consequentialist tradeoff of costs and benefits is salient in determining an individual’s behavioral reactions [23]. The method of economics can effectively solve these kinds of problem. It views the concept of privacy as not absolute but, rather, subject to interpretation in “economic terms and assess the costs and benefits of a firm who implemented the privacy protection. The game theoretic on personalization has shown that personalization based on personal information not only can cause competition to be localized to individual consumers [24], [25], but also can solve the contradiction of personalization-privacy. Such a calculus perspective of privacy suggests that, when requested to provide personal information to corporations, consumers would perform a risk–benefit analysis to assess the outcomes they would face in return for the information, and respond accordingly [26]-[29]. Results of this study can make incremental contributions to the existing literature.

IV. Conclusion

With the popularity of personalized service, firms need more and more information of consumers to provide higher efficiency and more accurate personalized services. However these behaviors of firms will lead to consumer concern about their privacy. For these reason, the stream of modern privacy research had its genesis in the 1970s. In the next decades, a large number of useful studies have been conducted and published. However, because of its disjointed nature, the overall research stream has been suboptimized. We suppose that future studies should involve more behaviors and psychological feature of consumers. We believe that our recommendations for future research in privacy should lead to a more cohesive stream of literature that yields actionable steps for individuals, managers, and regulators. In conclusion, information privacy is a very current and exciting research domain that will continue to evolve as new technologies and new initiatives such as social networking or virtual worlds further push the limit of access to information.

APPENDIX A

TABLE I. STUDY OF PRIVACY PROTECTION

Method	Author	Content
Privacy Protection Algorithm	Gediminas Adomavicius et al. 2005	Algorithms base on project
	Lu Liu et al. 2008	Protect data privacy
	Li Yu et al. 2007	Protect data privacy
	Ping-feng Liu et al. 2007	Protect data privacy
	Resnick er al. 1994	Project of grouplens
	Marlin et al. 2001	User Rating Profiles
	Sarwar Bet al. 2001	recommendation algorithms
	Savia et al. 2006	Two-Way Latent Grouping Model
	J. S. Lee et al. 2005	Dimensionality of ITEM
	Agrawal et al. 2000	ID3 decision tree based on the interference of privacy
	Alexandre Evfimievski 2012	Security association rules
	Kantarcioglu 2002	Data mining based on privacy protection
	Cranor 1999	PITS and PETs
	Gkoulalas-Divanis et al. 2009	LBS
	Honget al. 2004	LBS
Privacy Protection Policy	McGinity 2000	GiliSoft Privacy Protector
	Awad and Krishnan2006	Fair information practices
	Culnan and Armstrong 1999	Fair information practices
	Hui et al. 2007	Fair information practices
	Xu et al. 2009	Fair information practices
	Robert Pitofsky 2000 8	Fair Information Practices in the Electronic Marketplace
	Culnan, M. J. 2003 9	P3P
	Eisenhardt 1989	Grounded theory approach
	Liu and Arnett 2002	Privacy policy
	Peslak 2006	Privacy policy
	Xu et al. 2008a	infromation privacy concern
	Sipior and Ward 1996	legislation of privacy
	Ateniese and Medeiros 2002	Privacy policy and economics
	Kim 2005	legislation of privacy and trust

ACKNOWLEDGE

The work was supported by Key Research Institute of Humanities and Social Science (12ZS0112).

REFERENCES

[1] G. Linden, B. Smith, and J. York, "Amazon. comrecommendations: Item-to-item collaborative filtering," vol. 5, pp. 20-25, 2003.

[2] J. Gomez, T. Pinnick, and A. Soltani. (2009). *Know Privacy*. UC Berkeley school of information report. [Online]. Available: http://escholarship.org/uc/item/9ss1m46b

[3] N. F. Awad and M. Krishnan, "The personalization privacy paradox: An empirical evaluation of information transparency and the willingness to be profiled online for personalization," *MIS Quarterly*, vol. 5, pp. 13-28, 2006.

[4] B. Kasanoff, "Making it personal: How to profit from personalization without invading privacy," *U.S: Basic Books*, pp. 44-45, 2001.

[5] M. Brenton. (1964). *The Privacy Invaders*. HeinOnline. [Online] U.S. Available: http://heinonline.org/HOL/LandingPage?handle=hein.journals/barraba5&div=39&id=&page=

[6] R. Pitofsky, "Fair information practices in the electronic marketplace," *A Report to Congress*, U.S., pp. 30-32, 2000.

[7] M. J. Culnan and R. J. Bies, "Consumer privacy: Balancing economic and justice considerations," *Journal of Social Issues*, vol. 59, no. 2, pp. 323-342, 2003.

[8] L. F. Cranor, P. Guduru, and M. Arjula, "User interfaces for privacy agents," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 13, no. 2, pp. 135-178, 2006.

[9] J. Reagle and L. F. Cranor, "The platform for privacy preferences," *Communications of the ACM*, vol. 42, no. 2, pp. 48-55, 1999.

[10] L. F. Cranor, "What do they indicate?: Evaluating security and privacy indicators," *Interactions*, vol. 13, no. 3, pp. 45-47, 2006.

[11] J. Kolter and G. Pernul, "Generating user-understandable privacy preferences," in *Proc. the International Conference on Availability, Reliability and Security*, March 2009, pp. 16-19.

[12] R. E. DeGrande and S. Donizetti, "Privacy protection without impairing personalization by using the extended system masks and the extended contextualized P3P privacy policies," in *Proc. the 12th Brazilian Symposium on Multimedia and the Web*, Brazil, 2006, pp. 11-16

[13] R. Agrawal and R. Srikant, "Privacy-preserving data mining," *ACM Sigmod Record*, vol. 29, no. 2, pp. 439-450, 2000.

[14] W. P. Ge, W. Wang, H. F. Zhou, and S. Bole, "Classification of mining based on privacy protection," *Computer Research and Development*, vol. 43, no. 1, pp. 39-45, 2006.

[15] A. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, "Privacy preserving mining of association rules," *ACM*, pp. 90-100, 2002.

[16] C. Clifton, M. Kantarcioglu, J. Vaidya, X. Lin, and M. Y. Zhu, "Tools for privacy preserving distributed data mining," *ACM SIGKDD Explorations Newsletter*, vol. 4, no. 2, pp. 28-34, 2002.

[17] C. Jensen and C. Potts, "Privacy policies as decision-making tools: An evaluation of online privacy notices," *ACM*, pp. 22-26, 2004.

[18] C. Liu and K. P. Arnett, "Raising a red flag on global WWW privacy policies," *Journal of Computer Information Systems*, vol. 43, no. 1, pp. 117-127, 2002.

[19] A. R. Peslak, "Internet privacy policies: A review and survey of the Fortune 50," *Information Resources Management Journal*, vol. 18, no. 1, pp. 29-41, 2005.

[20] A. R. Peslak, "Privacy policies of the largest privately held companies: A review and analysis of the forbes private 50," *ACM*, pp. 14, 2005.

[21] A. R. Peslak, "Internet privacy policies of the largest international companies," *Journal of Electronic Commerce in Organizations*, vol. 4, no. 3, pp. 46-62, 2006.

[22] F. Belanger, J. S. Hiller, and W. J. Smith, "Trustworthiness in electronic commerce: The role of privacy, security, and site attributes," *The Journal of Strategic Information Systems*, vol. 11, no. 3, pp. 245-270, 2002.

[23] P. H. Klopfer and D. I. Rubenstein, "The concept privacy and its biological basis," *Journal of Social Issues*, vol. 33, no. 3, pp. 52-65, 1977.

[24] Y. Chen, C. Narasimhan, and Z. J. Zhang, "Individual marketing with imperfect targetability," *Marketing Science*, vol. 20, no. 1, pp. 23-41, 2001.

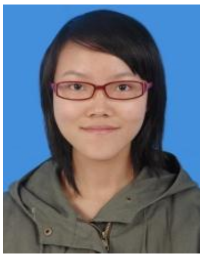
[25] V. Choudhary, A. Ghose, T. Mukhopadhyay, and U. Rajan, "Personalized pricing and quality differentiation," *Management Science*, vol. 51, no. 7, pp. 1120-1130, 2005.

[26] R. K. Chellappa and R. Sin, "Personalization versus privacy: An empirical examination of the online consumer's dilemma," *Information Technology and Management*, vol. 6, no. 2, pp. 181-202, 2005.

[27] M. J. Culnan, "Consumer awareness of name removal procedures: Implication for direct marketing," *Journal of Inter-active Marketing*, vol. 9, pp. 10-19.

[28] T. Dinev, M. Bellotto, P. Hart, V. Russo, I. Serra, and C. Colautti, "Privacy calculus model in e-commerce: A study of Italy and the United States," *European Journal of Information Systems*, vol. 15, no. 4, pp. 389-402, 2006.

[29] G. R. Milne, "Consumer participation in mailing lists: A field experiment," *Journal of Public Policy and Marketing*, vol. 16, no. 2, pp. 298-309, 2010.



Li Chen was born in 1989. She received a B.S. in 2008 from Guangdong University of technology. Currently she is a master student in Guangdong University of technology. Her major is Management Science. She now focuses on privacy in e-business and data mining in social network. Chen has written several papers.

Hong-wei Liu received a B.S in Sun Yat-Sen University in 1983 and received M.E.in South China University of Technology. He is the professor working for Guangdong University of technology.