Proposing a model on Security challenges in cloud computing especially Social Media and Social Sites

Fayaz Ahmad Lone[#1], Dr. Amit Kumar Chaturvedi[*2]

#Ph.D. Scholar, Mewar University, Chittorgarh
Rajasthan, India

Abstract *- For big data application security challenges, Social media is the next battleground. Big social data can serve as an early warning system of trouble brewing and a leading indicator of imminent action by a potential troublemaker. Mostly Social media listens for early detection of adverse events and product quality surveillance. It may also be used for product recommendation. In this paper, we will focus to uncover ideas and open problems with big data security in cloud computing especially Social Media and Social Sites. We have also proposed a model Information Measuring for Social Media [IMSM] Model for improving the contents and quality of information on social sites.*

Keywords – Social sites, cloud, big data, security, information quality.

I. INTRODUCTION

This is the era of communication and presently we have so many options of communication, voice telephony, video calling, email, online chatting etc. Several people maintain relationships online through social networking sites like Facebook, LinkedIn, Twitter and others. The rapid growth of these social networks has given rise to marketing and customer relationship opportunities for businesses, and large datasets for analytics. So, lot of text data, pictures, videos, advertisements etc are uploaded on the websites. For the future of social networking there will be a continued focus on user privacy and data control. Since Cloud databases are used for storing contents and continuously photos, videos, messages, images etc uploaded on these cloud databases. There is no check or control on the contents uploaded on internet, especially when we discuss in concern with the social media and social sites.

In this paper, we will focus to uncover ideas and open problems with big data security in cloud computing especially Social Media and Social Sites. Basically, Big data is a set of very large and complex data, which is difficult to capture, process, and analyse using current computing infrastructure. It is characterized by volume, velocity, variety, variability, veracity, and value. In present time, every internet user have an account on at least one social media to connect with its known persons, society, or groups. Estimated 300 million ne photos / day unloaded on facebook; 300 million Instagram users share 60 million photos/ day, and more than 100 hrs / min video unloaded to YouTube.

Figure 1 : Social Sites or Media using Big Data

For big data application security challenges, Social media is the next battleground. Big social data can serve as an early warning system of trouble brewing and a leading indicator of imminent action by a potential troublemaker. Mostly Social media listens for early detection of adverse events and product quality surveillance. It may also be used for product recommendation. Recently big data growth is noticed to be increased from GB networks, gigapixel cameras, data-intensive Internet of Things (IoT), etc. The big data usage is increased also and more than 35% of all data could be considered useful. With all the advantages of big data, it also have some challenges like storage, transfer of data, how to clean, search, visualise, analyse, and privacy i.e. how to secure big data.

Cloud computing provides a scalable computing infrastructure, platform and software services through which large datasets can be stored and analysed. Big data transfer challenge requires new protocol and algorithms.

Current innovations aimed at tackling massive flow challenge include: [1] GridFTP: high performance, secure, reliable protocol for bulk data transfer optimized for high-bandwidth wide-area networks

[Foster3]. [2] Globus: a hosted provider of high-performance, reliable, and secure data transfer, synchronisation, and sharing [Foster2].

The cloud allow users and organizations to rely on external providers. This leads to new security and privacy problems i.e. users lose control over their data.

II. BIG DATA ANALYTICS TOOLS

There are varieties of applications and tools developed by various organizations to process and analyze Big Data. The Big Data analysis applications support parallelism with the help of computing clusters. The following are major applications in the area of Big Data analytics.

A. **MapReduce**

MapReduce is a programming model for computations on massive amounts of data and an execution framework for large scale data processing on clusters of commodity servers. It was originally developed by Google and built on well-known principles in parallel and distributed processing [11].

MapReduce program consists of two functions – Map function and Reduce function. MapReduce computation executes as follows:

1. Each Map function is converted to key-value pairs based on input data. The input to map function is tuple or document. The way key-value pairs are produced from the input data is determined by the code written by the user for the Mapfunction.
2. The key-value pairs from each Map task are collected by a master controller and sorted by key. The keys are divided among all the Reduce tasks, so all key-value pairs with the same key wind up at the same Reduce task.
3. The Reduce tasks work on one key at a time, and combine all the values associated with that key in some way. The manner of combination of values is determined by the code written by the user for the Reduce function.

The MapReduce model hide details related to the data storage, distribution, replication, load balancing and so on. Furthermore, it is so simple that programmers only specify two functions, which are map function and reduce function, for performing the processing of the Big Data. MapReduce has received a lot of attentions in many fields, including data mining, information retrieval, image retrieval, machine learning, and pattern recognition.

B. **Hadoop**

Hadoop is a free, Java-based programming framework that supports the processing of large data sets in a distributed computing environment. It is part of the Apache project sponsored by the Apache Software Foundation. Hadoop was inspired by Google's MapReduce Programming paradigm [10].

Hadoop is a highly scalable compute and storage platform. But on the other hand, Hadoop is also time consuming and storage-consuming. The storage requirement of Hadoop is extraordinarily high because it can generate a large amount of intermediate data. To reduce the requirement on the storage capacity,Hadoop often compresses data before storing it.

Hadoop takes a primary approach to a single big workload, mapping it into smaller workloads. These smaller workloads are then merged to obtain the end result. Hadoop handles this workload by assigning a large cluster of inexpensive nodes built with commodity hardware. Hadoop also has a distributed, cluster file system that scales to store massive amounts of data, which is typically required in these workloads. Hadoop has a variety of node types within each Hadoop cluster; these include DataNodes, NameNodes, and EdgeNodes.

C.. **IBM InfoSphere BigInsights**

It is an Apache Hadoop based solution to manage and analyze massive volumes of the structured and unstructured data. It is built on an open source Apache Hadoop with IBM big Sheet and has a variety of performance, reliability, security and administrative features.

D. Hive and Pig Data Model

In the case of Pig all data objects exist and are operated on in the script. Once the script is complete all data objects are deleted unless you stored them. In the case of Hive we are operating on the Apache Hadoop data store. Any query you make, table that you create, data that you copy persists from query to query. You can think of Hive as providing a data workbench where you can examine, modify and manipulate the data in Apache Hadoop. So when we perform our data processing task we will execute it one query or line at a time. Once a line successfully executes you can look at the data objects to verify if the last operation did what you expected. All your data is live, compared to Pig, where data objects only exist inside the script unless they are copied out to storage. This kind of flexibility is Hive's strength.

The outcome of the Big Data research is that Hadoop and MapReduce is used for obtaining solutions for optimizing job scheduling and organization of indexes and layout rendered. The combinations of Hive, Hadoop and Mahout may be used to build a random forest based decision tree model to detect botnet in peer to peer network. Rapidminder and Hadoop proposed an architecture

called Radoop to scale the data and network size. Hadoop and IBM smart analytic system introduced new architecture to support the analytical system. Hadoop proposed a self-tuning system called starfish.

III. RELATED WORK

M. K. Kakhani, S. Kakhani, and S. R. Biradar, have discussed research issues in Big Data Analytics. Recently, Big Data has attracted a lot of attention from academia, industry as well as government. It is a very challenging research area. Big Data is term defining collection of large and complex data sets that are difficult to process using conventional data processing tools. Every day, we create trillions of data all over the world. These data is coming from social networking sites, scientific experiments, mobile conversations, sensor networks and various other sources. We require new tools and techniques to organize, manage, store, process and analyze Big Data. This paper systematically presents various research issues related to Big Data analytics.[1]

S.J.Samuel, Koundinya RVP, K. Sashidhar, and C.R. Bharathi, presented a survey on big data and its research challenges. There has been an ever-increasing interest in big data due to its rapid growth and since it covers diverse areas of applications. Hence, there seems to be a need for an analytical review of recent developments in the big data technology. This paper aims to provide a comprehensive review of the big data state of the art, conceptual explorations, major benefits, and research challenging aspects. In addition to that, several future directions for big data research are highlighted [2].

C.L. Hepsiba, J.G.R. Sathiaseelan discussed the Security Issues in Service Models of Cloud Computing. They explained that — Cloud computing is an emerging technology for providing computing resources and storage to all kinds of users. This technology is facing lot of challenges including data and network security, interoperability, legal and compliance issues. In security issues, there exist numerous risks for the data processed or stored in the cloud environment. Cloud data are may be used by unauthorized access or users. This paper is mainly focused on security issues for cloud service models like and their solutions [3] .

Gayathri. K.S, T. Thomas, J. Jayasudha, presented security issues of Media Sharing in Social Cloud. A social network, as the name suggests is a network of individuals created for social communication. It is a massive platform where people interact with others anywhere in the world online based on some relationship. The relationship can be friendship, family membership or others with some shared interests. One can share anything like profiles, photos, videos etc, with friends. To make communication more effective, content sharing has been widely adapted. With the rapid growth of social networking sites, there has been an increased use of content sharing through social networks. With more and more contents getting shared, the security and rights of shared content are getting compromised. Several existing media sharing sites like YouTube, Flickr, Facebook etc, face several issues with content sharing such as upload limitations, unsupported tracking model and copyright violations. Cloud computing allows the distribution of services such as storage and processing. A social network can be blended with a cloud network. Such a social cloud provides the functionalities of both cloud computing and social networking. This paper aims to examine the security issues associated with media sharing on cloud based social networks [4].

A.Singh, K. Chatterjee, presented a survey on Cloud security issues and challenges. The cloud computing provides on demand services over the Internet with the help of a large amount of virtual storage. The main features of cloud computing is that the user does not have any setup of expensive computing infrastructure and the cost of its services is less. In the recent years, cloud computing integrates with the industry and many other areas, which has been encouraging the researcher to research on new related technologies. Due to the availability of its services & scalability for computing processes individual users and organizations transfer their application, data and services to the cloud storage server. Regardless of its advantages, the transformation of local computing to remote computing has brought many security issues and challenges for both consumer and provider. Many cloud services are provided by the trusted third party which arises new security threats. The cloud provider provides its services through the Internet and uses many web technologies that arise new security issues. This paper discussed about the basic features of the cloud computing, security issues, threats and their solutions. Additionally, the paper describes several key topics related to the cloud, namely cloud architecture framework, service and deployment model, cloud technologies, cloud security concepts, threats, and attacks. The paper also discusses a lot of open research issues related to the cloud security [5].

Sanjay P. Ahuja, Bryan Moore, presented a Survey of Cloud Computing and Social Networks. In recent years there has been rapid growth in cloud computing and social networking technologies. Cloud computing shifts the computing resources to a third party, eliminating the need to purchase, configure and maintain those resources. With the incentive of lowered operational costs in software,

hardware and human effort, many companies are considering the use of cloud services. Likewise, social networks have seen massive growth, with millions of Internet users actively participating across various social networking websites. Even corporations have begun using social networks as a means to market and reach their customers. This paper will survey the current issues in cloud computing and social networks and how these technologies are being used together [6].

Christos Stergioua, Kostas E. Psannis, Byung-Gyu Kimb, Brij Gupta, discussed Secure integration of IoT and Cloud Computing in December 2016. Mobile Cloud Computing is a new technology which refers to an infrastructure where both data storage and data processing operate outside of the mobile device. Another recent technology is Internet of Things. Internet of Things is a new technology which is growing rapidly in the field of telecommunications. More specifically, IoT related with wireless telecommunications. The main goal of the interaction and cooperation between things and objects which sent through the wireless networks is to fulfill the objective set to them as a combined entity. In addition, there is a rapid development of both technologies, Cloud Computing and Internet of Things, regard the field of wireless communications. In this paper, we present a survey of IoT and Cloud Computing with a focus on the security issues of both technologies. Specifically, we combine the two aforementioned technologies (i.e Cloud Computing and IoT) in order to examine the common features, and in order to discover the benefits of their integration. Concluding, we present the contribution of Cloud Computing to the IoT technology. Thus, it shows how the Cloud Computing technology improves the function of the IoT. Finally, we survey the security challenges of the integration of IoT and Cloud Computing[7].

In the 4th International Conference on Recent Trends in Computer Science & Engineering, Gunasekaran Manogarana, , Chandu Thotab, M. Vijay Kumarc presented MetaCloudDataStorage Architecture for Big Data Security in Cloud Computing. The cloud is increasingly being used to store and process the big data. Many researchers have been trying to protect big data in cloud computing environment. Traditional security mechanisms using encryption are neither efficient nor suited to the task of protecting big data in the Cloud. In this paper, we first discuss about challenges and potential solutions for protecting big data in cloud computing. Second, we propose MetaCloudDataStorage Architecture for protecting Big Data in Cloud Computing Environment. This framework ensures efficient processing of big data in cloud computing environment and gains more business insights [8].

Md. Tanzim Khorshed, A.B.M. Shawkat Ali, Saleh A. Wasimi, presented a survey on gaps, threat remediation challenges and some thoughts for proactive attack detection in cloud computing. The long-term potential benefits through reduction of cost of services and improvement of business outcomes make Cloud Computing an attractive proposition these days. To make it more marketable in the wider IT user community one needs to address a variety of information security risks. In this paper, we present an extensive review on cloud computing with the main focus on gaps and security concerns. We identify the top security threats and their existing solutions. We also investigate the challenges/obstacles in implementing threat remediation. To address these issues, we propose a proactive threat detection model by adopting three main goals: (i) detect an attack when it happens, (ii) alert related parties (system admin, data owner) about the attack type and take combating action, and (iii) generate information on the type of attack by analyzing the pattern (even if the cloud provider attempts subreption). To emphasize the importance of monitoring cyber attacks we provide a brief overview of existing literature on cloud computing security. Then we generate some real cyber attacks that can be detected from performance data in a hypervisor and its guest operating systems. We employ modern machine learning techniques as the core of our model and accumulate a large database by considering the top threats. A variety of model performance measurement tools are applied to verify the model attack prediction capability. We observed that the Support Vector Machine technique from statistical machine learning theory is able to identify the top attacks with an accuracy of 97.13%. We have detected the activities using performance data (CPU, disk, network and memory performance) from the hypervisor and its guest operating systems, which can be generated by any cloud customer using built-in or third party software. Thus, one does not have to depend on cloud providers' security logs and data. We believe our line of thoughts comprising a series of experiments will give researchers, cloud providers and their customers a useful guide to proactively protect themselves from known or even unknown security issues that follow the same patterns [9].

S. .J. Samuel, Koundinya RVP, K. Sashidhar, and C.R. Bharathi, "A SURVEY ON BIG DATA AND ITS RESEARCH CHALLENGES". There has been an ever-increasing interest in big data due to its rapid growth and since it covers diverse areas of applications. Hence, there seems to be a need for an analytical review of recent developments in the big data technology. This paper aims to provide a comprehensive review of the big data state of the art, conceptual explorations, major benefits, and research challenging aspects. In addition to that,

several future directions for big data research are highlighted [11].

IV. PROPOSED IMSM [INFORMATION MEASURING FOR SOCIAL MEDIA] MODEL

We are in the age of information, on every topic, issue, or thing in world we have information on internet. But all the information uploaded on the internet is not useful. Some information may misled us and create a negative effect. Social sites are very popular these days to connect anyone with its known people or groups. Marketing companies and institutes, who are engaged in spreading negative information to society may also use this communication network for their use.

So, information to be uploaded over the internet and people who are engaged in sharing or uploading information on the social sites should be measured and then only allowed to share the information among others. If the feedback or response from the user's is not in favour, such information should be dropped from social sites.

Here we have proposed a model ***Information Measuring for Social Media [IMSM]*** in figure 2. According to this model, Intention to use social networking sites or social media for sharing and circulating the information should be measured on the basis of (1) Usefulness of information, (2) Ease of information Access, and (3) Subjective norms of the social network or site or media.

Usefulness of information will be measured on the basis of ***Information quality*** and ***People Attitude***. ***Information quality*** will be marked on four factors credibility, availability, relevance of information, and perception of external control over the information. ***People Attitude*** will be evaluated on the trustworthiness, profession already filed with site, personal experience notified by the site, user rating by people connected with through their responses recorded with the site, and attitude or behaviour towards information being shared.

Ease of information access is an important issue, this leads to the presentation of information. People over the social media always like to attractive presentation of information. So, technology is the very important factor for measuring easy of information access. This factor should be measured on the usability factor of that technology, and Accessibility.

Third and very important factor is the subjective norms of the social sites or media. This is important to understand why that site is created, what are objectives and motivation behind creation of that group. This should be evaluated on the basis of Privacy policy, security constructs, self-identity, and motivational level to its intended users.

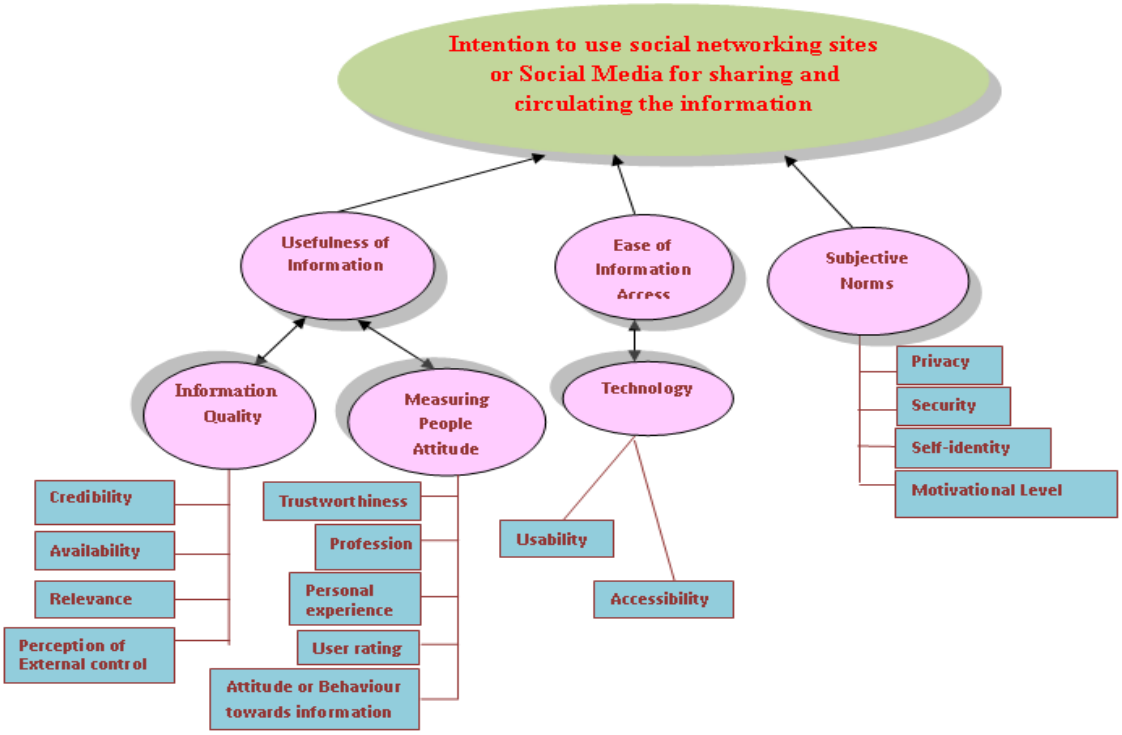

Figure 2: Information Measuring for Social Media [IMSM] Model

V. CONCLUSION

Cloud databases are the maintained on third party servers. Generally, social sites doesn't have their own servers, they hire these servers from cloud service providers on pay per use basis. Social media or sites have also given rise to several applications of data analytics. This trend is

expected to continue with the idea that organizations can derive useful information such as trends and user profiling. Social media has a strong tie with big data as these services produce and consume big data.

After analyzing the literature and various research outcomes, it is observed that there should be a check or control over the people who are uploading the data, videos, pictures etc. on the factors like trustworthiness, profession, personal experience, user rating, and attitude or behaviour. The usefulness of the information should also be checked. All the social sites should also frame norms for privacy, security, self-identity, and motivational level to measure the contents uploaded on the site. Social sites should ensure the ease of information access. Future researcher may work for proposing a model for testing the information quality or People's attitude.

ACKNOWLEDGEMENT

We are thankful to all the people who directly or indirectly helped us to prepare this research paper.

REFERENCES

[1]. M. K. Kakhani, S. Kakhani, and S. R. Biradar, "Research Issues in Big Data Analytics", International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 2, Issue 8, August 2013 ISSN 2319 – 4847.

[2]. Jimmy Lin, Chris Dyer," Data-Intensive Text Processing with MapReduce", Manuscript prepared April 11, 2010.

[3]. C.L. Hepsiba, J.G.R. Sathiaseelan , "Security Issues in Service Models of Cloud Computing", International Journal of Computer Sc. and Mobile Computing, Vol. 5 Issue. 3, Mar 2016, pp. 610-615, ISSN 2320-088X.

[4]. Gayathri. K.S, T. Thomas, J. Jayasudha,, "security issues of Media Sharing in Social Cloud", International Conference on Modelling, Optimisation and Computing (ICMOC-2012), Elsevier, Procedia Engineering 38 (2012), pp. 3806-3815.

[5]. A.Singh, K. Chatterjee, "Cloud security issues and challenges: A survey", Journal of Network and Computer Applications 79 (2017), Elsevier, pp 88-115

[6]. Sanjay P. Ahuja, Bryan Moore, "A Survey of Cloud Computing and Social Networks", Network and Communication Technologies; Vol. 2, No. 2; 2013, ISSN 1927-064X E-ISSN 1927-0658, pp 11-16.

[7]. Christos Stergioua, Kostas E. Psannis, Byung-Gyu Kimb, Brij Gupta, "Secure integration of IoT and Cloud Computing", Elsevier, Future Generation Computer Systems, pp. 1-12, Dec 2016.

[8]. G. Manogarana, , C. Thota, M. Vijay Kumar, "MetaCloudDataStorage Architecture for Big Data Security in Cloud Computing", 4th International Conference on Recent Trends in Computer Science & Engineering, Procedia Computer Science 87 (2016), pp. 128 – 133.

[9]. Md. Tanzim Khorshed, A.B.M. Shawkat Ali, Saleh A. Wasimi, "A survey on gaps, threat remediation challenges and some thoughts for proactive attack detection in cloud computing", Future Generation Computer Systems 28 (2012) pp. 833–851.

[10]. Tom White," Hadoop: The Definitive Guide", O'Reilly Media, Inc, 2009

[11]. S. .J. Samuel, Koundinya RVP, K. Sashidhar, and C.R. Bharathi, "A SURVEY ON BIG DATA AND ITS RESEARCH CHALLENGES", ARPN Journal of Engineering and Applied Sciences, VOL. 10, NO. 8, MAY 2015

[12]. M.M. Potel, C A Dhote, D.H. Sharma, "Homomorphic Encryption for Security of Cloud Data", Procedia Computer Science 79 (2016), pp. 175-181.

[13]. Zhong Wang, Qian Yu, "Privacy trust crisis of personal data in China in the era of Big Data: The survey and countermeasures", Computer Law & Security Review 31(2015), pp.782-792

[14]. N. Kshetri, "Big data's impact on privacy, security and customer welfare", Telecommunicaitons Policy 38 (2014), pp.1134-1145.

[15]. S.Akter, S.F.Wamba, A.Gunasekaran, R. Dubey, S. J.Childe, "How to improve firm performance using big data analytics capability and business strategy alignment?", International Journal of Production Economics, 182 (2016), pp.113-131.

[16]. A. Siddiqa, I.A.T. Hashem, I.Yaqoob, M.Marjani, S. Shamshirband, A. Gani, F.Nasaruddin, "A Survey of big data management: Taxonomy and state-of-the-art", 71 (2016), pp. 151-166.

[17]. R. Shaikh, M. Sasikumar, "Trust Model for Measuring Security Strength of Cloud Computing Service", Procedia Computer Science 45 (2015), pp 380-389

[18]. Vennila.S, and Priyandarshini J. , "Scalable Privacy Preservation in Big Data A Survey", Procedia Computer Science 50 (2015), pp. 369-373

[19]. B. H. Krishnaa, Dr. S. Kiranb, G. Muralia, R.P.K. Reddy, "Security Issues In Service Model Of Cloud Computing Environment", Procedia Computer Science 87 (2016),pp 246-251

[20]. N. Khana, A. Al-Yasirib, "Identifying Cloud Security Threats to Strengthen Cloud Computing Adoption Framework", Procedia Computer Science 94 (2016), pp 485-490

[21]. S.A. Hussain, M.Fatima, A. Saeed, I. Raza, R. K. Shahzad, "Multilevel classification of security concerns in cloud computing", Applied Computing and Informatics, 2210-8327, 2016, King Saud University Published by Elsevier B.V.

[22]. C. Saadia, H. Chaouib, "Cloud Computing Security Using IDS-AM-Clust, Honeyd, Honeywall and Honeycomb", Procedia Computer Science 85 (2016), pp 433-442