PROBLEMS AND PROSPECTS OF BIG DATA: SECURITY ASPECTS

Arpita Banerjee[1] and C. Banerjee[2]

[1]Assistant Professor, St. Xavier's College, Jaipur

[2]Assistant Professor, Amity Institute of Info. Technology, Amity University, Jaipur

Abstract

The prevalence of computing and electronic communication technologies has led to the fastest growth of data from both digital and analog sources. New skills to collect, evaluate, distribute and keep secure massive amounts of data increases new fears about the nature of secrecy and to the means by which individual privacy might be conceded or protected. The term 'privacy' incorporates not only avoiding observation or keeping one's individual matters and relationships as secret, but also the capability to share information selectively but not publicly. The promise of big data collection and analysis says that the derived data can be used by the businesses for decision-making processing so that both the individual as well as the society can be benefitted. Threats to privacy emerge from intentional or unintentional disclosure of personal data. This paper investigates problems faced by big data due to lack of inbuilt security. The author analyzes the security problem based by the big data irrespective of its use.

Keywords : Big Data, Security, Threats, Big Data Analysis

Introduction

Today's environment is a threat based environment which mainly enforces the three Vs of big data: volume, variety, and velocity. All the three Vs are growing at an unexpected rate and have required a mechanism of how security providers can maintain threats. The term big data refers to rapid growth in large compilation of data or a set of collection of data whose focus is beyond the capacity of usual database software tools to detect, accumulate, manage and scrutinize. This type of data can be in a well organized, disorganized, refined or unprocessed form coming from diversified sources. It is big in the scale of analysis that can be tested to the various types of data, basically to make assumption and draw result. [1]. Everyday approximately around 2.5 quintillion bytes of data is being collected which is nearly 90% of the data in the world. It is widely noted that the threats inherent in big data are being resist by both the public and private sectors. [1][2][7]

Consistent big data include data and metadata collected by units like government, private sectors and individuals etc. It not only includes proprietary and open data, but also data about individuals collected incidentally or accidentally inthe course of other activities. In a privacy context, the term 'big data' typically means data about one or a group of individuals or that might be analyzed to make inferences about individuals. Data might be text, audio, video, sensor based or some combination of the above or they collect data directly from some source or data derived by some process of analysis. They might save the data for a long period of time or they might analyze and reject as the data get channelized.[1]

For example, oil and gas companies collect data from refinery sensors and seismic exploration; mobile communications companies collect data from cell towers; electric power utilities collect data from power plants and distribution systems. Businesses collect large amount of user-generated data from prospects and customers with the help credit card numbers, social security numbers. They also collect data of customers from their buying habits and patterns of usage. [8]

The entry of big data and the need to circulate this information within an organization has created an enormous new target for hackers and other criminals. This data which was previously unusable by organizations is now highly valuable, is subject to privacy laws and compliance regulations and must be protected. [8]

The security threat outlook is growing in numerous ways including progress in the rapid bulk of threats. The graph of increasing threats can be seen from the fact that where in the 1990s, the user of personal computers, on an average, used to receive hardly one or two spam posts per day, which has now changed to approximately 200 billion spam messages sent per day[2]. According to a review, it was found that the IT industry had encountered a malware in 2008 which had occurred after 15 years. [4]

According to a survey by Trend Micro, it is estimated that the threat overview for users has view an increase of six to seven orders of magnitude in the last few years.[2]

According to various researchers, the illusion of financial gain has motivated delinquents to implement new and inventive methods and to become more careful and accurate with each passing year. Nowadays, criminals are more experienced, intelligent, well-versed with technology and evolving their expertise and tools in present time. For example, criminals of today uses many quality controls procedures and methods for creating malware, spams etc. These intelligent persons test these procedures on several machines with different operating systems to ensure it bypasses detection. Meanwhile, server-side distinct threats drive speedy progress and circulation and are not detectible using old-style methods. Everyday hundreds of malware are being multiplied and increases in thousands. And presently, malware is no longer constrained to personal computers but entering into mobile devices also. [2][6]

The circulation points for spam, viruses, malware, and other malicious tools are continuously accumulating, while geospatial threats have become very common. An IP address threat affected many computer users of Italy recently. The result showed its effect not only the users of Italy but whosoever who got access to that IP address. Not only this, there are many threats which occur but are left obscure. This entails software security company recognition to become more thoroughly perceived on geographically distributed areas. Similarly, now the individual in place of community, country, cities, companies, or demographic groups have become the target of Spear phishing threats making their detection further complicating.[5]

The need to accomplish, preserve and develop this enormous bulk and variability of data on a consistent base presents software security providers with an unusual velocity challenge. The variability of the internet over time enhances to the complexity of the problem. Cybercriminals consistently alter authentic and valid sites into fraudulent sites without any fear and hesitation. In one of the examples among many such

transformations, in early 2012, I Frame redirection was installed on a popular news site by the cybercriminals in the Netherlands. [7]Rest of the section is as follows- Section II encompasses problems of big data, whereas Section III presents prospects of big data and in Section IV conclusion is presented.

Issues of Big Data

It is assumed that as technology progresses over time, the magnitude of large volume of information that qualify as big data will also grow, resulting in rapid increase in the rate of threat.[1] The threat scenario has evolved simultaneously, with the number of threats increasing by instructions of magnitude in short phases. Due to this evolving threat, the number of refined and polished methods and computing power that criminals can now have at their disposal, and with the creation of big data, the software security companies are struggling with challenges on an exceptional scale. [2] Protecting computer users from the assault and attack of threats is not an easy task, but if threat revealing and prevention tools, methods techniques process etc are weak, the result will be insufficient and inaccurate[7][5]. To add security know-hows into a big data environment, certain strategies, policies methods etc need to balance with the data.

Data came into existence, collected and possibly processed immediately, communicated or stored, (locally, remotely, or both), copied, or analyzed, communicated to users or archived or discarded. Technology at any of these stages can affect privacy positively or negatively.

- **Privacy**

  Privacy is different from security in many aspects. First and the foremost thing is that these policies cannot be in build in the code of any software using big data precisely. Perhaps this is because the beliefs and predilections of human beings have larger diversity than the useful possibility of proclamations about computer security. Indeed how to codify human privacy preferences is an important and emerging area of research. [17] Security generally deals with preparing today's task force to meet the threats and challenges which will be faced by the cyber world tomorrow. But privacy policies are used for building the future policies for facing future threats using future platforms. These platforms encompass not just hardware and software, but also new and different kinds of data and algorithms. [8][9]

- **Complexity**

  Complexity is the most basic and intrinsic characteristic found in the approval, employment and in big data technologies. About half of the users of big data find analyzing and deploying big data as the most complex task needed to enhance their enterprise's cyber defense. The users of big data do not know how a big the solution will be and its effects on their information technology environment and who has the right expertise to manage the new technologies. To successfully implement big data solution, new and innovative technology is a basic requirement that will store, organize, and further analyze massive and diverse data sets. Inter-operability among existing data environments and new technologies is liable upon selecting the accurate technologies and having the right proficiency to implement them. [7][8]

- **Cost**
  The growing stealth and sophistication of attacks can place a stress on even the utmost substantial security budgets, which are already becoming weak in handling risk like insecure mobile devices and apps, data breaches; social engineering tactics; insider negligence; and use of insecure cloud services and many more. [8][21]A survey done by many researchers coined that many enterprises whether small or big quote inadequate budgets due to lack of adoption of big data analytic tools and methods. Generally, cost along with storage, computers, data tools and data visualization frameworks, etc. plays the biggest factor for these enterprises in deploying a big data solution. But with the emergence of big data solutions offered by many software security companies and cloud services, it may be perceived that costs to include and manage big data solutions will fall and adoption rates will increase with a rapid pace. Besides, for one time investment on these big data solution, there are some hidden costs also which are known as opportunity costs. Purchasing new technologies affect an enterprise's ability to spend and sustain other technologies, essential to their security such as firewalls and detection software etc.[9][17]

- **Data Policies and Data Security**
  When data is allowed to travel through a medium or network, it needs to follow certain protocols or rules or some set of policy issues of transmission across organizational boundaries but these rules are not limited to privacy policy, security rules, protection against intellectual property and risk liability.[9][11]Evidently, privacy is an issue whose impact particularly to users of big data is developing as the value of big data becomes more specious. Besides data policies, security of data is also a very important issue. For example, tools are used to manage sensitive and confidential data. Some of the recent surveys have proved that there are not only personal or organizational data breaches but also nation-based data and security breaches. With severe breaches on the growth, talking about security issues of data through technical and policy process will become obligatory.[12][13]

- **Technology and Techniques.**
  New technologies for extracting, storage and maintenance of big data are needed by organization to capture value from big data. New problems and increasing computing power will affect the development of new analytical techniques. [13][14]

- **Organizational Change and Talent**.
  There is an absence of understanding of the value of big data among many organizational leaders as in how to unlock this value. In this present world of competition, whether established companies or new arrivals, all are likely to weight the value of big data. But it is a well-known fact that many organizations do not have the talent in form of experienced man power to derive perceptions from big data. In addition, many organizations today do not formulize work flows and incentives in ways to augment the efficient use of big data for making better decisions and apply more informed action.[14][15]

Challenges of Big Data

1. New methods of tracing threats are needed to process the voluminous data, upcoming from the world and to stay ahead of a sophisticated, aggressive, and ever-growing threat scenario. The traditional

rules of safeguarding the data no longer apply. Scaling up to control the changes in the threat perception is required, but it must be done logically. A brute force approach is not rationally viable. Successful protection and security of theses vast data is highly dependent on the accurate grouping of methodologies, human insight, an expert understanding of the threat and the efficient dealing of big data to create actionable intellect.[4][9][21]

2. It is the misconception that intelligence always comes with experiences. Due to this, many enterprises whether large or small lack specially trained analysts to design these big data systems and use the results of the analysis.[10][14]
3. In the search for new and innovative ways to store and exploit big data, companies need to ensure that they have mechanisms in place which allow them to meet government compliancy regulations for data protection, especially for data at rest.[4][10]

4. Envisaging the next threat can prevent an attack that could possibly cause monetary damages of millions of dollars. Precise and exact prediction of previous history is very important as it helps the organization in decision making. The key for success to many renowned software companies is the past analysis of company's behaviors and on the basis of this predict future behavior. This requires employing of effective mechanisms to archive historical information, access and provide prompt reporting and details.

5. The user's psychology also plays a very importing role in identifying the threats and attacks. Every time the user follows a fixed set up or arrangement which may consist of visiting a news site, comes across several ad servers and cataloging on to any social networking site.[17][18]. If that pattern suddenly changes, diverting the user to a domain which he had never visited earlier, can immediately be prioritized for further analysis. There may be many complex co-relations which can be identified only by some specific type of system that are able to process a very large number of database searches happening per second.[15][21]

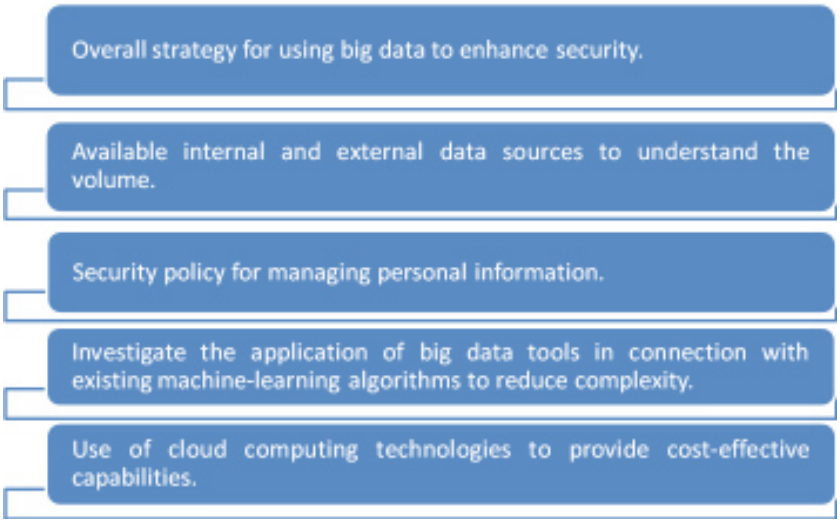
6. While security of big data is a numbers game, but then human intervention is prima fascia. Skilled analysts need to continuously develop the combination of methodologies, apply the human intuition to complex problems and identify trends that computers miss.[14]

7. Various law enforcing companies are working directly with the ISP involved in an attack to drive a better end result.[19]

8. Ultimately, the first and the foremost thing is creating security awareness among the users and the well implications of the security and data policies for safety and security of big data.[9][14][21]

9. Security should be incorporated from the very beginning in all the application which is handling big data like Hadoop, Cloud, etc. So this big data application is facing many challenges but proper implication of these measures can reduce the security risk to some extent.[13][14]

A pictorial representation of issues of Big Data is summarized below:



Brief layout of issues of Big Data

Conclusion

Successful protection relies on the right blend of policies, methods, and human insight, an expert understanding of the threat overview and the proficient treating of big data to create actionable intelligence. Understanding the organization of big data, analyzing and studying the complex relationships, using specialized algorithms for searching relevant data and engaging routine models are some of the critical components. Besides, the companies which are handling big data and the government plays a very important and prominent role in preventing such threats by making new and effective security policies and updating the old data policies. Technical enhancements of privacy can be effective only when accompanied by regulations or laws because unless some penalties are enforced, there is no end to the increase of the measures counter measures "game" between violators and protectors. Rules and regulations provide both deterrence of harmful actions and incentives to deploy privacy protecting software technologies. From everything already said, it should be obvious that new sources of big data are abundant; and that they will continue to grow; and that they can bring enormous economic and social benefits.

Works Cited

[1] James Manyika, Michael ChuiMay 2011Big data: The Next Frontier for Innovation, Competition, and Productivity

[2] Banerjee, C., &Pandey, S. K. (2010). Research on Software Security Awareness: Problems and Prospects. ACM SIGSOFT Software Engineering Notes, 35(5), 1-5.

[3] Deb ShinderBig Data: The Security Perspective (Part 1), [Published on 7 Aug. 2013

[4] By Mike Ferguson Enterprise Information Protection - The Impact of Big Data

[5] Report To The President Big Data And Privacy: A Technological Perspective

[6] Josh Halliday (10 January 2011). "Email Spam Level Bounces Back After Record Low". guardian.co.uk.http://www.guardian.co.uk/technology/2011/jan/10/email-spam-record-activity.

[7] Mark Bouchard ,Big Data for Advanced Threat Protection Key Criteria for Cutting Through the Clamor

[8] Zettaset, The Big Data Security Gap: Protecting the Hadoop Cluster

[9] Banerjee Arpita, Banerjee C. (2014). Cyber Security Awareness Through Education: Problems and Prospects. IMPETUS an Interdisciplinary Research Journal.2(1)

[10] Addressing Big Data Security Challenges: The Right Tools for Smart Protection

[11] Ponemon Institute LLCSponsored byMicrosoft CorporationNovember 2014 Enhancing Cyber security with Big Data: Challenges & Opportunities

[12] "A Special Report on Managing Information: Data, data Everywhere," The Economist, February 25, 2010; and special issue on "Dealing with data," Science, February 11, 2011

[13] James Manyikaet, al.Big data: The Next Frontier for Innovation, Competition, and Productivity

[14] Banerjee C., Banerjee Arpita, PandeyS. K. (2013): Software Security Awareness: Comparison of Artifacts Based Awareness Tools and Techniques.SGVU Journal of Engineering & Technology, 1(1), 33-38

[15] Challenges and Opportunities with Big Data A community white paper developed by leading researchers across the United States

[16] VenkataNarasimhaInukollu et, al "Security Issues Associated With Big Data Incloud Computing, International Journal of Network Security & Its Applications (IJNSA), Vol.6, No.3, May 2014,Enhancing Big Data Security, www.advantech.com

[17] Conducted by Ponemon Institute LLC November 2014Enhancing Cybersecurity with Big Data: Challenges & Opportunities

[18] Arpita Banerjee, et.al. Software Security Awareness Framework: Management Perspective

[19] C. Banerjee, et. al. IT Security Practices In An Organization: Balancing Technology And Management Perspective

[20] Banerjee, C., &Pandey, S. K. (2009). Software Security Rules, SDLC Perspective. arXiv preprint arXiv:0911.0494.

[21] Banerjee Arpita., Banerjee C, Poonia Ajeet. Security Threats of Social Networking Sites:An Analytical Approach , International Journal of Enhanced Research in Management & Computer Applications, ISSN: 2319-7471Vol. 3 Issue 12, December-2014, pp: (1-4)