



“ANALYSIS OF SECURITY ISSUES AND CHALLENGES IN BIG DATA AMONG SMB’S”

Ms Neetu Kumari

M. Tech Research Scholar,
Department of Computer Science, Rajasthan
College of Engineering for Women,
Jaipur, Rajasthan

ABSTRACT

The current research study entitled “Analysis of Security Issues and Challenges with Big Data in SMB’s” . Innovation of new technology and its adoption contributes not only increasing the business values but as well facilities the business growth and agility to an organization. The development of the current big data is still faced with many problems especially security and privacy protection. Currently many organizations realize the big data security issues and actively take actions on big data information security problems. Information security is critical important for Internet enterprises. System security adopts techniques such as redundancy, network

separation, access control, authentication and encryption. Security issues are caused by openness, boundless, freedom of the networks, the key to solve such issues are making network free from them and turning network into controllable, manageable inner system. In current research paper respondents opinion were statistically analyzed with One Way ANOVA with the help of SPSS Software and the obtained P value was highly significant therefore the results concluded that null hypothesis H_0 : There is no significant relationship between Big data Security, risks and Benefits among SMB’s in India is rejected and alternate hypothesis which states that H_1 : There is a

significant relationship between Big data Security, risks and Benefits among SMB's in India is accepted and proved. This paper summarizes the characteristics of big data information security, and focuses on conclusion of security problems under the

big data field and the inspirations to the development of information security technology. Finally, this paper outlooks the future and trend of big data information security.

Key Words: - *Big Data, Data Security, Data Privacy, SMB's, Technology, India.*

1. INTRODUCTION:

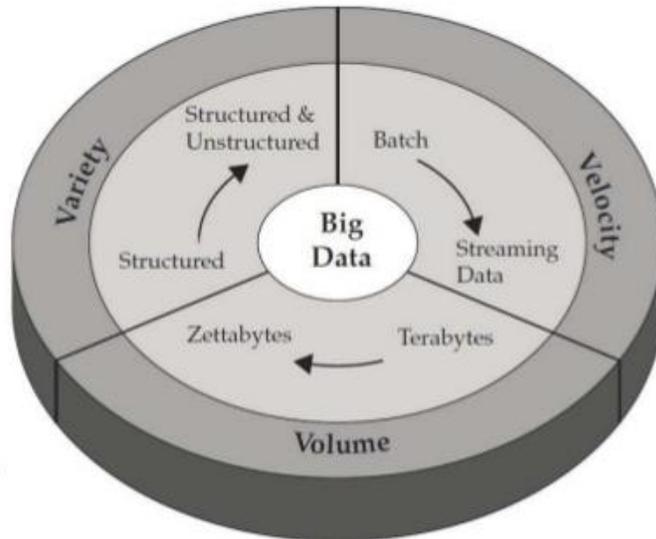
Big data is a collective term referring to data that is so large and complex that it exceeds the processing capability of conventional data management systems and software techniques. However with big data come big values. Data becomes big data when individual data stops mattering and only a large collection of it or analyses derived from it are of value. With many big data analyzing technologies, insights can be derived to enable better decision making for critical development areas such as health

care, economic productivity, energy, and natural disaster prediction.

The term Big Data appeared for the first time in 1998 in a Silicon Graphics (SGI) slide deck by John Mashey having the title Big Data and the Next Wave of Infra Stress. The first book mentioning Big Data is a data mining book that came to fore in 1998 too by Weiss and Indrukya. The first academic paper having the word Big Data in the title appeared in the year 2000 in a paper by Diebold.

Characteristics of Big data are elaborated in Figure 1 below:

- ▶ **Volume** – challenging to load and process (how to index, retrieve)
- ▶ **Variety** – different data types and degree of structure (how to query semi-structured data)
- ▶ **Velocity** – real-time processing influenced by rate of data arrival



The Indian SME industry has grown significantly over the years owing to a rise in better opportunities that can sustain the growth of such businesses. Be it the growing investment opportunities or greater adoption of innovative technology, Indian SMEs are now emerging as one of the most important market players in recent times. They are also expected to surpass traditional corporates in terms of growth and revenue generation within the next couple of years.

According to a report by The Ministry of Micro, Small and Medium Enterprises, the Indian SME sector has emerged as one of the fastest growing industries in the country over the past couple of decades. **This industry is also playing a vital role in**

facilitating employment generation and is helping in the industrialisation of rural and backwards areas, thus, reducing regional disparities, and assuring more equitable distribution of national income and wealth.

It has taken over 50 years of collective effort from both the government and the people of India to facilitate the growth of the SME sector so far. However, along the way, some industry experts identified **clearly defined the issue of Big data Security and Privacy within this space** that require immediate resolution in order to ease business operations in India

2. REVIEW OF LITERATURE

A detailed Literature has been reviewed to make the study relevant. Few key observations obtained from Literature cited are elaborated below:

Lindell et al. (2002) presents an improved implementation of the two party cases, using Yao's garbled circuits (GCs). Du and Atallah gives statement of SMC problem and various applications (Du, 2001b). This paper gives guidelines for SMC research with different applications where SMC can be applied efficiently.

Verykios et al. presents various approaches to protect sensitive rules during transaction processing. (Clifton, 2002) Clifton et al. gives tools for privacy preserving data mining; in this random number mechanism is used to preserve privacy of individuals. In this, if two parties collaborate they can get the data of third party (Verykios, 2003).

Agrawal et al. presented new protocols for different functions intersections, size and equi-join. And demonstrated that these protocols revealed insignificant information apart from what can be revealed from the query result. They presented a method to compute equi-join

size but this methodology outflows some information about tuples which are combined, on the basis of duplicates distribution (Agarwal, 2003).

Maurer presented the role of cryptography to achieve security in databases and addresses the issue of specifying and accomplishing confidentiality in a framework where the database is not fully trusted (Maurer, 2004). Verykios et al. gives an overview of privacy preserving data mining techniques. A detailed review and classification hierarchy of previous published work has been given (Verykios, 2004).

Zhan et al. (2004) present the randomized response techniques to perform privacy preserving data mining operations. In this paper authors considered multi-group i.e. attributes are partitioned in specific number of groups. Brickell et al. (2005) present a SMC based algorithm to compute shortest distance and secure union in the environment where parties are "honest but curious".

Trevathan (2005) present a model to conduct secure and anonymous online auctions. Methods are proposed to detect fraudulent in e-commerce. The proposed models have been implemented on online

auction server. It can be used for various real life online applications.

Karr et al. (2005) presented the case, when data is stored in distributed databases and regulated by various statistical organizations then what is the way of accomplishing, secure linear regression for “horizontally partitioned data”. They also proposed the methods for the records that use the secure sum protocols, MPC protocol, to find the least squares estimators for disjoint sets of data.

Liu et al. (2006) explore probability of using multiplicative random projection matrix for protecting distributed data privacy during data mining. Raymond et al. (2006) present (α, k) anonymity prototype to protect identification and associations of critical information in data. In this paper, quasi-identifier and equivalence class concepts are used for global and local recoding. This work, experiments different variables set and comparative study of the result.

3. RESEARCH METHODOLOGY

RESEARCH METHODOLOGY		
Objectives of Research	of	<ul style="list-style-type: none"> • To analyze issues and Challenges of Big Data Security among SMB’s. • To provide insight on the effects of perceived risks, requirements and benefits in big data security for SMB’s.
Hypothesis of Research	of	<i>H₀</i> : There is no significant relationship between Big data Security, risks and Benefits among SMB’s in India.
Research Design		Exploratory – To know the parameters and formulate the hypotheses. Analytical – To analyze the parameters found out.
Selected SMB’s under study		Dr B Lal Clinical Labrotory Private Limited, Jaipur Gravita India Limited, Jaipur Elektrolites (Power) Pvt. Ltd., Jaipur
Sampling Design		Stratified Random sampling Method
Sample Size		(a)Employees of Selected SMB’s (Sample Size= 300)
Data collection Techniques		Primary Data collection – A framed set of questionnaire for customers of two wheelers in Rajasthan Secondary Data Collection – Research reports of IT Companies, SMB, Big Data Entrepreneurs Companies, Annual reports, Management books, journals, research papers etc.

Analytical tools For Pilot Study	Cronbach's alpha for reliability and Kaiser Meyer's Rank Test for Variability
Statistical Analysis for hypothesis testing	Chi Square Test, Multivariate ANOVA, Students't' test.

4. PROBLEM STATEMENT

With the fabulous development of information technology, big data application prompts the development of storage, network and computer field. It also brings new security problems.

The development of the current big data is still faced with many problems especially security and privacy protection. Currently many organizations realize the big data security issues and actively take actions on big data information security problems. Information security is critical important for Internet enterprises.

This security challenge caused by big data has attracted the attention of information security and industrial community domain. This paper summarizes the characteristics of big data information

security, and focuses on conclusion of security problems under the big data field and the inspirations to the development of information security technology. Finally, this paper outlooks the future and trend of big data information security.

5. RESULTS AND ANALYSIS

5.1 Demographic Details of respondents

Demographic study means study of both quantitative and qualitative aspects of selected human population. Quantitative aspects include composition, age, gender, size, and structure of the population. Qualitative aspects are the research specific factors such as current usage of big data. Demographic variables of current research study are evaluated in table 1 below.

TABLE 1 DEMOGRAPHIC DETAILS OF CUSTOMERS AS RESPONDENTS

Sample characteristic	Category	No of Respondents HERO(N=300)
Gender	Male	69%
	Female	31%
Age Group (Years)	18-25	16%
	26-30	24%
	31-35	31%
	36-40	19%
	Above 40	10%

Company size	Percentage
1 - 10	4.55%
10 -50	27.27%
50 - 100	4.55%
100 -250	9.09%
higher than 250	54.55%
Total	100%
Business Sector	Percentage
Finance	50.00%
ICT sector	27.27%
Consumer goods	9.09 %
Other	13.64%
Total	100%
Role in the company	Percentage
Owner	4.55%
IT manager	31.82%
Employee	63.64%
Total	100%

In short, small and mid-sized businesses have focused an appropriate level of concern on the business risks that affect those most. The next-and far more important-question is how they will reduce those concerns, address the risks by actually implementing solutions to protect their businesses

5.2 Analysis of Security issues of Big data in SMB

The protection gap

But despite awareness of the risks they face and clarity about the best ways to mitigate them, a striking number of small and mid-sized businesses not only trail the state of the art, but lack even the most basic protection for their business information. Figure 2 shows the status of planning and implementation across the segment

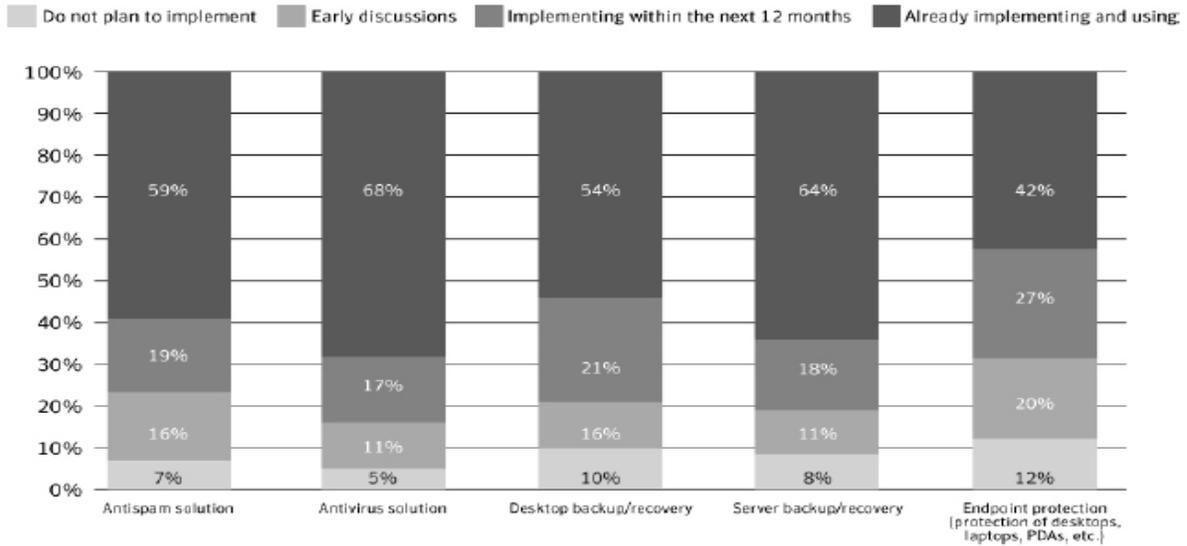


Figure 2 : Status of Planning and Implementation across the Segment

More than a third of these businesses operate with no protection against viruses and spam. Many others are protected only by half-measures: backup/recovery for servers but not desktops for example, or antivirus point solutions that can't protect mobile endpoints or defend against fast-changing, fast-moving, complex threats that use multiple techniques to attack digital assets. As Ray Boggs, Vice-President of

SMB research for IDC puts it, "Of course SMBs know better, but they are too often focused on business opportunities outside the company to pay attention to the risks they are taking right at home."

What's stopping them? Through the survey, SMBs report the familiar constraints of staffing, time, and budget, as shown in Figure 3

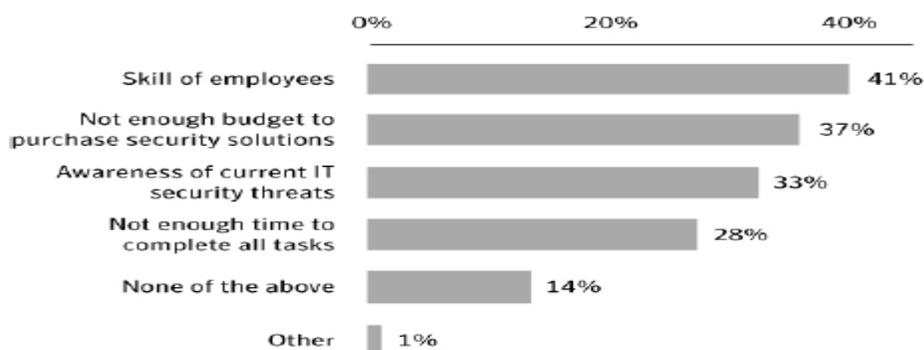


Figure 3 Root Cause for the protection Gap

These limits are especially severe at businesses that lack even a single in-house dedicated IT staff member-true for 42% of survey participants. What's more, their median IT security and storage budgets hover around \$4,500 per year-barely enough to keep up with obsolescence, much less growth. The survey did reveal one promising trend-despite downdrafts throughout the economy as a whole, 90% of SMB survey participants reported their IT security and storage budgets trending up, or at least not in decline.

Consequences

To assess the economic scale of the risks these firms face, the survey asked participants who had suffered security breaches or data loss to inventory the conditions responsible. And as Figure 3 details, those conditions strongly resemble the inventory of SMB security risks reported in Figure 4:

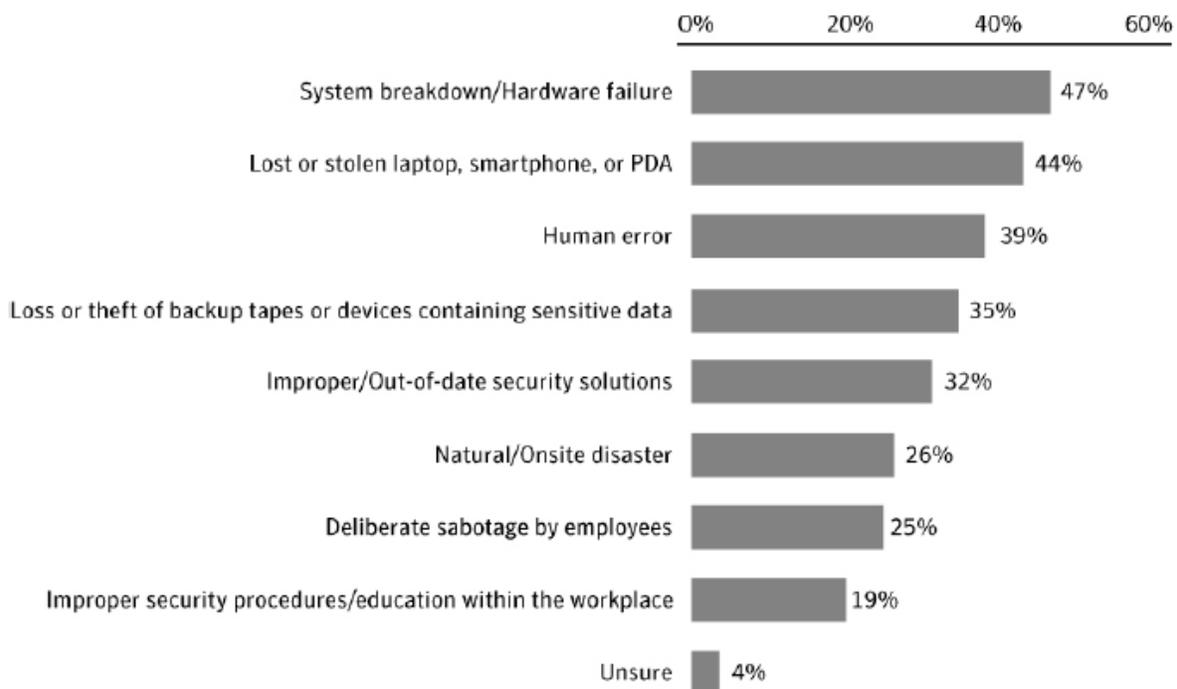


Figure 4 Causes Cited for Security breach of Data Loss

Not surprisingly, a cross-comparison of Figure 4 against Figure 2 shows that security incidents and data losses are concentrated exactly where SMB gaps and shortfalls leave vulnerabilities. Here are just a few examples:

- 44% of breaches involve compromised mobile devices-endpoints like laptops and PDAs that are overlooked in many SMB security plans
- 39% involve sabotage, human error, or poor procedures-while SMB security concentrates on threats from outside the network

:

- 35% of breaches involve failures in backup processes-known vulnerabilities for SMB servers, desktops, and laptops Solid, regular backup practices can mitigate the risks of inevitable hardware failures, but lapses may have serious consequences

Hypothesis Testing

In current research study on results obtained above of respondents the values were statistically analyzed above Likert's scale values with one way ANOVA by using SPSS and results are as mentioned

ANOVA							
GROUP			Sum of Squares	df	Mean Square	F	P Value (Sig)
SMB's	Big Data Security	Between Groups	1.472	5	.294	1.320	.005
		Within Groups	58.610	5	.292		
		Total	60.082	5			
	Big Data Risk & Benefits	Between Groups	5.322	5	1.064	2.162	.007
		Within Groups	57.118	5	.284		
		Total	62.440	5			

Respondents opinion were statistically analyzed with One Way ANOVA with the help of SPSS Software and the obtained P value was highly significant therefore the results concluded that null hypothesis H_0 : There is no significant relationship between Big data Security, risks and Benefits among SMB's in India is rejected and alternate hypothesis which states that H_1 : There is a significant relationship between Big data Security, risks and Benefits among SMB's in India is accepted and proved.

Few observations based on respondents reported for protection of Big data in SMB are as described:

- **Protection principles**

Losses and business risks like these are not necessary. Even when staff, time, and budget constraints stand in the way of a systematic solution, small and mid-size businesses can improve their security posture with simple, cost-effective protection measures like these:

- **Stay informed**

Some of the best things in IT security are free. Information resources won't keep technical defenses up to the minute, but periodic reports like the Symantec Internet Security Threat Report can keep even the smallest business aware

of trends in the threat environment, and how best to defend against them.

- **Back up data**

They may be tedious and time-consuming, but backups-even manual backups-offer some of the highest returns available among IT initiatives. Protection against natural disaster, hardware failure, and above all human error gives a business continuity and confidence in the face of a wide range of risks. Include off-site storage of encrypted data as part of a mature backup and recovery program.

- **Protect from the inside**

Employee error, fraud, and vandalism can compromise a company's most sensitive and valuable information-and legally required disclosures can savage its reputation. Simple policies and controls-starting with elimination of duplicate or portable data stores-can substantially improve your security posture. The Payment Card Industry offers excellent guidance on data protection, appropriate for members and nonmembers alike.

- **Don't forget physical security**

By far the oldest component of data protection, physical security still ranks high in importance. Policies for screen-locking, end-of-day shutdown, asset tagging and

tracking, and others are easy to implement—often as simple as keeping the right doors locked. And every one of them cuts the odds of the worst-case data-loss scenario, when a device containing critical data falls under a thief's control for an extended time with low chance of exposure.

When the time comes to invest in upgrading your electronic protection, here are three additional principles to consider:

- **Use layered security**

Threats escalate, and even sophisticated protections can fail against new attacks. Multi-layer defenses protect against local breakthroughs or single-point failures of any one technology or method. The latest defense-in-depth strategies combine antivirus and anti spam software with firewalls, intrusion prevention, device and application control, and patch management solutions.

- **Deploy comprehensive security**

Depth is critical, but don't neglect breadth. Security plans should cover desktops, laptops, and messaging servers. Mobile devices—whether carried in by outsiders or taken out by employees—are the most difficult to protect. But new endpoint protection technologies quarantine connections until new devices demonstrate

compliance with all relevant security policies, and ensure that security products are regularly updated to block new threats.

- **Use solution providers for needed expertise**

You are exposed to a single company's security and threat environments, but your local solution provider sees tens—even hundreds. If staffing and time constraints are keeping you from effective information protection, your local IT consultant or reseller can help you explore a cost-effective way forward.

6. CONCLUSION AND RECOMMENDATIONS

To handle big data and to work with it and obtaining benefits from it a branch of science has come up and is evolving, called Data Science. Data Science is the branch of science that deals with discovering knowledge from huge sets of data, mostly unstructured and semi structured, by virtue of data inference and exploration.

Symantec's latest security offering, Symantec Protection Suite (SPS), is designed, scaled, and priced to meet SMB security and data protection requirements.

The only suite offering comprehensive protection across laptops, servers, messaging gateways, and backup and recovery environments, SPS delivers proven protection for business information and computers, helps defend against aggressive new malware and spam threats, and backs up and quickly recovers computers and information in the event of a problem.

Information security in big data environment is a promising field in

information security. This paper introduces impact to information security from SMB's aspects of big data security. In general, improving system efficiency and provide general cloud storage functions on premise to ensure user data and access authority are the research direction of future safe big data utility and computing. At present, more things need to be done in cryptograph searching and reduplicate data removing.

REFERENCES

Ahmed, M., Chowdhury, A. S. M., Ahmed, M., Rafee, M. H. et al. (2012), 'An advanced survey on cloud computing and state-of-the-art research issues.', *International Journal of Computer Science Issues (IJCSI)* 9(1).

Ahronovitz, M., Amrhein, D., Anderson, P., de Andrade, A., Armstrong, J., Arasan, B., Bartlett, J., Bruklis, R., Cameron, K. and Carlson, M. (2011), 'Cloud computing use cases white paper'.

Aljabre, A. (2012), 'Cloud computing for increased business value',

International Journal of Business and Social Science 3(1), 234{239.

Arendt, L. (2008), 'Barriers to ict adoption in smes: how to bridge the digital divide?', *Journal of Systems and Information Technology* 10(2), 93{108.}

Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkatasubramanian. L-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data*, 2007, 1(1):1-52

- Bessani A, Correia M, Quaresma B, et al. DEPSKY: Dependable and secure storage in a cloud-of clouds [C] //proc of the 6thConf on Computer System. New York: ACM, 2011:31-46
- Chen Mingqi, Jiang He. USA Information Network Security New Strategy Analysis in Big Data [J]. Information Network Security. 2012(8):32—35
- Goel S., Hofman J.M., Lahaie S., Pennock D.M. and Watts D.J.. Predicting consumer behavior with Web search. National Academy of Sciences, 2010, 7 (41): 17486– 17490
http://www.wired.com/science/discoveries/magazine/16-07/pb_theory
- Lei Zou, Lei Chen and M. Tamer zsu. k-automorphism: a general framework for privacy preserving network publication. // Proceedings of the 35th International Conference on Very Large Data Bases (VLDB'2009), Lyon, France, 2009: 946-957
- Mao Ye, Peifeng Yin, Wang-Chien Lee, and Dik-Lun Lee. Exploiting geographical influence for collaborative point-of-interest recommendation.//Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval(SIGIR'11), Beijing, China, 2011: 325-334
- Meng Xiao-Feng, Ci Xiang. Big Data Management: Concepts, Techniques and Challenges. Journal of Computer Research and Development, 2013, 50(1): 146-169 (in Chinese)
- Narayanan A, Shmatikov V. How to break anonymity of the Netflix prize dataset. ArXiv Computer Science e-prints, 2006, arXiv:cs/0610105: 1-10
- Study Finds Web Sites Prying Less: Shift May Reflect Consumer Concerns[EB/OL]. <http://www.CNN.com>, 2002-03-18A survey of data disclosing in 2010 by Verizon[EB/OL].[2012-05-10].
- Sweeney L..K-Anonymity: a model for protecting privacy. InternationalJournal on Uncertainty, Fuzziness and Knowledge-based Systems, 2002, 10 (5): 557-570
- Sweeney L..K-Anonymity: Achieving k-Anonymity Privacy Protection using Generalization and Suppression.
- Ying X. and Wu X.. Randomizing social networks: a spectrum preserving

approach. //Proceedings of the SIAM
International Conference on Data
Mining (SDM'08), Georgia, USA,
2008: 739-750