"AN IMPERATIVE ANALYSIS OF SECURITY ISSUES AND CHALLENGES WITH BIG DATA IN SMB'S"

Ms. Reena Singh
Research Scholar,
Department of Computer Science, Suresh Gyan Vihar
University,
Jaipur (Rajasthan)

Prof (Dr). Dinesh Goyal
Director Center for Cloud Infrastructure and Security and
Principal, Engineering,
Suresh Gyan Vihar University,
Jaipur (Rajasthan)

Abstract: The current research study entitled "An Imperative 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. MAny organizations including leading SMB's of India understand the concept of security issues of big Data and started taken sincere actions against the same. Information security is critical important for Internet enterprises. System security adopts techniques such as redundancy, network separation, access control, authentication and encryption. 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. Thus the current research endevour concludes the various linguistic characteristics of big data analytics and security concerns of big data, which finally recommends the security techniques and its utilities among all the developing Big data Analytics especially in SMB's in India. Concluding these researches will be helpful postulating future trends in Big Data Analytics Security issues.

KeyWords:- 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. As every thing has its own Pros and Cons , Big data also have big problems and benefits.. "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."[1]
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. [2] 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
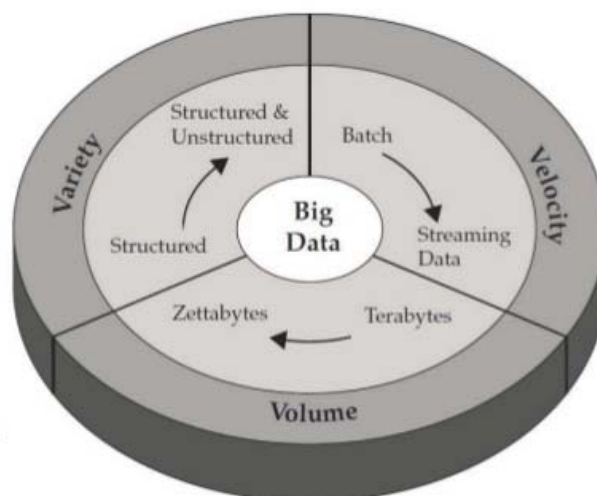
Figure 1 Characteristics of Big data

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.[3]

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. [4]

With the remarkable improvement of data innovation, huge information application prompts the advancement of capacity, system and PC field. It additionally brings new security issues. This security challenge caused by huge information has pulled in the consideration of data security and mechanical group area. This paper outlines the attributes of enormous information data security, and concentrates on finish of security issues under the huge information field and the motivations to the advancement of data security innovation. At long last, this paper viewpoints the future and pattern of huge information data security.

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) [5] presents an enhanced usage of the two party cases, utilizing Yao's confused circuits (GCs). Du and Atallah gives proclamation of SMC issue and different minimum squares estimators for disjoint arrangements of data.[12]

applications (Du, 2001b). This paper gives rules for SMC look into with changed applications where SMC can be connected efficiently.[6]

Agrawal et al. exhibited new conventions for various capacities crossing points, size and equi-join. Furthermore, exhibited that these conventions uncovered unimportant data separated from what can be uncovered from the inquiry result. They introduced a technique to process equi-join measure yet this approach outpourings some data about tuples which are consolidated, on the premise of copies appropriation (Agarwal, 2003).[9]

Maurer displayed the part of cryptography to accomplish security in databases and addresses the issue of determining and finishing classification in a structure where the database isn't completely confided in (Maurer, 2004). Verykios et al. gives an outline of protection saving information mining methods. A point by point audit and arrangement progression of past distributed work has been given (Verykios, 2004).[10]

Zhan et al. (2004) exhibit the randomized reaction systems to perform protection saving information mining operations. In this paper creators considered multi-amass i.e. properties are parceled in particular number of gatherings. Brickell et al. (2005) display a SMC based calculation to process most brief separation and secure union in the earth where parties are"honest however inquisitive".

Trevathan(2005) display a model to lead secure and mysterious online closeouts. Strategies are proposed to recognize false in web based business. The proposed models have been actualized on online sale server. It can be utilized for different genuine online applications.[11]

Karr et al. (2005) introduced the case, when information is put away in conveyed databases and controlled by different measurable associations then what is the method for achieving, secure direct relapse for "on a level plane parceled information". They likewise proposed the techniques for the records that utilization the safe total conventions, MPC convention, to locate the

3. 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.

4. RESULTS AND ANALYSIS

4.1 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 status of two wheeler, etc. 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. Another important query arises is the mechanism to reduce these security of big data concerns and predicting and postulating the suggestive ways which can be beneficial for protecting the business specially its big data security. These issues are elaborated below:

4.2 Security challenges and issues investigation among SMB's[19]
Be that as it may, regardless of attention to the dangers they face and lucidity about the most ideal approaches to relieve them, a striking number of little and moderate sized organizations trail the best in class, as well as need even the

most fundamental security for their business data. Figure 2 demonstrates the status of arranging and execution over the fragment [14]
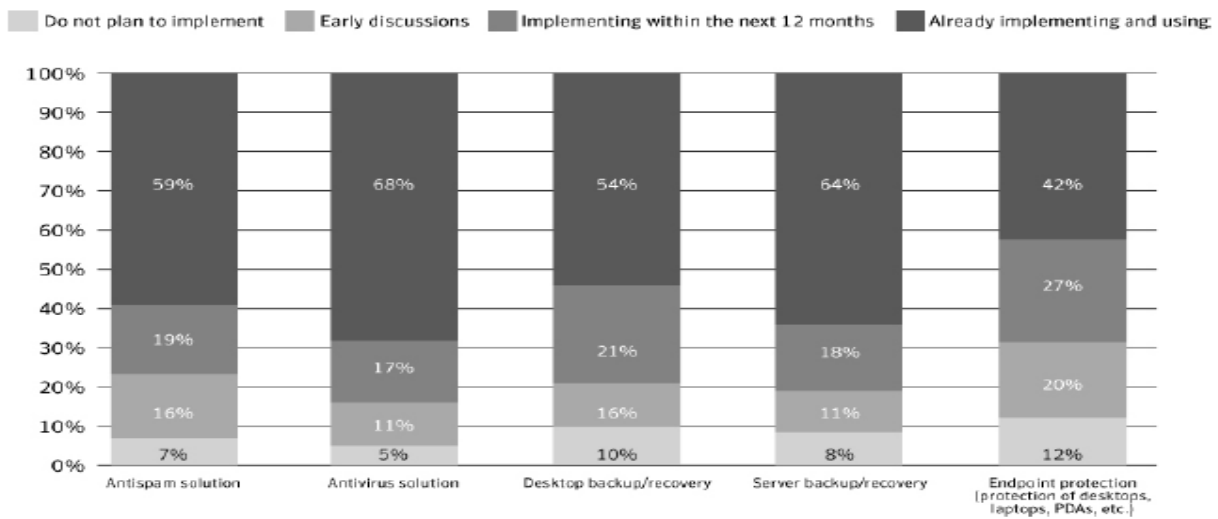

Figure 2 : Status of Planning and Implementation across the Segment

It can be stated that more than 33% of these organizations work with no assurance against infections and spam. Numerous others are ensured just considerably measures: reinforcement/recuperation for servers however not desktops for instance, or antivirus point arrangements that can't secure portable endpoints or safeguard against quick evolving, quick moving, complex dangers that utilization different systems to assault advanced resources. As Ray Boggs, Vice-President of SMB explore for IDC puts it, "obviously SMBs know better, yet they are time and again centered around business openings outside the organization to focus on the dangers they are taking comfortable." What's stopping them? Through the survey, SMBs report the familiar constraints of staffing, time, and budget, as shown in Figure 3
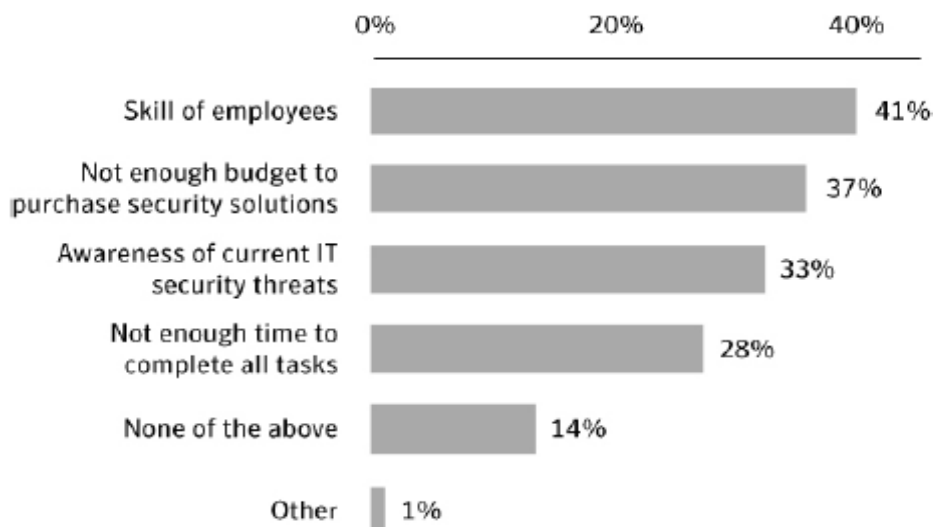

Figure 3 Root Cause for the protection Gap

5. IMPACT STUDY

To evaluate the monetary size of the dangers these organizations confront, the overview asked members who had endured security breaks or information misfortune to stock the conditions dependable. What's more, as Figure 3 points of interest, those conditions emphatically look like the stock of SMB security dangers detailed in Figure
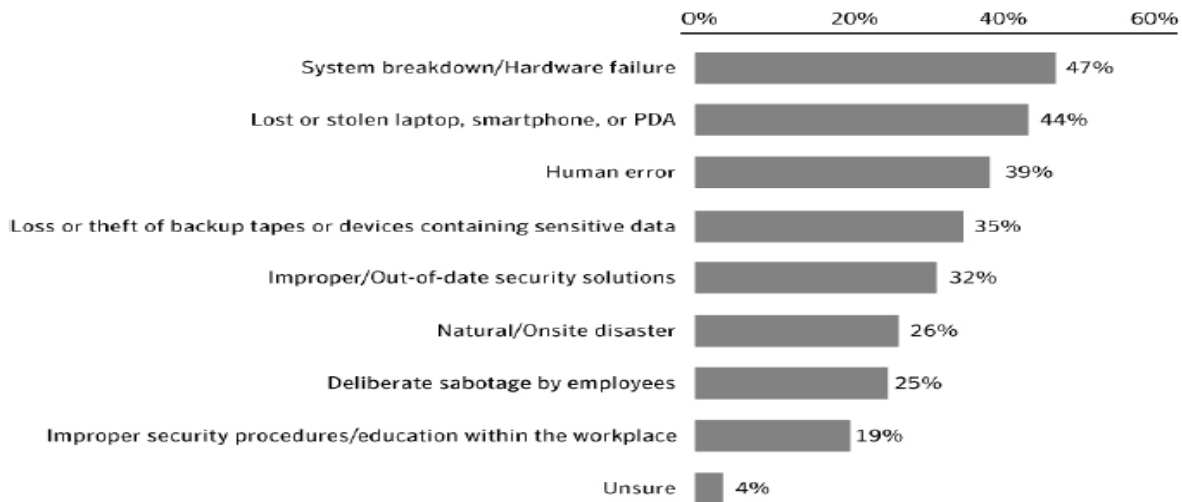
2:

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:[18]
• **Standards of Protection**
  Misfortunes and business dangers like these are a bit much. Notwithstanding when staff, time, and spending limitations obstruct an efficient arrangement, little and

medium size organizations can enhance their security pose with straightforward, financially savvy assurance measures may have impact.

• **Awareness about Security Details**
  A portion of the best things in IT security are free. Data assets won't keep specialized protections up to the moment, yet occasional reports like the Symantec Internet Security Threat Report can keep even the littlest business mindful of patterns in the danger condition, and how best to guard against them.

• **Data endorsement**
  They might be dull and tedious; however reinforcements even manual reinforcements offer a portion of the most noteworthy returns accessible

among IT activities. Insurance against cataclysmic event, equipment disappointment, or more all human mistake gives a business congruity and trust despite an extensive variety of dangers. Incorporate off-site stockpiling of scrambled information as a feature of a develop reinforcement and recuperation program.

- **Security And Protection Within Organization**
Worker mistake, misrepresentation, and vandalism can bargain an organization's most touchy and important data and lawfully required divulgences can savage its notoriety. Basic strategies and controls-beginning with end of copy or convenient information stores-can generously enhance your security act. The Payment Card Industry offers superb direction on information security, fitting for individuals and nonmembers alike.

- **Security in Physical Form**
By a long shot the most established part of information insurance, physical security still positions high in significance. Strategies for screen-locking, end-of-day shutdown, resource labeling and following, and others are anything but difficult to execute regularly as basic as keeping the correct entryways bolted. Furthermore, every one of them cuts the chances of the most pessimistic scenario information misfortune situation, when a gadget containing basic information falls under a criminal's control for an expanded time with low shot of presentation.

At the point when the time comes to put resources into overhauling your electronic security, here are three extra standards to consider:

- **Layered Security Mechanism**
Dangers heighten, and even advanced securities can come up short against new assaults. Multi-layer barriers secure against neighborhood leaps forward or single-point disappointments of any one innovation or technique. The most recent protection top to bottom techniques join antivirus and antispam programming with firewalls, interruption counteractive action, gadget and application control, and fix administration arrangements.

- **Installation of wide-ranging of Security Mechanism**
Profundity is basic, yet don't disregard broadness. Security designs should cover desktops, portable PCs, and informing servers. Cell phones whether conveyed in by pariahs or taken out by workers are the most hard to ensure. However, new endpoint insurance advances isolate associations until the point when new gadgets exhibit consistence with all significant security strategies, and guarantee that security items are routinely refreshed to piece new dangers.

- **Request for experts advice and suggestion on Security Issues**
You are presented to a solitary organization's security and risk conditions; however your nearby arrangement supplier sees tens-even hundreds. On the off chance that staffing and time limitations are keeping you from successful data assurance, your neighborhood IT advisor or affiliate can enable you to investigate a financially savvy route forward.

5. 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.[15]

Symantec's most recent security offering, Symantec Protection Suite (SPS), is outlined, scaled, and evaluated to meet SMB security and information assurance prerequisites. The main suite offering far reaching assurance crosswise over tablets, servers, informing doors, and reinforcement and recuperation situations, SPS conveys demonstrated insurance for business data and PCs, guards against forceful new malware and spam dangers, and moves down and rapidly recoups PCs and data in case of a problem.[16,17]

Data security in huge information condition is a promising field in data security. This paper acquaints affect with data security from SMB's parts of huge information security. As a rule, enhancing framework effectiveness and give general distributed storage works on start to guarantee client information and access specialist are the examination bearing of future safe enormous information utility and registering. At present, more things should be done in cryptograph looking and reduplicate information evacuating.

REFERENCES

i. 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)

ii. 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).

iii. 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

iv. 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

v. Lindell K .L., Anderson, P., de Andrade, A., Armstrong, J., Arasan, B., Bartlett, J., Bruklis, R., Cameron, K. and Carlson, M. (2002), `Cloud computing use cases white paper'.

vi. Due, A. (2001), `Cloud computing for increased business value', International Journal of Business and Social Science 3(1), 234{239.

vii. Clifton, L. (2002), `Barriers to ict adoption in smes: how to bridge the digital divide?', Journal of Systems and Information Technology 10(2), 93{108.}

viii. Varikios Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. L-diversity: Privacy beyond k-anonymity. ACM Transactions on Knowledge Discovery from Data, 2003, 1(1):1-52

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

x.   Agarwal 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, 2003:31-46

xi.   Zhan 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

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

xiii.   Raymond C.L. (2006) 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].

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

xv.   http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

xvi.   Sweeney L..K-Anonymity: Achieving k-Anonymity Privacy Protection using Generalization and Suppression.

xvii.   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

xviii.   Zhang Hongjun, Hao Wenning, He Dengchao & Mao Yuxing (2014); Survey of Research on Information Security in Big Data; PLA university of Industry and Technology Nan Jing, China; Brazilian Work Shop on Social Netqwork Analysis and Mining.

xix.   SMB Survey (2009) White paper; SMB Protection Gap,SMB security and data protection: survey shows high concern, less action. Symantec World Headquarters 20330 Stevens Creek Blvd. Cupertino, CA 95014 USA