International Journal of Engineering Science Invention
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org ||Volume 6 Issue 9|| September 2017 || PP. 20-27

A brief Investigation on Data Security Tools and Techniques for Big Data

Vinod Kumar[1], Pushpendra Kumar[2] RS Thakur[3]
[1, 2, 3]*(Computer Application Department, MANIT, BHOPAL, Madhya Pradesh, India)*
Corresponding Author: Vinod Kumar

Abstract: *It is an era of information and in this era of information, due to continuous development in the field of electronics and IT; the computational devices and storage becoming inexpensive. With these growing computational capabilities, data is generated from everywhere. These data are stored in databases for future references/decisive purposes. The term Big Data is used for these massive data having varieties, generated with velocity and measured in term of Tera, Peta, Exa, Zetta, Yotta Bytes. Enormous data regularly talks about these mixed bags of data sets: conventional endeavor created information, machine-produced or sensor data, and online networking data. Diversity of data sources, streaming nature of data acquisition, distinct data formats, large scale heterogeneous networking environment, proprietary technologies/software are the big causes that is why there is a need to think beyond traditional security and privacy solutions. The objective of this research contribution is to present an analytical and comparative study on existing data security tools and techniques for big data.*

Keywords: *Big Data, Data Security Tools, Disk Encryption, Techniques, File Encryption*

Date of Submission: 21-08-2017 Date of acceptance: 05-09-2017

I. Introduction

The giant amount of information generated from all over the world are all syndicated to trigger an emission of data that is being delivered with unlikely sum, rate, and differing qualities. Subsequently, associations requests for an effective strategy to safeguard, to make utilize, and accomplish real-time knowledge from "Big Data" [1] [2]. New tools and techniques for dealing with challenges of cyber threats are needed to compute the enormous 13volume of data sets growing from the machine world and to stay ahead of a complicated, violent, and constantly increasing threat, hazards, menace landscape. No ready-to wear methods are available that can directly be applied to solve enormity of this kind of problem. The convention tools as well as the techniques of security provisions can no longer be applied to deal with big data. Developing and scaling up the security and privacy tools to manage the changes in the threat horizon is needed, because of enormous, complex nature of data the work should be done very wisely keeping the tradeoff among hardware software and cost.

With the time, Big Data became the core competitive factor for enterprises to develop and grow. In the age of Big Data [3][4], data is generated from everywhere, some enterprises such as; information industrial enterprises will put more focus on the technology or product innovation for solving the challenges of big data, i.e., capture, storage, analysis and application. Enterprises like, manufacturing, banking and other enterprises will also benefit from analysis and manage big data, and be provided more opportunities for management innovation, strategy innovation or marketing innovation. High performance network capacity provides the backbone for high end computing systems. These high end computing systems plays vital role in Big Data. Persistent and Sophisticated network under attacks have dared and threatened security teams of organizations. Big Data analytics promises major benefits to the enterprises. Business organizations requires to support secure access of data for data analytics, so as to excerpt maximum value from piled information, but it may cause a big probability for security threats. Managing enormous amounts of datasets increases the security threats and level of potential breaches of data. Susceptible data are goldmines for criminals, vulnerable data can be theft and disclosed, it can invade boundary prepared and data security rules and regulations setup, aggregation of data across boundaries can break data dwelling laws. Thus secure solutions for susceptible data sets, however support analytics for meaningful and valuable insights, is indispensable for any Big Data edge. [5] Big data analytics will perform a key role in coming days for identifying security breaches and crimes [6]

One of the most effective and well proven ways of shield against the data breaches is to safe guard data itself by applying the encryption over data. While there is regularly the misconception that information encryption is a pleasant to have, there are expanding weights and difficulties. Enterprise server consists of sensitive data. It is proven to be very helpful in the organizations to encrypt the data of the enterprise server to prevent from those sufferings.

II. Network Security For Big Data At Glance

Enterprises awash in flood of unstructured, semi structured and structured data, which introduced a multitude of security and privacy issues for organizations to contend with. Today's undertaking security groups engaged and scanning for the main drivers of the assault frequently have a craving for searching for a needle in a hayrick. But as per a white paper [7], finding weighty information in context of big data is just like looking for the needles. Security has traditionally been all about the defense. The term network security means providing security when data is on fly, i.e. over network. Network activity checking remains a conclusive part of any undertaking's security methodology; however picking up connection into the massive measures of information gathered from Network, in an auspicious manner, is still an obstacle for some venture security groups. Episode responders are in the long run searching for conceivable approaches to authoritatively recognize dangers for assessing danger of disease and to step to remediate [8]. A new generation of methods and architectures designed specifically for big data technologies are needed that extract value from gigantic amounts of different data types through high-velocity capture, discovery and analysis. In its review, authors [9] illustrate efficient extraction of value from data and through a figure correlate three associated things: analytics, cloud-based distributed environment deployment, and Networked Society, and these will be inextricably linked.

It is observed that data generated by the many devices having spatial and temporal characteristics, are part of the networked society. When network society, cloud computing and different phases associated with big data are correlated and viewed in a single sleeve, these two figures (Fig. 1 and Fig. 2) are originated in current and future context, because networking is currently in a transition phase, from layer-based approaches to layer-less approaches. So from network security point of view focus should be assessed from current scenario to future requirements. Getting oneself abreast of current literature on Big Data and their idiosyncrasies with respect to security and privacy issues of/in Big Data is totally dependent on three Vs (variety, velocity and volume). Since a proliferation of data which is being generated by multitude of devices, users, and generated traffic, with incredible volume, velocity, and variety [17].Authors of a research paper discussed characteristics, architecture and framework for Big Data [18]. As per authors of same research paper, a big data framework consists of several layers, such as system layer, information gathering layer, processing layer, modeling/statistical layer, administration/access/query layer, visualization/presentation layer and so forth. Authors of a paper [19] highlights that traditional security approaches are inadequate since they are tailored to secure small-scale static data.

The three Vs of Big Data requests ultra-quick reaction times from security and protection arrangements/items. In same paper, Author highlights that these are the main 10 Big Data security and protection challenges from Big Data point of view: needs secure computations in distributed programming frameworks, non-relational data stores demands best security practices, required security at data storage and transactions logs, validation/filtering is required at input end-points, real time security/compliance monitoring is required, scalable and compostable privacy preserved data mining and analytics required, access control and secure communication must be cryptographically enforced, demands granular access control, required granular audits, and data provenance.

Due to dependency on Big Data and criticalness of data/information/knowledge in terms of human lives, there is a need to rethink particularly from a security viewpoint. Big Data breaches will be big too, with the potential for even more serious reputational damage and legal repercussions than at present. Security and privacy issues must be magnified by three Vs of Big Data. Diversity of data sources, streaming nature of data acquisition, distinct data formats, large scale heterogeneous networking environment, proprietary technologies/software are the big causes that why there is a need to think beyond traditional security and privacy solutions.

In a white paper [10], author focuses on intelligence-driven security for big data and states that rapid and massive growth information related to security creates new competencies to defend against the unknown threats. Authors of white paper [11] states that the intelligence is necessary for tackling security and privacy issues related to big data. In the same paper authors suggested that these four steps are required for security intelligence: Data collection, Data integration, Data analytics and Intelligent threat and risk detection (which incorporates constant danger recognition/assessment, Pattern matching, factual connection, security examination, and log data administration, i.e. screen and react with the assistance of complex connection innovations).

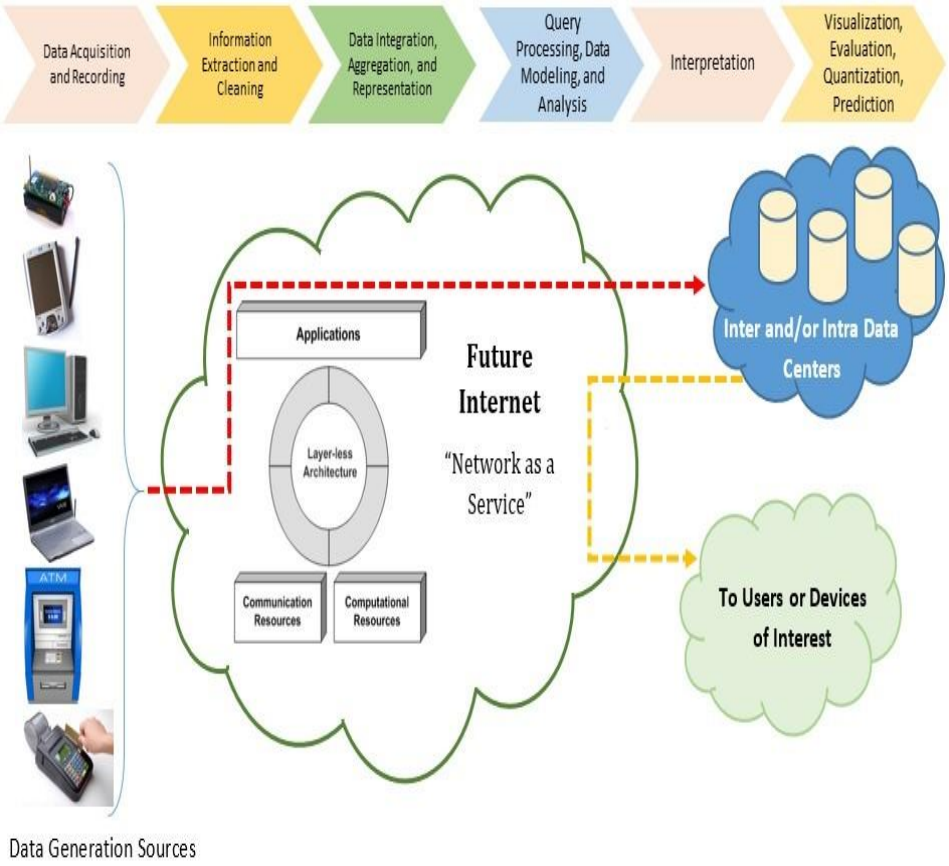

Figure 1. Big Data Aspects in Future Networking Consequence

III. Threats For Network Security In Big Data

The Threats or vindictive computerized events regularly cross different channels and use distinctive attack bearings. Usually data theft incorporate five stages: identify target and a way to enter network, drill/partially crack the network for entrance, discover the valuable information, plant an agent (tiny programs/patches) close to desired data, and finally data/information leak out. So we must have binaries included in an assault or document hashes, log information, order and control foundation, host possession, area, on-screen character meta-data, and so forth. For identification of treats in real-time- The best way for security during data on fly is fastest data transfer between source and destination with multilayer strongly encrypted wrapping of data packets.

According to an exploration firm [12], associations/ventures ought to adjust the abilities security in a comprehensive digital security technique customized to the threats and the risks particular to the association's requests. Associations/undertakings ought to search for clients or different substances, profile records, and search for odd exchanges against those profiles. As per the same research firm [12], big data demands will soon change the currently available security products/concepts. Specialists trust the development of enormous data investigation may give new devices in fighting digital security threats but an integrated prevention approaches is the best option against treats.

Since threat landscape is growing simultaneously with the three Vs of big data and demonstrating same qualities too, if risk location/countermeasures components are feeble, the outcome will be lacking. According to white paper [13], the right blends of components/techniques, a specialist comprehension of the danger scene, human knowledge and intelligent and quick handling of enormous data to make noteworthy insight. In this manner a decent comprehension and knowledge is obliged to associate how information is gathered and where, how to sorted out them, how to break down complex associations with the assistance of particular hunt calculations and utilizing need based custom models are the primary discriminating segments to accomplish security/protection in huge information. Due to the massive availability of available data in/from public domain, data leak-out can be costly and data hackers become more damaging [14].

Figure 2. A Simultaneous Growth in Big Data and Threats

Passwords and controlled access via permissions, Two-factor (or multi-factor) authentication, Firewalls, Data Leakage Prevention (DLP) Technology are basic available existing technological approaches that are already relatively mature with time but inadequate to tackle and fulfill demands of big data.Since confidentiality, authentication and integrity are the basic security primitives, for these and other specific needs for big data one integrated solution from a holistic point of view rather than solving the security and privacy issues on requirement basis. Due to gigantic and variety of data the traditional cryptography based security mechanism are not sufficient, to resolve this. Authors [15] suggest a picture data security in hybrid cloud. Since with the cross breed cloud touchy cloud can be put away in private mists while no sensitive information can be store out in the open mists. By bringing the security related data together at single centralized place, analysis can be performed that wasn't possible previously. This provides a competitive advantage over previous approaches here additional data sets can be correlated in different ways with existing data and new relationships will be found which was previously unimaginable.

Validation and assurance of end-to-end security, application specific security model, message-level security, policy oriented security, and security as a service are some security solutions and can be fruitful when shifted from current context to future context.

IV. Challenges For Network Security In Big Data

The only major reasons for the challenges faced in implementing security in big data are as follows:-

- The data which are collected, aggregated, and analyzed for big data analysis [21]. Big data repositories generally congregate information from a variety of sources. This variety of data makes secure access management- a challenge. The most important challenge is to provide secure access to this variety of data gathered.
- The infrastructure which is used to store and build the big data [21]. As it is very known that big data environment is distributed in nature. This distributed nature makes the make big data very complicated and extreme sensitive to the attack. It is also very complicated to provide the physical security to the distributed data stores.
- The technologies which are applied to analyze structured and unstructured big data [21]. The programming tools which is used in big data such as Hadoop and NoSQL databases was not developed by keeping security in mind. For example, in case of Hadoop there was no way to authenticate the user or any services and also doesn't provide any encryption mechanism of the data on fly.

V. Tools And Techniques Available To Answer Big Data Secuirty Issues

Big data can play a vital role in security management as well, as per the white paper [16], security management foundational concepts involve three aspects:

- A responsive "scale out" base that ready to react and fit for variable framework and advancing security menace.
- Examination and perception instruments to bolster security experts. It incorporates from essential occasion distinguishing proof with supporting points of interest, inclining of key measurements notwithstanding abnormal state perception, reproduction of suspicious documents with apparatuses to robotize testing of these records, and full recreation of all log and system data around a session to focus definitely what happened.

- Threats insight to associate reasons/example/effect of threats noticeable inside association with currently accessible data about threats outside the association.
- Devices to bolster four noteworthy utilization cases in enormous information arrangement: stream processing, interactive processing, batch and graph processing, which are all standard in any extensive huge data arrangement nowadays.
- Streaming is ordinarily required when we as of now have some known examples that we use to distinguish dangers, and we need to have the capacity to recognize those known dangers from a high-volume stream of information progressively.
- Intuitive processing is imperative to help our security analyst's velocity up the manual examination procedure to discover new examples in diverse extensive data sets.
- Cluster preparing alludes to run of the usual MapReduce-based machine learning and demonstrating work that data researcher's utilization to discover new threats examples in the information.

5.1 Traditional Defense against Attacks for Data at Rest

Basically there are three layers to implement security shown in fig [1]. They are as follows-

1) Layer 1

This layer basically provide the identification and authentication functionality. The basic idea is to provide the authentication of the user who is going to access the system. It determines whether the user is authorized user or not. It basically implements these things from UID, password, One-Time-Password (OTP), thumb scanner and all.

2) Layer 2

After the user gets authenticated, user may trigger some commands to do certain operations. This layer generally determines whether the user is having the right to access a particular service or not. This layer also hides the data from various users by encrypting the data using either file encryption or disk encryption or both.

3) Layer 3

This layer is responsible for logging the information which is required by system administrator to implement the security. Generally a record is maintained which consist of all the successful and unsuccessful login attempts, resources utilized by a specific user which will be further helpful for the admin to give access to the appropriate user.

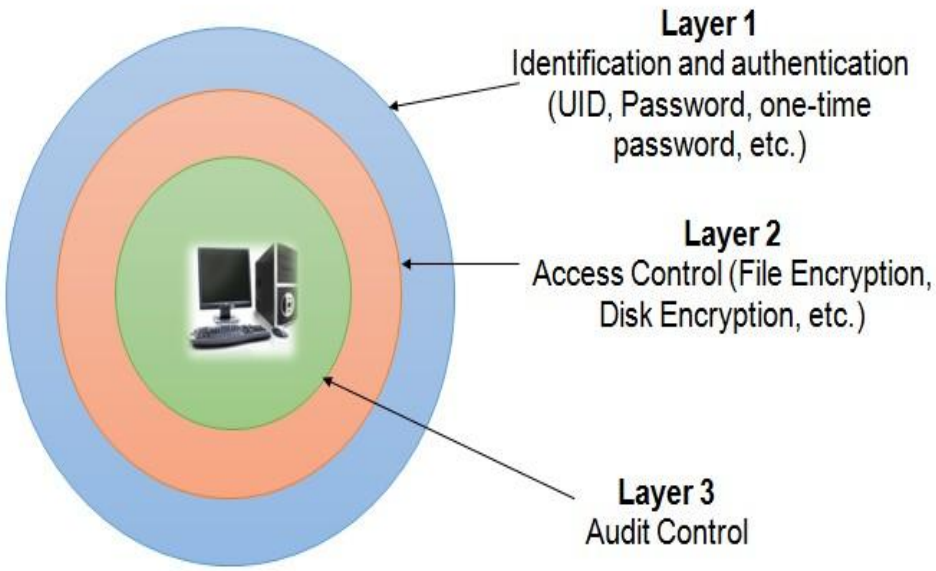

Figure 3. Layered Architecture of implementing security

5.2 File Encryption Tools

Setting the information in a difficult to reach registry and making the record disjointed by others will keep it secure as a rule. Encryption utilizes a calculation that shrouds the content's importance. List of some popular File Encryption tools are given below in table [1].

Table 1. List of Commonly Used File Encryption Tools

Tool Name	Type	Platform	Algorithm
AxCrypt	Open source	Windows	AES-128
Encrypt\Decrypt File Utility	Open source	Windows	AES-128
File Secure	Open source	Windows	AES
EncryptOnClick	Open Source	Windows	AES-256
Sophos	Commercial	Windows 8/MAC/Linux	AES-256
Cryptkeeper	Open Source	Linux	------
EMSAEZ Encryption	Open Source	Windows	Blowfish algorithm
Advanced File Encryption	Open Source	Windows	8192 bit symmetric key encryption

5.3 Disk Encryption Tools

To ensure privacy of the information put away on a computer disks a data security procedure called disk encryption is utilized. Disk encryption programming can straightforwardly work on a whole disk volume, a catalog, or even a solitary record, it is vital to separate it with (non-straightforward) document encryption programming which scrambles or unscrambles just individual documents and dependably the entire document (the decoded document is put away in a makeshift record in a decoded structure). Some prominent disk encryption is given underneath in table [2].

Table 2. List of Commonly Used Disk Encryption Tools

SN	Tool Name	Type	Platform	Algorithm
1.	Bitlocker	Open source	Windows	AES-256
2.	DiskCryptor	Open source	Windows	AES-256, Twofish and Serpent
3.	BitLocker Drive Encryption	Open source	Windows	AES-128/AES-256
4.	Cryptoloop	Open Source	Linux	Crypto API
5.	EncFS	Open Source	Linux	----
6.	CrossCrypt	Open Source	Windows	AES-256

5.4 Commercial Tools and Techniques

- IBM Threat Protection System is a robust and comprehensive set of tools and best practices that are built on a framework that spans hardware, software and services to address intelligence, integration and expertise required for Big Data security and privacy issues.
- HP ArcSight, another product that can strengthen security intelligence, able ready to conveys the propelled relationship, application assurance, and system barriers to shield today's cloud IT base from refined digital dangers.
- Another set of products (Identity-Based Encryption, Format-Preserving Encryption and many more) given by Voltage Security Inc., gives new effective systems to ensure information over its full lifecycle. RSA Security Management Portfolio, for Infrastructure, Analytics, and Intelligence can be another good option.
- Cisco's Threat Research, Analysis, and Communications (TRAC) devices are likewise a decent alternative in this class.

VI. Big Data For Security- Future Directions In Deployement

The sending of Big Data for extortion identification, and set up of security incident and event management (SIEM) frameworks, is appealing to numerous associations. The overheads of dealing with the yield of customary SIEM and logging frameworks are demonstrating a lot for most IT divisions and Big Data is seen as a potential friend in need. There are business swaps accessible for existing log administration frameworks or the innovation can be conveyed to give a solitary information store to security occasion administration and advancement.

Today logs are frequently disregarded unless an episode happens. Huge Data gives the chance to combine and break down logs naturally from numerous sources instead of in disengagement. This could give knowledge that individual logs can't, and possibly improve intrusion detection system (IDS) and intrusion prevention (IPS) through nonstop alteration and viably adapting "great" and "terrible" practices. Coordinating data from physical security frameworks, for example, building access controls and even CCTV, could likewise

fundamentally upgrade IDS and IPS to a point where insider assaults and social designing are calculated into the discovery process. This introduces the likelihood of essentially more propelled location of extortion and criminal exercises.

Big data will have an impact that will change most of the product categories in the field of computer security including solutions, network monitoring, authentication and authorization of users, identity management, fraud detection, and systems of governance, risk and compliance. Big data will change also the nature of the security controls as conventional firewalls, anti-malware and data loss prevention. In coming years, the tools of data analysis will evolve further to enable a number of advanced predictive capabilities and automated controls in real time.

This new time of registering obliges another way to deal with security. Here are three tips-

1. Turn to the cloud and mobile computing to improve security: Use the new options available through cloud and mobile to improve security. For instance, a cloud creates the opportunity to build in security right from the start. Crowd sourced threat intelligence provides the tips needed to stay ahead of cyber-attacks.
2. Use analytics for more brilliant protection and aversion: Conventional security innovations do not have the complex capacities expected to distinguish and ensure against today's information driven assaults. That is the reason associations need to utilize information to battle these assaults. Equip your security team so it can hunt for breaches by collecting security data from everywhere in the enterprise. Roll out security intelligence technologies that use real-time analysis, fraud prevention and anomaly detection. At the same time, organizations need to lock down their “crown jewels,” protecting their most crucial data by constantly monitoring who is accessing the data and from where. Finally, prepare for the inevitable. Enable your team with a “hunter mentality” to think like an attacker. Construct a coordinated plan for responding to an attack using the right tools, information and skills so you can limit the effects of an inevitable breach.
3. Build up a coordinated methodology: Today organizations utilize antivirus programming to get rid of malware and firewalls to keep out the "terrible fellows," At the point when important messages develop, it's frequently past the point of no return -competitive advantages are a distant memory or clients' charge card information has as of now been traded off.
4. Which is why it’s more crucial than ever for organization to craft a systematic approach to security? Create an integrated system that reaches across the different networks and devices you’re responsible for, tie existing security technologies into this foundation, and roll out security analytics on top of it. Then, to keep up with security risks that are evolving at hyper speed, it is required to cultivate a culture of vigilance and tap outside experts. Grade yourself against your peers and constantly test how well your organization is doing compared with industry standards. At the same time, reach out to outside security professionals and researchers. Partnerships build strength in helping shore up skills and pinpoint and deal with new threats. Security decisions should be made with data.

It’s more crucial than ever for your organization to craft a systematic approach to security? Create an integrated system that reaches across the different networks and devices you’re responsible for, tie existing security technologies into this foundation, and roll out security analytics on top of it. Then, to keep up with security risks that are evolving at hyper speed, it is required to cultivate a culture of vigilance and tap outside experts. Grade yourself against your peers and constantly test how well your organization is doing compared with industry standards. At the same time, reach out to outside security professionals and researchers. Partnerships build strength in helping shore up skills and pinpoint and deal with new threats. Security decisions should be made with data.

VII. Conclusion

With the advent of information technology for dealing with the data to store, preprocess, extract useful information, interpret and visualize, in both the cases of data at rest and data on fly in the network. The security risks and issues always has been subject of great concern for everyone. In Big Data, the implementation of security feature demands the completely abstract approach. Techniques like encryption, decryption, authentication, compression etc. causes the system to slow down exponentially. Therefore, still, there is a requirement of new approach which can tradeoff between security and resource performance and also its utilization.

References

[1]. The Economist, Nov 2011,"*Drowning in numbers–Digital data will flood the planet and help us understand it better*",http://www.economist.com/blogs/dailychart/2011/11/big-data-0
[2]. Seref Sagiroglu and DuyguSinanc, “Big Data: A Review”, pp. 42-47, 2013.
[3]. Jean-Pierre Dijcks, “Big Data for the Enterprise”, An Oracle White Paper, Oracle Corporation, June 2013.
[4]. Lohr S., Feb 11, 2012, "The Age of Big Data", New York Times, http://www.nytimes.com/2012/02/12/sunday-review/big-

datasimpact-in-the-world.html
[5]. Voltage Security, "Big Data, Meet Enterprise Security", White paper http://www.voltage.com/solution/enterprise-security-for-big-data/
[6]. ResearchFirm,https://www.gartner.com/doc/2773117?ref=SiteSearch&sthkw=big%20data%20security&fnl=search&srcId=1-3478922254
[7]. "Gettting Real About Security Management And Big Data: A Roadmap for Big Data in Security Analytics", White Paper, RSAs and EMC Corporation,www.EMC.com/rsa.
[8]. Arbor Networks Blog on "Next Generation Incident Response, Security Analytics and the Role of Big Data", http://www.arbornetworks.com/corporate/blog/5126-next-generation-incident-response-security-analytics-and-the-role-of-big-data-webinar, Feb. 2014
[9]. Mona Matti and Tor Kvernvik, "Applying Big-data technologies to Network Architecture", Ericsson Review, 2012
[10]. S. Curry, E. Kirda, Sy, L. Stemberg, E. Schwartz, W. H. Stewart, and A. Yoran, "Big Data Fuels Intelligence-driven Security", RSA Security Brief, Jan. 2013
[11]. "Big security for big data", Business white paper, HP, December 2012.
[12]. Gartner- Research Firm, http://www.gartner.com
[13]. "Addressing Big Data Security Challenges: The Right Tools for Smart Protection", White Paper, Trend Micro Incorporated, September 2012. www.trendmicro.com.
[14]. Schmitt, C., Shoffner, M., Owen P., Wang, X., Lamm, B., Mostafa, J., Barker, M., Krishnamurthy, A., Wilhelmsen, K., Ahalt, S., &Fecho, K. "Security and Privacy in the Era of Big Data", White Paper, ARENCI/National Consortium for Data Science, Vol. 1, No. 2 in the RENCI White Paper Series, November 2013.
[15]. Xueli Huang and Xiaojiang Du, "Achieving Big Data Privacy via Hybrid Cloud", IEEE INFOCOM Workshop on Security and Privacy in Big Data, pp. 512-517, 2014.]
[16]. "Gettting Real About Security Management And Big Data: A Roadmap for Big Data in Security Analytics", White Paper, RSAs and EMC Corporation, www.EMC.com/rsa.
[17]. "Big security for big data", Business white paper, HP, December 2012.
[18]. FiratTekiner, and John A. Keane, "Big Data Framework", In proceedings of IEEE International Conference on Systems, Man, and Cybernetics, 2013.
[19]. A.Cardenas, Y. Chen, A. Fuchs, A. Lane, R. Lu, P. Manadhata, J. Molina, P. Murthy, A. Roy, and S. Sathyadevan, "Top Ten Big Data security and Privacy Challenges", White Paper, Cloud Security Alliance, November 2012. http://www.cloudsecurityalliance.org/
[20]. J. Chauhan, "Penetration Testing, Web Application Security", Lecture Notes, Available at: http://www.ivizsecurity.com/ blog /penetration- testing/top-5 -big-data-vulnerability-classes/
[21]. Jeff markey "An article on how to manage big data security" at http://data-informed.com/manage-big-datas-big-security-challenges/

International Journal of Engineering Science Invention (IJESI) is UGC approved Journal with Sl. No. 3822, Journal no. 43302.

Vinod Kumar. "A brief Investigation on Data Security Tools and Techniques for Big Data." International Journal of Engineering Science Invention (IJESI), vol. 6, no. 9, 2017, pp. 20–27.