Towards Practical Privacy Preserving Technology Adoption Analysis Service Platform

Zongwei Luo

E-Business Technology Institute, the University of Hong Kong, Pokfulam, Hong Kong
zwluo@eti.hku.hk

Abstract

Technology adoption analysis is one of the key exercises in managing technology innovation and diffusion. In this paper, we present a service platform for technology adoption analysis, with aim tailored to provide service provisioning to potential technology users and providers. With two service models provided in this platform, a practical privacy preserving framework is developed to help relieve privacy concerns of the platform participants. To illustrate the feasibility of the privacy preserving framework of this platform, an adoption process for RFID technology adoption analysis in logistics and supply chain management is presented to identify key sensitive attributes for background knowledge leading to unique identification of an individual or company.

Keywords: Privacy preservation, RFID, Service platform, Technology adoption

1. Introduction

Radio Frequency Identification (RFID), as an enabling technology for modernizing today's logistics and supply chain management, is being experimented and piloted for quite a few years. Although the potentials benefits for improving the supply chain efficiency are conceptually appealing, there exist many concerns preventing RFID adoption, from different perspectives including both technical and social factors. Furthermore, there are no easy to use tools available to provide RFID adoption analysis services to both potential RFID technology users and providers.

Meanwhile, with the rapid technology advances, business intelligence tools, e.g. data mining and knowledge discovery, are available to infer trends and patterns from adoption analysis data. The concerns about protecting the privacy of individuals as well as companies' trade secret has made it difficult to obtain valuable data on which the adoption analysis depends. Thus, privacy protection has to be considered in a technology adoption analysis tool.

Organization of the paper is as follows. In Section 2, the service platform for technology adoption is presented. In Section 3, the privacy in technology adoption analysis is reviewed. In Section 4, the privacy preserving framework for the technology adoption analysis service platform is developed. In Section 5, an RFID technology adoption analysis process is presented

to illustrate the feasibility of this analysis service platform. Section 6 concludes the paper with summaries.

2. Adoption analysis service platform

Several theories, such as Diffusion of Innovation theory [33], the Technology Acceptance Model [34], the Theory of Reasoned Action [35], the Theory of Planned Behavior [36], and Social Cognitive Theory [37], have been developed to explain adoption and acceptance of technologies. How can we leverage them including DOI to develop tools to clearly explain why it is hard to see successful cases reported for RFID adoption and then a roadmap to speed up the RFID adoption?
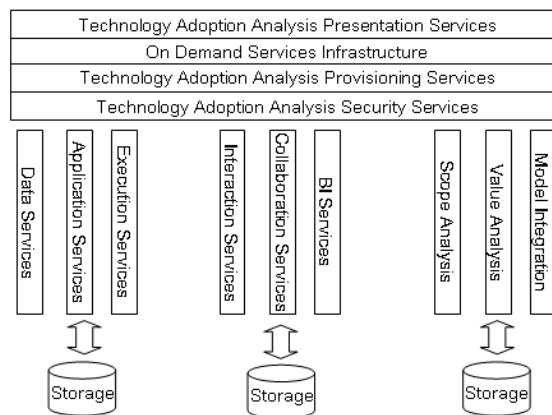
2.1 Platform architecture

Figure 1. Technology Adoption Analysis Service Platform

In [7, 45], an analysis approach has been developed to identify the adoption status based on DOI methodology. In [46], through identifying parameters for describing the value perception, a value analysis framework is developed to further prioritize those parameters and mark those most important parameters as key value indicators (e.g. cost, lead time, etc.) which can be used for value analysis in adoption. We develop our service platform for technology adoption analysis based on this value analysis framework [7, 46]. The architecture of this service framework is shown in Figure 1.

The key functional components in this service platform are the value analysis, model integration, and scope analysis. The value analysis components provide services to identify metrics to evaluate perceived value from different technology adoption parties. The value

metrics with highest priority, i.e. the most important, to the potential technology adaptors are called key value indicators. The key value indicators are strongly associated with adoption models, e.g. DOI, scope evaluation models, e.g. requirement analysis, and value analysis methods, e.g. ROI. In the value analysis framework, adoption models include DOI which guides the overall adoption analysis. Value analysis or evaluation models help identify what are treasured or concerns when adopters consider their RFID adoption. Scope evaluation models help focus the attention to the right concerns in the adoptions

2.2 Platform service models

Like other business intelligence, statistical and benchmarking exercises, the analysis data input to the service platform has great impact on the adoption analysis results. The data gathering methods like survey, polling, and focus group often need collaborative efforts from multi-party collaborations. The concerns about protecting the privacy of individuals as well as companies' trade secret have made it difficult to obtain value data for adoption analysis. Thus, privacy protection has to be ensured for all the parties involved in the adoption analysis. To help toward this, two service models are developed in this service platform for dealing with different privacy preserving needs:

- Service platform as software provisioning and execution services: in this way, data will be provided to the service platform. Privacy of the data will be preserved by the platform. During the adoption analysis, the key value indicators are processed within an analysis boundary or scope via applying suitable scope evaluation.
- Service platform as software packaging and deployment services: in this way, the technology adoption analysis algorithms are packaged as plug in components to offer to the platform participants. When these participants process their analysis data, it is not necessary for them to upload their data into service platform. To obtain the key value indicators, the necessary processes orchestrating those analysis algorithms will be packaged and downloaded into the participants' own computing environment.

3. Privacy in technology adoption analysis

Confidentiality and privacy protection has increasingly become a major concern in information gathering and sharing [1]. In the technology adoption analysis, there are three main steps which could cause potential privacy concerns:

- Analysis data gathering.
- Data analysis.
- Data publishing.

Fortunately, privacy concerns have attracted a large number of research activities leading to various privacy preserving methods.

In [4] a study is reported that a large number (about 87%) of the United States population could be uniquely identified using attributes like gender, date of birth, and 5-digit zip code. Following this, a k-anonymity method is proposed to ensure each data record is indistinguishable from at least k-1 other records with respect to certain "identifying" attributes. This allows publish data about individuals without revealing sensitive information. Since its introduction, many k-anonymity algorithms are developed, e.g. [5, 8], for data publishing. However, there are a few problems for k-anonymity methods:

- K-anonymity table generation is NP-hard [8], which has motivated approximation algorithms for producing k-anonymous tables [5].
- In a table created through k-anonymity method, an attacker can discover the values of sensitive attributes when there is little diversity in those sensitive attributes [3].
- K-anonymity does not guarantee privacy against attackers using background knowledge [3].
- Furthermore, k-Anonymity methods proposed in the literature do not adequately address preserving utility for the data analysis [2].

Miklau et al. [11] further illustrated utility preserving data publishing is a difficult task. They showed that to ensure perfect privacy, the views that are published should not be related to the data used to compute the secret query. Thus, an alternative is to publish partial documents which hide sensitive data [13]. Another privacy protection method is to control the access to the published data. Miklau et al. [12] presented cryptographic techniques to ensure that only authorized users can access the published document.

Secure multiparty computation (SMC) [14, 15, 16, 20] is a more general way for privacy protection. In SMC, n participants compute a common function on private inputs, without disclosing any information about each input other than the answer itself. It has been discussed in many collaborative applications including electronic auctions [30], card playing [31], data mining [32], and secure supply chain collaboration [10].

Du [17] proposed methods for secure two-party computations. Techniques employing randomization to guarantee privacy are proposed in [18, 19]. In [18], a commodity server model has been used for privately computing the scalar product of two vectors [19].

Yao [20] and Goldreich et al. [21] proposed generic techniques like circuits for privacy preserving multiparty computation. Generic constructions, however, tend to be impractical due to their complexity [22, 29]. In [27, 28], branching programs are used instead of circuits as function representation.

Recent research has focused on finding more efficient privacy-preserving algorithms for problems such as computation of approximations [23], auctions [24], set matching and intersection [25], surveys [26], and various data mining problems.

To summarize, although many privacy preserving methods have been proposed, all suffer from computation efficiency problem. For k-anonymity methods, another major problem is utility preserving, i.e. table created though k-anonymity shall be of good use toward target analysis.

4. Privacy preservation framework

There two major conflicts in the proposed privacy preserving methods:

- Utility conflict: Privacy preserving needs vs. analysis utility.
- Computation conflict: Privacy preserving needs vs. computation complexity.

These conflicts pose big challenges to develop privacy preserving protocols for practical use, especially in a semi honest environment. In [42], an AC-framework is proposed to achieve substantial practical utility over a semi-honest protocol. In [2], algorithms are developed towards computation efficient utility preserving k-anonymous protocols.

Data Privacy Preserving
Operation Privacy Preserving
Privacy Preserving via Access Control

Figure 2. Privacy preserving framework

In light of the utility and computation conflicts, the privacy preserving framework (see Figure 2) would be built on top of these privacy research results, and is developed by providing three types of privacy preserving capabilities: data privacy preserving (DPP), operation privacy preserving (OPP), and privacy preserving via access control (PPAC). DPP ensures data privacy protection when the service platform acts as application hosting environment to offer software as a service. OPP ensures messages exchanged among interactive parties are randomized, thus protecting trace privacy when service platform acts as software/services package environment to offer deployment services. Privacy preserving via access control ensures privacy sensitive data will be access controlled for the two service models.

To put privacy preserving methods into practical use in the service platform, the following principles are established for developing the DPP framework for technology adoption analysis data gathering:

- Background information research shall be conducted about individuals and companies to participate in the technology adoption study. Methods are available today such as clustering processes to identify a set of attributes leading to unique identification of an individual or company. This set of attributes is considered key sensitive attributes or quasi-identifiers. Please note that this set of attributes might not be minimal due to computation efficiency consideration in data clustering algorithms

available today. This approach we take is very similar to [43]. The differences are in [43] quasi-identifiers of data records are first clustered and then cluster centers are published. In our approach, we utilize these key sensitive attributes from background to study the correlation with the analysis data and those to be published data.

- Data value range is designed such that a fine granularity shall be achieved to prevent little diversity in those data entries. If little diversity is found in the data gathered [3], test will be formed to identify the correlation between the attribute with little diversity and those key sensitive attributes.
- Analysis results to be published shall be tested considering the key sensitive attributes. Again correlation relationship will be analyzed between the results with the key sensitive attributes.

It can be seen, in a trusted environment, i.e. the service platform for technology adoption analysis, if the key sensitive attributes are stable, privacy preserving could be made practical for no unique individual or companies would be revealed in the published results. However, if the service platform is not trusted, what are the differences? That means, the service platform itself could potentially reveal the undesirable information regarding privacy preserving.

To deal with this problem, access control has to be enforced in the service platform such that the key sensitive attributes would not be made public even in the service platform. That means the key sensitive attributes from Step 1 are made as secret. Step 2 and step 3 shall be performed in a way that the results will only whether correlation is positive.

Thus, the correlation study in Step 1, 2 and 3 has to be performed under secure multi-party computation principles. Methods provided in SMC for secure computation of surveys, e.g. [26] could be applied here. Let $f(x, y)$ be the correlation computation. Both parties who have the result data and the party of the key sensitive attributes would supply the private inputs x and y. Result data is made secret too before they are released. If we apply secure linear regression in the correlation analysis, the computation complexity for SMC is $O(n*m*k)$, where m is the number of players, $1=<k<=m$ and n is the number of data points [40]. It is easily to conclude that, in a 2 party SMC, this correlation study via linear regression is O(linear regression).

We can also populate data for those key sensitive attributes, and join them with the to be published result data, and apply clustering algorithms, e.g. [43]. If the new set of key sensitive attributes is the same as or of now big difference from the old set, we then have to apply k-anonymity algorithms to the to be published data.

That is, under stable key sensitive attributes for background information, the correlation study in the service platform could achieve O(linear regression) computation complexity in SMC protocol. The data gathering computation complexity now depends on the clustering algorithms chosen. However, the key sensitive attributes from the clustering might not be stable. That is, different set of key sensitive attributes would appear subject to the scope of background information selected for analysis. This requires that a good scope be chosen such that the key sensitive attributes are of good use towards unique identification of individuals or companies. However, we expect the number of set won't be significant enough to impact the correlation's O(linear regression) computation complexity in SMC.

So far, we have assumed that the key sensitive attributes are secret. However, there exist potential threats that the secret could be at risk. To address this problem, access controls to the secret and analysis results are to be enforced. This would lead to accountability protocols [42] for a party who correctly followed the protocols can be proven to have done so and consequently prove that someone else must have improperly disclosed data. However, in [42], although a party's malicious behavior could be detected, it may learn things that it should not and damage the result. To address this problem in the service platform, instead of revolving it via SMC, statistical significance is required for the data gathered. Then ongoing partial computed results will not be revealed to the participating party until the results are published. In this way, a malicious party's data could pose limited impact towards the final results. And this would discourage the malicious party for partial results won't be available, making it impossible to guess meaningful information based on partial results. Further, it is assumed that multiple data submissions from a single party could be detected and prevented.

To summarize, the privacy preserving framework for data privacy preserving in the service platform leverages k-anonymity and SMC to provide a trustworthy with data privacy preserving technology adoption analysis environment. A set of principles are designed to create an environment to make k-anonymity and SMC more practical to use while avoiding known k-anonymity and SMC pitfalls.

In the service platform, a randomization of trace generation is utilized to protect the operation privacy such that the interaction trace between the party and the service platform is protected. This would raise the privacy compromising barrier to give the service platform participating parties more confidence that even the service platform is difficult to trace their interactions since the platform log will contain traces with randomized data. It would also prevent malicious party from using replay attack to guess the exchanged messages. The following is a lightweight protocol developed for randomization of the operational traces in the interactions among parties in the service platform.

The randomization starts with a mutual authentication among the parties in the intended conversation. Once the authentication is complete, secret information and transaction identifier between the parties are agreed upon, say S_i and T_i. Public hash functions in hash chains, H and G are also agreed upon. The secret information and transaction number are kept secret. The messages M_i exchanged between parties Ps and Pr are randomized with a random number Ri. This message exchange protocol could be further simplified such that a lightweight mutual authentication protocol, e.g. [44], is used after a party login to the service platform via PKI mechanisms. Thus, each time, the parties to interact will first mutual authenticate each other via a light weigh mutual authentication protocol. By applying the above randomization process, the message exchanged among parties in the service platform are protected, losing utility for log data mining or knowledge discovery for revealing operation traces in the service platform. Besides messages, party identifiers as well as other sensitive information could also be randomized.

Access control helps protect the secret information in order to enable data and operation privacy preserving. During the discussion in data and operation privacy preserving, a few assumptions are made that sensitive information, e.g. key sensitive attributes in data privacy preserving, and secret information, e.g. shared secret among interactive parties in operation privacy preserving are access controlled.

5. Adoption analysis example

So far, we have described a service platform with privacy preserving capabilities. To illustrate the feasibility of this service platform, a process is constructed for identifying key sensitive attributes in background information research. This process is packaged as plug in and put in the local environment for execution.

To study RFID adoption in retailers in China, we place the retailers in supply chains to identify potential impacts from RFID adoption on the retailers. To achieve this, we have to document or model the supply chains operations in which the retailers participate. Therefore, we adopt the processes defined in a reference model - the Supply Chain Operations Reference-model (SCOR) [38] for the technology adoption analysis in logistics and supply chain management. To simplify the illustration, we will illustrate implementation of this approach through studying RFID adoption in retailers in China. We will place the retailer RIC in a three echelon supply chain (retailer, distribution center and manufacturer) to identify potential impacts from RFID adoption on the retailers. Output of the preparation will be the case study requirements, i.e. whether or how retailer RIC benefits from the RFID adoption and the impact on the retailer operational processes. SCOR models are

downloaded from the service platform to the local environment.

Retailer RIC sits at the very end in the supply chains, resulting in its sensitivity to the changes and fluctuations of the supply chains it participates by introducing RFID. To examine the impact of RFID on the whole supply chain is not a simple task. The activity-based costing is based on the concept of allocating the indirect and overhead cost to the activity level so as to improve the accuracy of cost accounting. Hence we intend to "allocate impact" of RFID on the RIC supply chain based on supply chain activities. Based on RIC supply chain, according to the interaction and relationship among the activities defined in SCOR model [38], we first classify them into different categories, based on the criteria whether RFID adoption would lead any potential impact on them, including product tracking, receiving & shipping, asset management, inventory control, packaging, shelf management and check out, data collection, regulation and customer requirement compliance, and returns management.

From these activity categories, we would further identify value activities which have potential RFID impact. For example, in Inventory control category, there are activities like ES.4 Manage Product Inventory, D4.1 Generate Stocking Schedule, and ED.4 Manage Finished Goods Inventories from [38]. These activities are identified and included because real time RFID data could potentially lead to timely information for managing various inventories, either to reduce the inventory level or the lead time to generating stocking schedules. Similar to the activity-based costing concept of allocating the indirect and overhead cost to the activity level, we "allocate impact" of RFID on the RIC supply chain based on value activities to identity metrics for impact evaluation.

From the metrics identified, we select the attributes which have non-negligible impact on the revenue or expense of retailer RIC based on the dynamic behavior by introducing RFID into the operations. These attributes are weighted so priority of each could be ranked. We rank each attribute and allocate reasonable probabilities to them according to the retailer RIC's business operations. These probabilities assigned most likely would vary from case to case, from company to company, and from time to time. Each retailer can make its own decision on the probabilities according to their situation (such as innovator DOI adoption stage) and make adjustment when necessary according the new development in RFID technology and their supply chain operations.

6. Conclusion

In this paper, we present a service platform for technology adoption analysis, with aim tailored to provide service provisioning to potential technology users and providers. Two service models are developed to offer privacy preserving capabilities in the service platform. In this platform, a practical privacy preserving framework based on the two services models is developed to help relieve privacy concerns of the platform participants. To illustrate the feasibility of this service platform, a process is constructed for identifying key sensitive attributes in background information research. This process is packaged as plug in and put in the local environment for execution. The process includes five steps, i.e. preparation, activity analysis, metrics identification, metrics ranking, and further data processing.

References

[1] Time. The Death of Privacy, August 1997.

[2] Rhonda Chaytor, Utility-Preserving k-Anonymity, Master of Science Thesis, Department of Computer Science, Memorial University of Newfoundland
October 2006

[3] Machanavajjhala, A., Gehrke, J., Kifer, D., and Venkitasubramaniam, M. 2006. l-diversity: Privacy beyond k-anonymity. In ICDE.

[4] Sweeney, L. 2000. Uniqueness of simple demographics in the u.s. population. Tech. rep., Carnegie Mellon University.

[5] Aggarwal, G., Feder, T., Kenthapadi, K., Motwani, R., Panigrahy, R., Thomas, D., and Zhu, A. 2004. k-anonymity: Algorithms and hardness. Tech. rep., Stanford University.

[6] Sweeney, L. 2002. k-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 10, 5, 557–570.

[7] Zongwei Luo, et al., Model Integration for Technology Adoption Analysis, IEEE ICAL, 2007

[8] Meyerson, A. and Williams, R. 2004. On the complexity of optimal k-anonymity. In PODS.

[9] Samarati, P. 2001. Protecting respondents' identities in microdata release. In IEEE Transactions on Knowledge and Data Engineering.

[10] M. Atallah, H. Elmongui, V. Deshpande, and L. Schwarz. Secure supply-chain protocols. In IEEE International Conference on Electronic Commerce, pages 293{302, 2003.

[11] Miklau, G. and Suciu, D. 2004. A formal analysis of information disclosure in data exchange.In SIGMOD.

[12] Miklau, G. and Suciu, D. 2003. Controlling access to published data using cryptography. In VLDB. 898–909.

[13] Yang, X. and Li, C. 2004. Secure XML publishing without information leakage in the presence of data inference. In VLDB. 96–107.

[14] Goldreich, O., Micali, S., and Wigderson, A. 1987. How to play any mental game. In STOC'87: Proceedings of the 19th ACM Conference on Theory of Computing. 218–229.

[15] Ben-Or, M., Goldwasser, S., and Wigderson, A. 1988. Completeness theorems for non-cryptographic fault-tolerant distributed computation. In STOC '88: Proceedings of the 20th ACM Symposium on Theory of Computing. 1–10.

[16] Chaum, D., Crepeau, C., and Damgard, I. 1988. Multiparty unconditionally secure protocols. In STOC '88: Proceedings of the 20th ACM Symposium on Theory of Computing. 11–19.

[17] Du, W. 2001. A study of several specific secure two-party computation problems. Ph.D. thesis, Purdue University.

[18] Beaver, D. 1997. Commodity-based cryptography. In STOC '97: Proceedings of the 29th ACM Symposium on Theory of Computing. 446–455.

[19] Du, W. and Zhan, Z. 2002. A practical approach to solve secure multi-party computation problems. In New Security Paradigms Workshop 2002.

[20] A. Yao. How to generate and exchange secrets. In Proc. 27th Annual IEEE Symposium on Foundations of Computer Science (FOCS), pages 162–167. IEEE, 1986.

[21] O. Goldreich, S. Micali, and A. Wigderson. How to play any mental game. In Proc. 19th Annual ACM Symposium on Theory of Computing (STOC), pages 218–229. ACM, 1987.

[22] Justin Brickell Donald E. Porter Vitaly Shmatikov Emmett Witchel,, Privacy-Preserving Remote Diagnostics, CCS007, October 29 - November 2, 2007, Alexandria, Virginia, USA.

[23] J. Feigenbaum, Y. Ishai, T. Malkin, K. Nissim, M. Strauss, and R. Wright. Secure multiparty computation of approximations. In Proc. 28th International Colloquium on Automata, Languages and Programming (ICALP), volume 2076 of LNCS, pages 927–938. Springer, 2001.

[24] M. Naor, B. Pinkas, and R. Sumner. Privacy preserving auctions and mechanism design. In Proc. 1st ACM Conference on Electronic Commerce, pages 129–139. ACM,

[25] M. Freedman, K. Nissim, and B. Pinkas. Efficient private matching and set intersection. In Proc. Advances in Cryptology - EUROCRYPT 2004, volume 3027 of LNCS,pages 1–19. Springer, 2004.

[26] J. Feigenbaum, B. Pinkas, R. Ryger, and F. Saint-Jean. Secure computation of surveys. In Proc. EU Workshop on Secure Multiparty Protocols, 2004.

[27] E. Goh, L. Kruger, D. Boneh, and S. Jha. Secure function evaluation with ordered binary decision diagrams. In Proc. 13th ACM Conference on Computer and Communications Security (CCS), pages 410–420. ACM, 2006.

[28] M. Naor and K. Nissim. Communication preserving protocols for secure function evaluation. In Proc. 33rd ACM Symposium on Theory of Computing (STOC), pages 590–599. ACM, 2001.

[29] O. Goldreich. Secure multi-party computation. http://www.wisdom.weizmann.ac.il/home/oded/public html/pp.html, 2001.

[30] O. Baudron and J. Stern. Non-interactive private auctions. In Financial Crypto'01. Springer, 2001.

[31] O. Goldreich, S. Micali, , and A. Wigderson. How to play any mental game. In Annual ACM Symposium on Theory of Computing, pages 218{229, 1987.

[32] Y. Lindell and B. Pinkas. Privacy preserving data mining. In Advances in Cryptology { CRYPTO'00, pages 36{54, 2000.

[33] Rogers E. (2003) Diffusion of Innovations, 5th ed., The Free press

[34] Davis R. (1989) "Perceived usefulness, perceived ease-of-use and user acceptance of information technology," MIS Quarterly (13)3, pp. 319-339

[35] Ajzen I. and M. Fishbein, (1980) Understanding Attitudes and Predicting Social Behavior, Englewood Cliffs, NJ: Prentice-Hall

[36] Ajzen I. (1982) "From intentions to action: a theory of planned behavior" in J. Kuhl and J.Beckmann (eds), Action Control: from Cognition to Behavior, New York: Springer-Verlag, pp. 11-39

[37] Compeau D. and C. Higgins (1995) "Application of social cognitive theory to training for computer skills," Information Systems Research (6)2, pp. 118-143

[38] Supply-Chain Council (2006), Supply Chain Operations Reference Model SCOR Version 8.0, 2006

[39] D. Beaver. Foundations of secure interactive computing. In Proc. Advances in Cryptology - CRYPTO 1991, volume 576 of LNCS, pages 377–391. Springer, 1992.

[40] Mikhail Atallah Marina Bykova Jiangtao Li Keith Frikken Mercan Topkara, Private Collaborative Forecasting and Benchmarking, WPES'04, October 28, 2004, Washington, DC, USA.

[41] Curtin J., R. Kauffman, and F. Riggins (2007) "Making the 'MOST' out of RFID technology: a research agenda for the study of the adoption, usage and impact of RFID", Information Technology and Management (8)2, pp 81-110

[42] Wei Jiang and Chris Clifton, "AC-Framework for Privacy-Preserving Collaboration", 2007 SIAM International Conference on Data Mining (SDM07), Minneapolis, Minnesota, April 26-28, 2007.

[43] Gagan Aggarwal1, Tomas Feder, et. al. Achieving Anonymity via Clustering, PODS006, June 26-28, 2006, Chicago, Illinois, USA.

[44] Zongwei Luo, Terry Chan, Jenny Li, A Lightweight Mutual Authentication Protocol for RFID Networks, ICEBE 2005

[45] Zongwei Luo, Benjamin Yen, Tan Zhining, Ni Zhicheng, RFID Technology Adoption Analysis in China, ICEBE 2007

[46] Tan Zhining, Ni Zhicheng, MSc Thesis, Department of Computer Science, the University of Hong Kong, 2007