International Journal of Security and Its Applications
Vol.8, No.3 (2014), pp. 143-152
http://dx.doi.org/10.14257/ijsia.2014.8.3.16

Privacy Protection in E-commerce: Identity-based Anonymous Privacy Agent

Dan Guo[1,2]
[1]*College of Computer and Information Engineering, Harbin University of Commerce, Harbin, China*
[2]*College of Computer Science and Technology, Harbin Engineering University Harbin, China*
hrbcucomputer@126.com

Abstract

An identity-based anonymous agent privacy protection scheme is proposed to deal with the problems of privacy information disclosure. In this paper, some analysis about the status quo of the privacy protection research in China and some partitioning of privacy data in e-commerce are given. We further illuminate the whole process of privacy protection by an e-commerce privacy protection scheme based on the identity of the anonymous agent. At last, the simulation results are analyzed. Experimental results show that the novel scheme can ensure the anonymous reliability, integrity and efficiency to realize privacy protection in e-commerce process.

Keywords: *privacy; security; e-commerce; anonymous; agent*

1. Introduction

With the rapid development of the internet business, a lot of privacy information as the trade information is stored, forward, release, and at the same time, the purpose of the various agencies try to mining useful privacy information from these mass data, in order to meet their needs. Therefore, to protect privacy data and prevent sensitive information leakages become the current challenges facing. The current privacy protection researches focus on the research of privacy data hidden. No matter use what kind of technology is hard to meet the integrity in data hiding process and efficiency in access process of win-win. Proof of identity is the basic characteristics of trusted computing. The presence of privacy and specific identity is a one-to-one mapping relationship; therefore, identity itself is the bottleneck of privacy inference proof process. If the complex privacy hidden problem is transformed into simple identity hidden problems, the calculation for this hidden process will be far less than the process for data itself hidden.

2. Research Actuality in China

A. Research Aim of Privacy Protection

Privacy protection technology, a new research hotspot, is worth being investigated further both in theoretical research and practical application .Considering the following two aspects of the privacy protection technology for implement data privacy protection [1,2]. How to ensure that no privacy was leaked in the process of data applications. How to make the protection of privacy is more advantageous to the data application. At present, focusing on

ISSN: 1738-9976 IJSIA

how to design better privacy protection principle and algorithm to achieve the balance of the two is the main work of privacy protection.

B. Research Branch of Privacy Protection

Domestic research on privacy protection technology can be summed up in three aspects, which based on data distortion or data encryption technology. The one is privacy protection technology based on cryptography, which adopt techniques of cryptography theory process privacy information to solve the anonymity problem in data mining, such as secure multi-party computation [3-4].In addition, based on the limited release of privacy protection technology was studied and put forward the support multiple constraints k - anonymous method[5].One another is Identifier anonymous protection, in order to protect the personal information which be used for specific identifiers such as name, id number, social security code to delete or cryptographic operations. But the attacker can get other information through other channels deduction and identify the user such as national, date of birth, sex, address and so on. According to the above ,the method can be adopted such as a conditional release data as don't published some sensitive mean threshold of the data or data generalization [6, 7], etc. The third direction is data distortion, that is initial data should be distorted , disrupt, randomization process by data cleansing before data being used such as classification mining algorithm based on privacy protection [8], association rule mining [9, 10], distributed data privacy protection collaborative filtering recommendation [11], grid access control [12], etc. . Although this kind of method can guarantee the results of the statistical properties of the whole, it is often at the expense of the incomplete data.

No privacy protection technology is suitable for all applications. Although many new methods [13-16] hard to due to a certain category simply having combined several technology, can be good use certain types of technology to solve the problem of privacy protection, introduction of other defects is inevitable.

3. Privacy Data Partitioning in e-commerce

A. Privacy Data Hierarchy

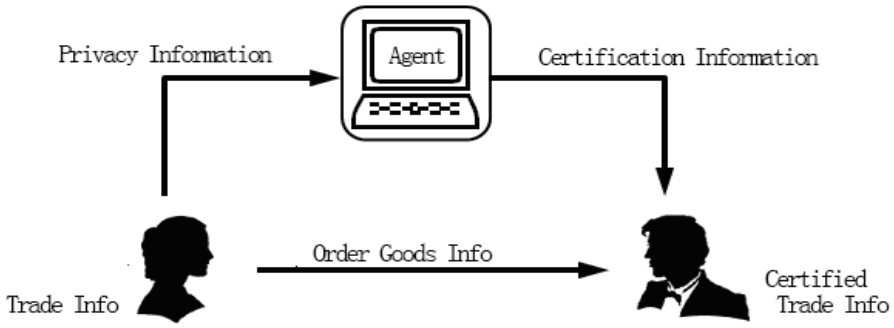

Figure 1. This is the Privacy Agent Process DFD

There are three aspects of the object involved, user, agent and merchants, as shown in Figure 1. Above all, user and agent establish agency relationship. In the subsequent transactions, the privacy agency responsible for providing users with verifiable information and guarantee privacy won't be leaked. Merchants confirm trade is valid with the authentication information provide by privacy agent party. Therefore, privacy agent process

design starts from the e-commerce process between users and business, to embed the privacy agent existing in the electronic commerce system.

Privacy data hierarchy based on the definition of e-commerce privacy protection consists of four layers. Private or personal information protection is a hard nut to crack, because not all the data need to be kept secret. Privacy of data is identified as privacy or not depends on the data owner himself. For example, one person A always receive form unknown product recommendations which analysis through a series shopping search records of A in some shopping web, but she doesn't like to accept any online from the unknown group recommendations. So, the best way is to keep her shopping search records as confidential data. On the other hand, the person B just upset for getting more recommendations in his search, received an online commodities recommended would make him a better choice. So, he doesn't mind revealed his shopping search records. Privacy means different things to different people. Therefore, according to the different problems this paper put forward the following data privacy at different levels.

- Sensitivity security level for protected privacy data itself
- Subjective intention security level for privacy protection request main body
- Operation set security level for set of e-commerce process operation which possible lead to privacy disclosure
- Attainability security level by actual implemented protection strategies

B. Privacy Data Partitioning

In the specific application, usually private data include the sensitive data and the characteristics of data representation. Privacy data usually refers to what we call the privacy of sensitive data, such as personal consumption record, age, address, family members and the education experience, wages, the patient's illness records, the company's financial information, etc. However, the definition of privacy is diverse in given sets of different data and the data owner. Such as conservative patient see disease information for privacy but open patient do not see it as a privacy data.

In this paper, the analysis of the data analysis is as follows: The analysis on the data that may be used for processing to lead to privacy leakage; user privacy security demand level data analysis; action set analysis of e-commerce privacy protection.

The data that may be used for processing to lead to privacy leakage have four types of data form. Personal data, any data that can be used to identify a person such as name, address, telephone number. Sensitive data, the data of any disclosure of race, religion, philosophy and other beliefs, political views, join a political party, travel, and personal health condition, such as medical history, travel, etc. Identification data, personal data, is the data subjects that allow direct identification of such as id number, health insurance number, social security number, bank card number, etc. Anonymous data, any data that not associated with any logo or identifiable data themes such as sex, blood type, education, history, disease, *etc*.

User privacy security demand level data analysis. According to user requirements for protection of privacy, determine the different level of demand, to match the level of operational safety, for security and protection measures. Each user according to his own will may put forward a variety of privacy protection demand with its own characteristics in some interval may be very stable, may also with the change of the understanding of privacy information to frequent changes. Therefore, demand level data analysis must ensure that clear division at all levels should possess strong transformability, in order to ensure the maneuverability of the privacy protection.

The action set analysis of e-commerce privacy protection. Usually network shopping operation involves several steps, purchaser to apply to buy account, select a product, provide payment information and a real address, wait for logistics distribution, once receiving the product or service immediately confirm the order, complete the transaction. In the whole process of e-commerce transactions, each link is likely to be involved in the operation of the privacy, and privacy protection action set analysis is to separate out some levels of action set, such as correlation is weak, medium and strong, and so on , from numerous operation sets to establish one-to-one relationship between the security level of definition phase.

4. Privacy Anonymous Agent in e-commerce

The architecture design of privacy agent cluster is shown in figure 2.Privacy agent cluster consists of privacy agent system and neighborhood systems, in which privacy agent system is an internal system to privacy agent implement and neighborhood system is an external coordination system to realize the communication between different agencies. Master control system of privacy agent system consists of transaction processing machines, data processing machines, spare processing machines and related auxiliary cluster. Various processing machines together to complete the agent system master control function in agent system operation. The sub agent system controlled and managed by the master control system through the trusted platform, providing credible network interface to the agent system to ensure the high internal credible performance of it, through which processing data can exchange data with the neighborhood system.

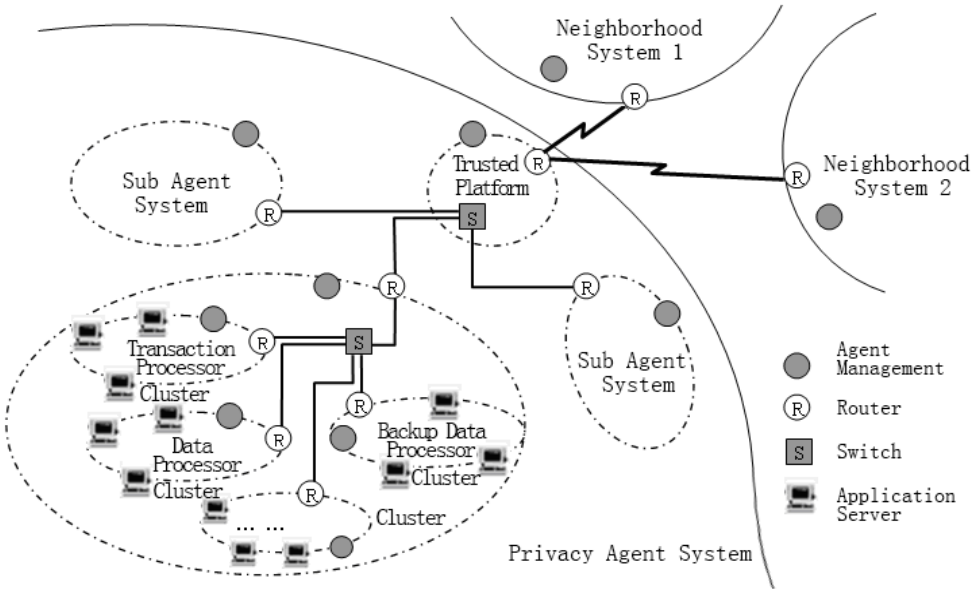

Figure 2. This is the Architecture of Privacy Agent System

The agent process management is the mainly responsible of transaction processor, which requires a high-performance processor to make sure the efficient and stable of the system. As a private transaction processor, transaction processor has to make sure the security of the data that is involved in the transaction. Therefore, safety measures can be adopted by the security protection gate combining with relevant security protection strategy, locking protected data to ensure data security follows physically.

Privacy agent machine, data processor, involving a lot of privacy data storage, usually is set as a secret database storage machine. Considering the data access can use distributed database. Considering integrity can use multistage storage mode or multistage backup storage. Considering the data security some encryption processing can be set.

According to the different backup cycle, privacy agent machine, backup data processor, back up data in accordance with the different levels. Once system is attacked, backup information about system and log files can be used to restore the system and original data.

User agent machine U disk lock be used for storing personal digital certificate that should not be read. As the user successfully applied to privacy agent, as the U disk lock is issued by a private agency. Only a U disk lock user can request privacy agent modification, cancellation, recovery, and so on.

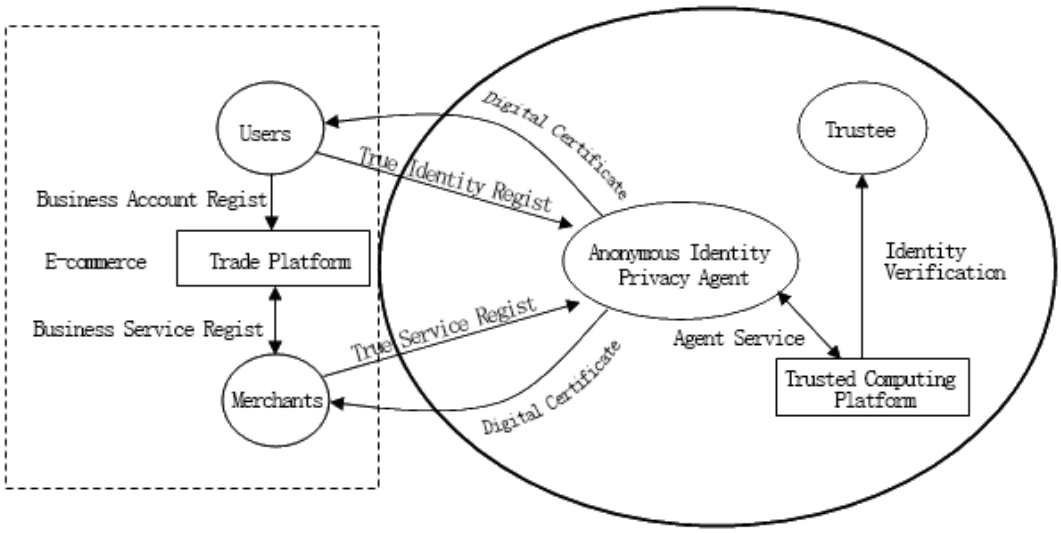

Figure 3. This is the Model based on the Identity Agent of the Privacy Protection Services

The model based on the identity agent of the privacy protection services is shown in Figure 3, in which the privacy agent process is achieved by the method of privacy agent multiple group blind signature certificate signing algorithm.

The most common e-commerce identification problem is the user's identity must be first identified, when a user logs in an e-commerce platform. And the important guarantee of privacy agency protocol is an agent as the third party can provide credible authentication to ensure that the information have been published by user does not involve any privacy information in the certification process, and the identity authentication agent for many of the same user operation to match the identity by the different hidden labeled to ensure could not find the true identity after services on many occasions, that is, the agent machine every time users are assigned to a different identity as if every time is an independent service. What privacy agent should do is to agent the identity authentication of the user, the purpose of which is after the main body of the privacy was been agented hidden the data utilized by network will not be associated with the main body the data belong to, and the disrelation as which does not affect the process of e-commerce transactions.

The tripartite relationship of the agent protocol is shown in Figure 4.

The secure identification protocol should at least meet the following two conditions:

Certifier C can prove to the verifier V that he is a P;

In addition to the necessary authentication information verifier V cannot obtain any other useful information about C.

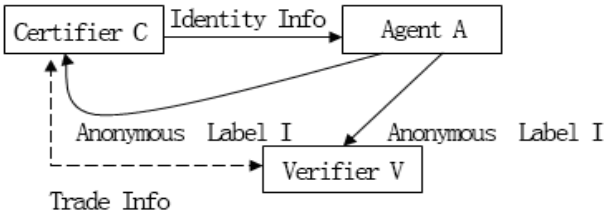

Figure 4. This is the Tripartite Relationship of the Agent Protocol

The usual identification protocol is described as follow:

Certifier C provides relevant certification information shows his identity to the verifier V, including personal privacy information. The verifier V verification certificate is P's true identity. The verifier V and certifier C establish a connection having been certified, the two sides in the transaction with the identity of the certified authorization for a deal.

The tripartite relationship of the agent protocol is shown in Figure 4, in which the privacy anonymous agent identification protocol in e-commerce is described as follow:

According to the problem should be resolved, identity zero knowledge proof protocol analysis can be part of the agency agreement for privacy. Zero knowledge proof can make the verifier believe that a statement is correct but without any useful information should be provided to the verifier. Zero knowledge proof is essentially a protocol involving two or more parties, namely two or more parties required to complete a task in a series of steps. Certificate to prove to the verifier and make them believe they know or have a particular message, but prove to verifier cannot leak any information about prove to message. Certifier C provides relevant certification information shows his identity to the agent A, including personal privacy information. The verifier A verification certificate is C's true identity. The verifier A and certifier C establish a connection having been certified. In a deal, certifier C makes a trade request through Agent identity given by privacy agent A to the verifier V. Verifier V verify the virtual identity of certifier C provided by privacy agent A. The verifier V and the privacy agent A establish a connection having been certified. Then the verifier V and the certifier C establish a connection having been certified. Verifier V in the trading and certifier C use this virtual identity with the identity of the certified authorization for a deal. Certifier C and verification V, in the whole process of e-commerce transactions, did not show any personal privacy information.

The privacy agent multiple group blind signature algorithm is described as follows:

The agent blind signature adopts by the paper allows privacy agent blind the identity message of user first, and then let t merchants signs the message having been blinded. At last user eliminates the blind factor and gets the signature signed by merchants. Between the user and anonymous agent cluster, the signature algorithm adopts the agent multi-signature algorithm based on agent blind signature algorithm. In ordinary agent signature scheme, signature agent on behalf of the original signature user can sign a signature, which only represent one original signature user, a signature agent can also on behalf of multiple original signature users sign the deal. The similarity of ordinary agent multi-signature and the agent

multi-signature in this paper is both sign by a third agent party. What essential difference in two signature protocol is ordinary agent multi-signature allows an agent exercise of some right for the original signer , however the agent multi-signature in this paper only agent the identity hidden and give the alias as the identity authentication but not a substitute for the original signer to exercise any right. In the agent process, after the user privacy identity is agented, anonymous system constantly updated identify labels, signature algorithm between which and anonymous agent machine adopts the way on agent multiple group blind signature to makes any one member of a group of different identity labels for the same user can on behalf of the user for anonymously signature. The agent multiple group blind signature is also used in anonymous agent process between anonymous agent cluster and merchants .In this scenario, any member of the agent cluster can on behalf of the entire cluster by the way of anonymously broadcasting multi-signature. What's more is the anonymous user identity authenticated by agent multiple group blind signature can be publicly identified.

Blind signature scheme implementation is described below:

i. let U is the receiver of the signature, S is the signer, p and q are the large prime numbers to meet safety requirements, p is a big factor of p-1,g is the primitive element of z_q^* ,S has a public-private key pairs(x,y),in which $y \equiv g^x \bmod p$,m is signed message,h is secure hash.

ii. The signature process is shown in Figure 5.

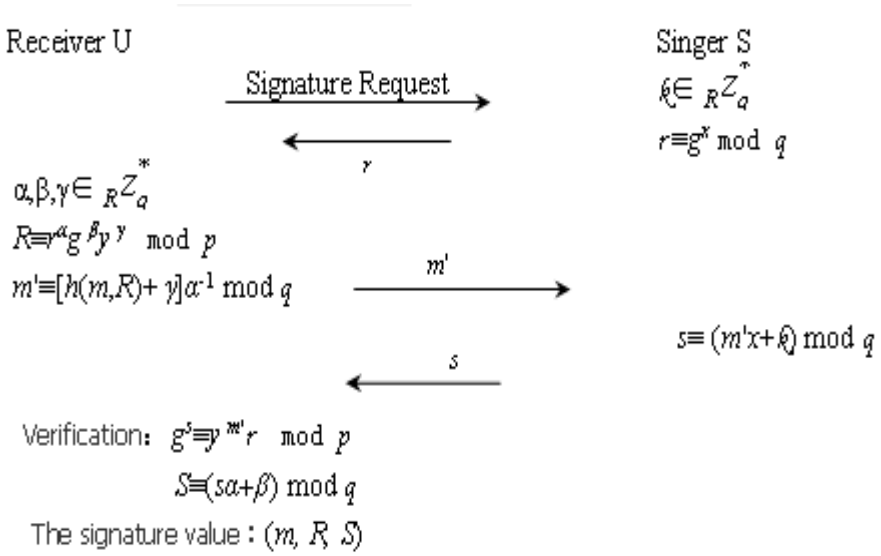

Figure 5. This is Signature Process of the Blind Signature Scheme

iii. After having gotten the signature, verifier verifies equation: $g^{s} \equiv y^{h(m,R)} R \mod p$. The signature is effective if equation is set up, otherwise is invalid.

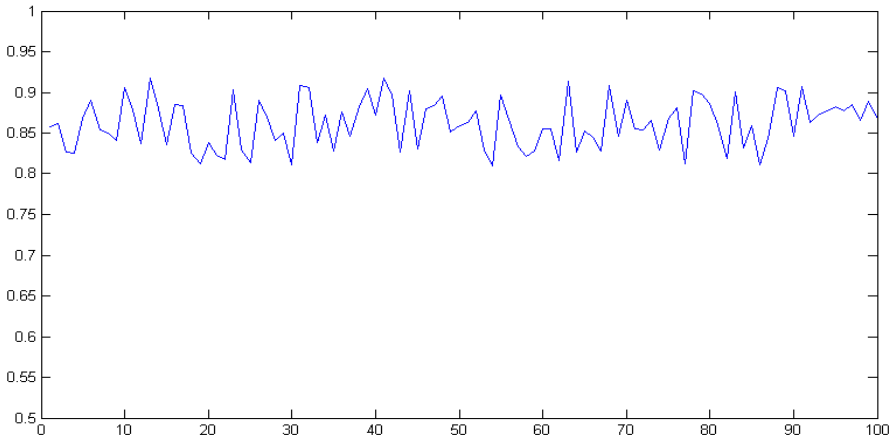

Figure 6. This is the Hiding Efficiency Curve of the Scheme

The privacy protection scheme based on the identity-based anonymous agent, as is defined above, has been successfully applied to our anti-DM project for preventing the leakage of buying track and the backtracking based on delivery address. The system simulation results for hiding efficiency are shown in Figure 6.

In the simulation experiments that agent for a thousand users, the average hiding efficiency range is between 81% and 92%. Experiment results show that the novel anonymous agent scheme is available for realizing the production, distribution and management of the anonymity identity in e-commerce and is reliable for ensuring the anonymous reliability, integrity and efficiency to realize privacy protection in small and medium- scale privacy agent of e-commerce. Consequently, further studies on large-scale privacy agent are continuing.

Acknowledgements

Professor Chunkuang Ma in Harbin Engineering University originally suggested the ideas on which this work was based on. Without the encouragement and dedication from Ms. Shusiang Chen and Mr. Yusueh Kowk and my husband Hongfeng Wang this thesis would not have been written. To them I offer my sincerest thanks and gratitude.

References

[1] V. S. Verykios, E. Bertino, I. N. Fovino, I. N. Provenza, Y. Saygin and Y. Theodoridis, “State-of-the-art in privacy preserving data mining”, CM SIGMOD Record, **(2004)**, pp. 50-57.
[2] R. Agrawal and R. Srikant, “Privacy preserving data mining”, Proceedings of the ACM SIGMOD Conference on Management of Data (SIGM OD), Dallas, Texas, (2000), pp. 439-450.
[3] A. C. Yao, “How to generate and exchange secrets”, Proceedings of the 27th IEEE Symposium on Foundations of Computer Science (FOCS), Toronto, Canada, **(1986)**, pp. 162-167.
[4] C. Clifton, M. Kantarcioglou, X. Lin and M. Y. Zhu, “Tools for privacy preserving distributed data mining”, ACM SIGKDD Explorations, **(2002)**, pp. 28-34.
[5] L. Sweeney, “K-anonymity:a model for protecting privacy”, International Journal on Uncertainty,Fuzziness and Knowledge-based Systems, **(2002)**, pp. 557-570.

[6] N. Li and T. Li, "t-closenes s: Privacy beyond k-anonymity and ldiversity", Proceedings of the 23rd Internat ional C on ference on Data Engineering (ICDE) . Is tanbul , Turkey, **(2007)**, pp. 106-115.
[7] C. C. Aggarwal, "On k-anonymity and the curse of dimensionality", Proceedings of the 31st Very Large Data Bases (VLDB)Conference. Trondheim, Norway, **(2005)**, pp. 901-909.
[8] Q. Wei Zhong, Z. De-Qing and J. Hai, "Research on privacy preservation mechanism for credentials and policies in grid computing environment", Journal of Computer Research and Development, (in Chinese), **(2007)**, pp. 11-19.
[9] Y. Xiao-Chun, L. Xiang-Yu, W. Bin and Y. Ge, "K-anonymization approaches for supporting mul tiple constraints", Journal of Software, (in Chinese), **(2006)**, pp. 1222-1231.
[10] A. C. Yao, "How to generate and exchang e secrets", Proceedings of the 27th IEEE Symposium on Foundations of Computer Science (FOCS), Toronto, Canada, **(1986)**, pp. 162-167.
[11] C. Clifton, M. Kant Arcioglou, X. Lin and M. Y. Zhu, "Tools for privacy preserving distributed datamining", ACM SIGKDD Explorations, **(2002)**, pp. 28-34.
[12] S. Merugu and J. Ghosh, "Privacy-preserving distributed clusteringusing generative models", Proceedings of the IEEE International Conference on Data Mining (ICDM), Melbourne, Florida, USA, **(2003)**, pp. 211-218.
[13] G. Miklau and D. Suciu, "A formal analysis of information disclosure in data exchange", Proceedings of the ACM SIGMOD Conference on Management of Data (SIGMOD). Maisondela Chimie, Paris, France, **(2004)**, pp. 575-586.
[14] A. Machanavajjhala and J. Gehrke, "On the efficiency of checking perfect privacy", Proceedings of the Sym posium on Principles of Database Systems (PODS), Chicago, Illinois, USA, **(2006)**, pp. 163-172.
[15] S. Merugu and J. Ghosh, "Privacy-preserving distribut edclusteringusing generative models", Proceedings of the IEEE International Conference on Data Mining (ICDM), Melbourne, Florida, USA, **(2003)**, pp. 211-218.
[16] R. Agrawal, R. Srikant and D. Thomas, "Privacy preserving OLAP", Proceedings of the ACM SIGMOD Conference on Management of Data (SIGMOD) . Baltimore, Maryland, **(2005)**, pp. 251-262.

Author

Dan Guo
College of Computer & Information Engineering
Harbin University of Commerce
1 Xue Hai Street, Song Bei district, Harbin city
China 150028
Phone (H) +86 045184892063 (Mobile) +86 13101605828
Email:hrbcucomputer@126.com