"Exploratory Study Big Data Security Analysis among SMB's"

Mrs Sonal Saxena [1], Ms Neetu Kumari [2]

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

Abstract -The industrial addition in a backward area can only be achieved by the quick development and promotion of small scale industries. Besides economic aspects, the social role of small scale and cottage industrial units are quite compelling in achieving different social goals such as removal of poverty, completion of self-reliance, attrition in disparities in income, wealth and standard of living and regional imbalances. Small and medium-sized enterprises (SMBs) play a important role in the economic development of nations. ***Therefore, it is vital to evaluate the Security of Big Data of SMBs to support that role.***

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.*

*****

I. INTRODUCTION:

Security and confidentiality are always collaborated together, but actually they are different. Basically, security means that data access is inhibited, and safe from attacks. It helps to contain confidentiality. Without security one cannot assure confidentiality. Confidentiality concern, deals with accredited usage of personal delicate data by authorized personnel and prevention from baleful participants during the data access and computing.

Due to digitization the data inflow-outflow is hiked over the limit. These data are kept in depots for undefined time durations. This data could be related to health, wealth or business. Here, the problem is accordant as, when personal data is considered discrete and corporations have various reasons to protect it. For ex. an individual may wish to keep his patrimonial health record personally as communicating it may lead to social negligence.

If this information is available digitally in the form of plain text, it may lead to emotional or personal outages. So it is necessary to keep the data confidential and secure.

Due to the large rise in the need for using Web applications all over the world, there have been big efforts from programmers to advance and apply new Web applications to be utilized by companies. In our study, we will pay handy attention to the security and effect research of Big data incisive taken by Indian government. We divide accordant efforts into representative divisions while maintaining our own independent understandings.

In the Big Data analysis context, so called analytics over Big Data is playing a major role. Analytics cover a large family of difficulties mainly arising in the context of Database, Data Warehousing and Data Mining research. Analytics discovery is intended to flourish complex procedures running over large-scale, huge in-size data repositories with the aim of avulsing useful knowledge hidden in such repositories. One of the most main application scenarios where Big Data arise is, without doubt, experimental computing. Here, scientists and researchers outgrowth huge amounts of data per-day via experiments (e.g., disciplines like high-energy physics, astronomy, biology, bio-medicine, and so forth). But extracting useful knowledge for decision making ambition from these massive, large-scale data repositories is almost impossible for actual DBMS-inspired research tools. From a methodological point of view, there are also research difficulties. A new methodology is needed for converting Big Data stored in composite and different-in-nature data sources (e.g., legacy systems, Web, scientific data repositories, sensor and stream databases, social networks) into an efficient hence well-illustrable format for aim the data analytics. As a aftereffect, data-driven approaches, in biology, medicine, public policy, social sciences, and humanities, can replace the traditional hypothesis-driven research in science.

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 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.**

II. 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.

III. RESEARCH METHODOLOGY

RESEARCH METHODOLOGY	
Objectives of Research	• 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	H_0: 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.

IV. 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.

V. 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

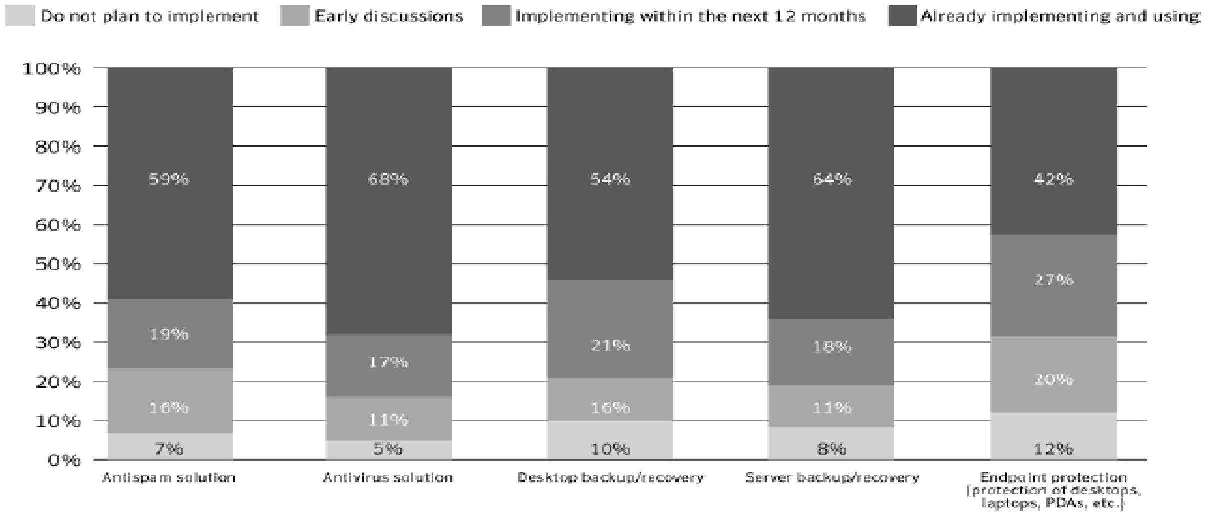

Figure 2 : Status of Planning and Implementation across the Segment

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

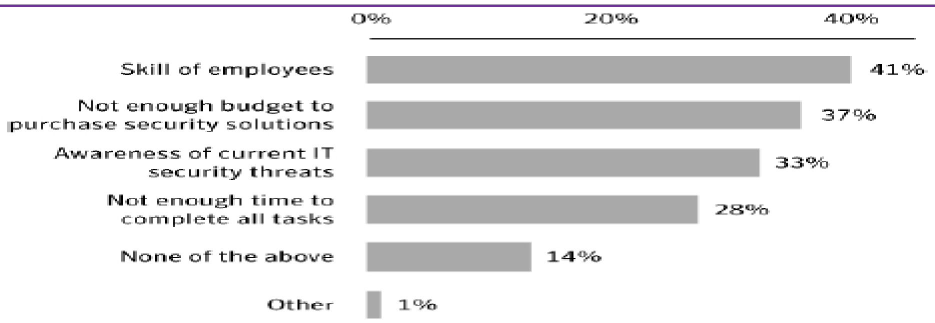

Figure 3 Root Cause for the protection Gap

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:

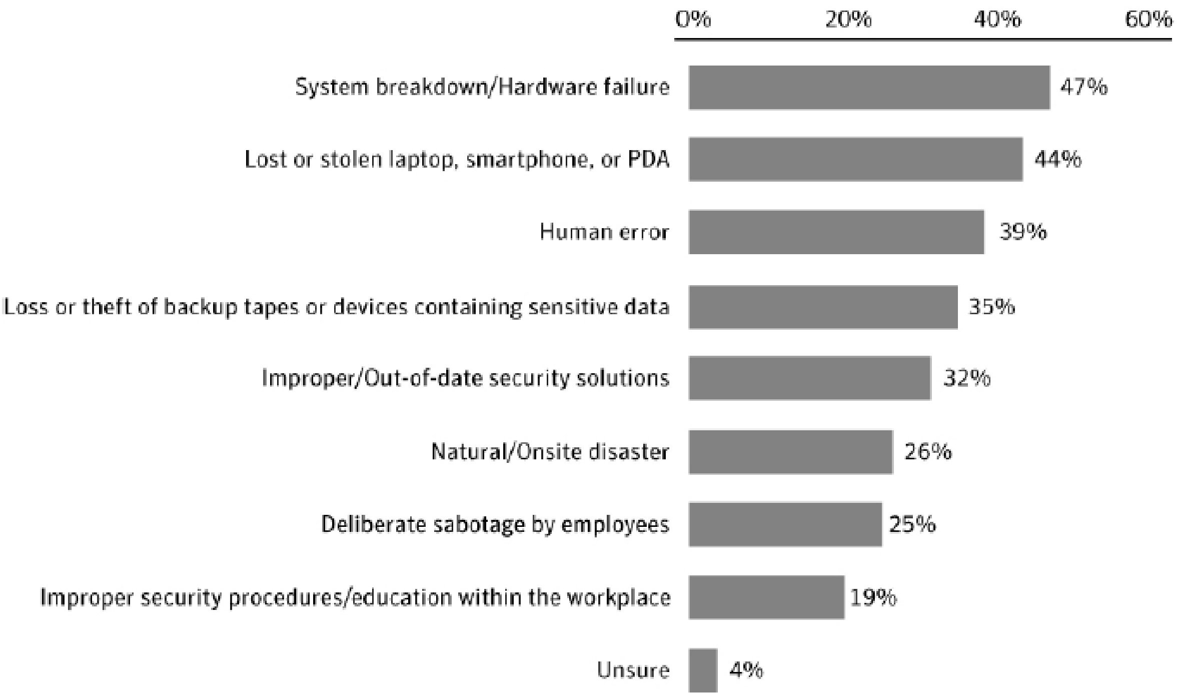

Figure 4 Causes Cited for Security breach of Data Loss

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 antispam 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.

VI. CONCLUSION AND RECOMMENDATIONS

Big Data is changing the way we perceive our world. The impact big data has created and will continue to create can ripple through all facets of our life. Global Data is on the rise, by 2020, we would have quadrupled the data we generate every day. This data would be generated through a wide array of sensors we are continuously incorporating in our lives. Data collection would be aided by what is today dubbed as the “Internet of Things”. Through the use of smart bulbs to smart cars, everyday devices are generating more data than ever before. These smart devices are incorporated not only with sensors to collect data all around them but they are also connected to the grid which contains other devices. A Smart Home today consists of an all encompassing architecture of devices that can interact with each other via the vast internet network. Bulbs that dim automatically aided by ambient light sensors and cars that can glide through heavy traffic using proximity sensors are examples of sensor technology advancements that we have seen over the years. Big Data is also changing things in the business world. Companies are using big data analysis to target marketing at very specific demographics. Focus Groups are becoming increasingly redundant as analytics firms such as McKinsey are using analysis on very large sample bases that have today been made possible due to advancements in Big Data. The potential value of global personal location data is estimated to be $700 billion to end users, and it can result in an up to 50% decrease in product development and assembly costs, according to a recent McKinsey report. Big Data does not arise out of a vacuum: it is recorded from some data generating source. For example, consider our ability to sense and observe the world around us, from the heart rate of an elderly citizen, and presence of toxins in the air we breathe, to the planned square kilometer array telescope, which will produce up to 1 million terabytes of raw data per day.

REFERENCES

[1]. 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).

[2]. 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'.

[3]. Aljabre, A. (2012), `Cloud computing for increased business value', International Journal of Business and Social Science 3(1), 234{239.

[4]. 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.}

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

[6]. 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

[7]. Chen Mingqi, Jiang He. USA Information Network Security New Strategy Analysis in Big Data [J]. Information Network Security. 2012(8):32—35

[8]. 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

[9]. http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

[10]. 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

[11]. 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

[12]. 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)

[13]. Narayanan A, Shmatikov V. How to break anonymity of the Netflix prize dataset. ArXiv Computer Science e-prints, 2006, arXiv:cs/0610105: 1-10

[14]. 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].

[15]. Sweeney L..K-Anonymity: a model for protecting privacy. InternationalJournal on Uncertainty, Fuzziness and Knowledge-based Systems, 2002, 10 (5): 557-570

[16]. Sweeney L..K-Anonymity: Achieving k-Anonymity Privacy Protection using Generalization and Suppression.