ISSN No. 0976-5697

Volume 8, No. 3, March – April 2017

International Journal of Advanced Research in Computer Science

REVIEW ARTICLE

Available Online at www.ijarcs.info

Security Challenges for Communications on IOT & Big Data

Dr. Pankaj Pathak
Sr. Asstt. Professor
Department of Information Technology
Balaji Institute Modern Management, Pune (MH) India

Dr. Nitesh Vyas
Asstt. Professor
Department of Computer Science,
Shri Vaishnav Institute of Management, Indore (M.P.) India

Someshwar Joshi
Asstt. Professor
Department of Computer Science,
Shri Vaishnav Institute of Management, Indore (M.P.) India

Abstract: Internet of Things or "IoT" refers to the highly interconnected network of heterogeneous devices where all kinds of communications seem to be possible, even unauthorized ones. In such a case, the security requirement for such network becomes critical. Presently abundant data is available and it is growing day by day. Big data is a collection of data sets which is very huge in size as well as complex. Presently Big data is one of the most burning topics in IT industry. It is going to play important role in future. Big data provides the new way for data management and use. The paper presents highlights on Internet of things and big data available on it. It also depicts the impacts and importance of these terminologies in the current scenario. Further we discusses the security challenges while we exploit the big data through IoT. Then we discuss various security enhancing techniques while communicating on Internet of things.

***Keywords*:** IOT, Big Data, security, communication, Data analytics

I. INTRODUCTION

When taking into account the monetary value created from technology, as well as the potential for new market opportunities, it is estimated that the Internet of Things will generate $14.4 trillion in net profit for enterprises over the next two decades. Organizations across all industries have started to develop and implement their own IoT strategies with the motive toward seizing the opportunity this new era presents.Internet of things (IOT) enables any device to be able to connect any other device using the internet. To highlight any device aspect the term internet of everything is also used. The Internet of Things is all about collecting data from various sources and making it useful in ways that enhance how we go about our business. The tremendous volume of data that will be coming in from devices presents a huge challenge for IoT solution providers. Big Data solutions will be overcoming this challenge by giving us the capacity to analyze data, and discover relevant trends and patterns.

Big data refers to collections of data sets with sizes beyond the ability of commonly used software tools such as database management tools or traditional data processing applications to capture and analyze within a stipulated time. Big data is characterized by '4 Vs': volume, variety, velocity and veracity. That is, big data comes in large amounts (volume), is a mixture of structured and unstructured information (variety) arrives at (often real-time) speed (velocity) and can be of uncertain provenance (veracity). Big data sizes are constantly increasing, ranging from a few dozen terabytes in 2012 to today many petabytes of data in a single data set. To meet the demands of handling such large quantities of data, new platforms of "big data" tools are being developed.Big data brings with it tangible benefits for any company willing to use it. The advantages of leveraging big data are real and oftentimes far-reaching, which is why so many organizations have adopted big data for their own operations.

For a long time, communication over the Internet has largely depended on the use of IP addresses to identify communicating parties. Some IoT uses cases will require a new technique of communication technologies that are able to provide greater security and more efficient communication. Pitfalls are still plentiful, and few represent as much of a problem as big data security. Businesses may be willing to use big data, but they must also be aware that security remains a top concern. This is in part because the technology is advancing so rapidly that the solutions to security problems often fall behind. If a business wants in on the enabling world of big data analytics, they'll need to be aware of some of the biggest security concerns first. IOT enabled devices would generate and transmit so much data that security issues [1] as well as managing the life cycle of those data are other dimensions that need to be addressed.

II. IMPACT OF IOT ON BIG DATA:

IoT and big data basically are two sides of the same coin. Managing and extracting value from IoT data is the biggest challenge that companies face. Big data is a terminology in both the tech and business worlds. Referring to the vast amounts of data generated by connected technology, big data is a tool that many businesses can use to make their advertising and other marketing efforts more effective. Data and using data for analytic purposes is not new, but what is new is the vast amounts of data now available to us, and that data has come available largely due to the Internet of Things (IoT).So, if the Internet of Things is not the only source of big data, are the two really connected? What is the impact of IoT on big data? The key is in how it is changing big data

and the way companies use that data.The IoT and big data are clearly growing apace, and are set to transform many areas of business and everyday life. But which particular sectors are likely to feel the IoT/big data disruption first? In its 2015 Internet of Things predictions, according to IDC, Presently over 50% of IoT activity is centered in manufacturing, transportation, smart city, and consumer applications, but within five years all industries will have rolled out IoT initiatives.

New generation of IoT applications is required to address specific business needs such as predictive maintenance, loss prevention, asset utilization, inventory tracking, disaster planning and recovery, downtime minimization, energy usage optimization, device performance effectiveness, network performance management, capacity utilization, capacity planning, demand forecasting, pricing optimization, yield management, and load balancing optimization. Fig.1 shows the process of obtaining large data through various application interfaces available on internet. The big data then processed by using big data analytics which further can be utilized by enterprises for their strategical decisions and to increase their sales performance.

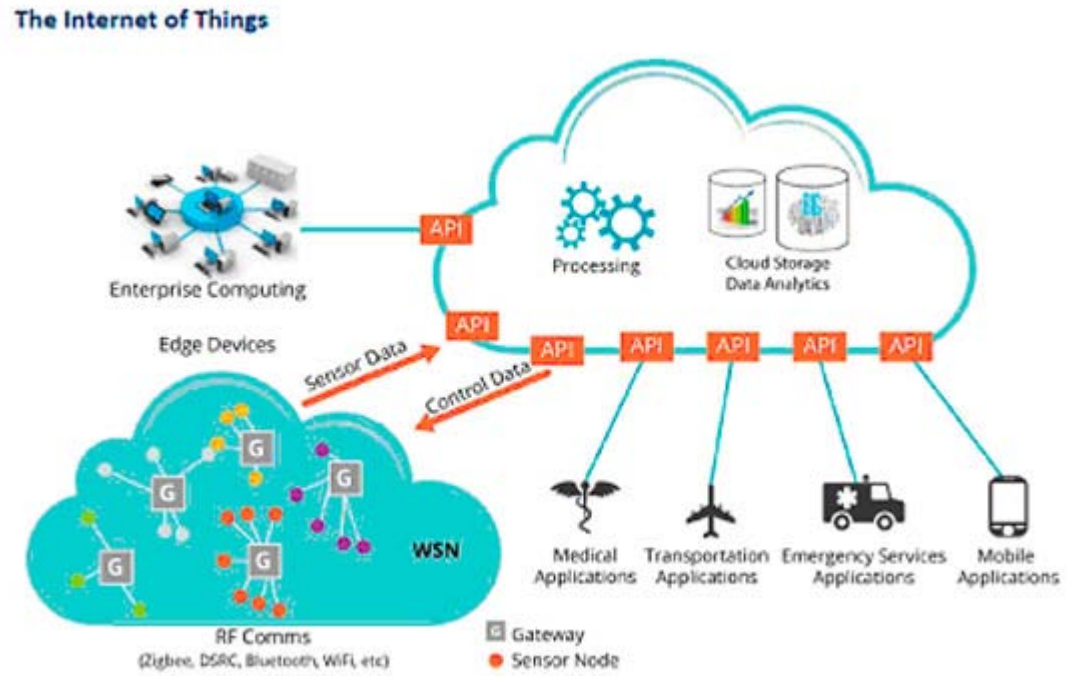

Fig.1 Internet of Things

III. IoT & Big Data : A New Competitive Advantage

The use of Big Data is becoming an important way for leading companies to outperform their peers. In most industries, established competitors and new entrants alike will leverage data-driven strategies to innovate, compete, and capture value. Organizations are setting up a proper analytics platform/infrastructure to analyze the IoT data.Big Data Analytics is aimed to enable organizations to make better decisions. Data Scientists, predictive modelers and other analytics professionals deal with huge amounts of transactional data and use Big Data Analytics to tap this data that may be untapped through conventional Business Intelligence programs.Big data analysiscan be done with the software tools commonly used as part of advanced analytics disciplines such as predictive analytics, data mining, text analytics and statistical analysis. Due to the Volume and Velocity of Big Data, data warehouses are unable to handle the processing demands posed by data sets that are being updated in real time and continually, such as the movements on social media websites. The newer technologies involved in Big Data Analytics involve Hadoop and related tools such as YARN, MapReduce, Spark, Hive and Sas well as NoSQL databases.

Big Data provides new growth opportunities and entirely new categories of companies, by analyzing and aggregating industry data. Many of these will be companies that stands in the middle of large information availability where data about products and services, buyers and suppliers, consumer preferences and intent can be caught and analyzed. The leaders have started aggressively cultivating the Big Data capabilities. An IoT device generates continuous streams of data in a scalable way, and companies must handle the high volume of stream data and perform actions on that data. In response to the actions can be event correlation, metric calculation, statistics preparation, and analytics. In a normal big data scenario, the data is not always stream data, and the actions are different. Building an analytics solution to manage the scale of IoT data should be done with these differences in mind.

IV. SECURITY CHALLENGES WITHIN IOT & BIG DATA:

The Internet of Things (IoT) has a data problem. Everyone is claiming to be the world's smartest something. But that sprawl of devices, lacking context, with fragmented user groups, is a huge challenge for the rapid growing industry.This paper describes the security challenges when organizations start moving sensitive data to a Big Data repository. It provides the different threats and the security control framework to address and mitigate the risk due to the identified security threats.

The collection, storage, manipulation and retention of massive amounts of data have resulted in serious security and privacy considerations. Various regulations are being proposed to handle Big Data so that theprivacy of the individuals is not violated. For example, even if personally identifiable information is removed from the data, when data is combined with other data, an individual can be identified. This is essentially the inference and aggregation problem that data security [5] researchers have been exploring for thepast four decades. This problem is exacerbated with the management of Big Data as different sources ofdata now exist that are related to various individuals.

In some cases, regulations may cause privacy to be violated. For example, data that is collected (e.g., email data) has to be retained for a certain period of time (usually 5 years). As long as one keeps such data, there is a potential for privacy violations. Too many regulations can also stifle innovation. For example, if there is a regulation that raw data has to be kept as is and not manipulated or models cannot be built out of the data, then corporations cannot analyze the data in innovative ways to enhance their business. This way innovation may be stifled.Therefore, one of the main challenges for ensuring security [6] and privacy when dealing with big data is to come up with a balanced approach towards regulations and analytics. That is, how can an organization carry out useful analytics and still ensure the privacy of individuals? Numerous techniques for privacy preservingdata mining, privacy-preserving data integration and privacy-preserving information retrieval have been developed. The challenge is to extend these techniques for handling massive amounts of data resides on network. Next the Big Data management strategies such as access methods and indexing and query processing have to be secure. So the question is how can policies for different types of data such as structured, semi structured, unstructured and graph data be integrated? Since Big Data may result from combining data from numerous sources, how can you ensure the quality of the data?

Complexity is inherent in the adoption, implementation, and maintenance of big data technologies. Most of the organizations cite complexity as the main barrier to deploying big data analytics to enhance their enterprise's cyber defense. Organizations deploying the Big Data Setup do not know how a big data solution would affect their legacy technology environment, and who has the right expertise to manage the new technologies. Successfully implementing a big data solution does require sophisticated technologies that will store, organize, and further analyze vast and varied data sets. Interoperability among existing data environments and new technologies is contingent upon choosing the right technologies and having the right expertise to implement them. Moreover, enterprises both large and small lack specially trained analysts to design these big data systems and use the results of the analysis.

The increasing stealth and sophistication of cyberattacks can put a strain on even the most generous security budgets, which are already spread thin addressing risk; insecure mobile devices and apps, including personally owned devices and apps entering the workplace, non-compliance with regulations, data breaches, social engineering tactics, insider negligence, and use of insecure cloud services. Both large and small enterprises cite insufficient budgets as a reason they have not yet adopted big data analytics. The various costs of deploying a big data solution can be many, and often include storage, computers, data tools, and data visualization frameworks. However, the emergence of big data solutions offered as cloud services, combined with a growing number of firms offering these services, may be seen as an indication that costs will fall and adoption rates will increase. Beyond the capital costs of adopting big data tools are the opportunity costs. When purchasing new technologies, it affects an enterprise's ability to invest in and maintain other technologies crucial to their security, such as firewalls and detection software.

Another security challenge for Big Data management and analytics is to secure the infrastructures. Manyof the technologies that have been developed including Hadoop, MapReduce, Hive, Cassandra, PigLatin,Mahout and Storm do not have adequate security protections. The question is, how can these technologiesbe secured and at the same time ensure high performance computing?Finally, the entire area of security, privacy, integrity, and data quality and trust policies have to be examined within the context of Big Data security. What are the appropriate policies for Big Data? How can these policies be handled without affecting performance? How can these policies be made consistent and complete?

We have listed just some of the challenges with respect to security and privacy for big data. That is, we cannot incorporate security into each and every big data technology that is being developed. We need to have a comprehensive strategy so that security can be incorporated while the technology is being developed. We also need to determine the appropriate types of policies and regulations to enforce before Big Data technologies are employed by an organization.

V. SECURITY ENHANCING TECHNIQUES

Security will be a major challenge as billions of devices join the Internet of Things, and different technologies will compete to provide appropriate solutions. Managing authentication on a large scale is a challenge already successfully met by the telecommunications industry in the form of mutual authentication with secret credentials in the SIM. This technology can now be extended to provide equally strong authentication and data integrity for the Internet of Things. Many security techniques have been proposed over the last fifteen years, ranging fromcryptographic techniques such as oblivious data structures that hide data access patterns to dataanonymization techniques that transform the data to make more difficult to link specific data records tospecific individuals. However, many such techniques either do not scale tovery large datasets and/or do not specifically address the problem of reconciling security with privacy. Atthe same time, there are a few approaches that focus on efficiently reconciling security with privacy andwe discuss them in what follows.

Privacy-preserving is the important concern while deploying the Big Data techniques.Cryptographic approaches [9], such as secure set intersection protocols, and may alleviate such concerns. But, these techniques do not scale for large datasets. Recent approaches based on data transformationand mapping into vector spaces, and combination of secure multiparty computation (SMC) and data sanitization approaches such as differential privacy, and k-anonymity have addressedscalability. Securitymodels and definitions also need to be developed supporting security analysis and proofs for solutions [3] combining different security techniques, such as SMC and differential privacy.

Privacy-preserving while doing data mining is typically performed onbig centralized data warehouses collecting all the data of interest. However, centrally collecting all dataposes several privacy and confidentiality concerns when data belongs to different organizations. An approachto address such concerns is based on distributed collaborative approaches by which the organizationsretain their own datasets and cooperate to learn the global data mining results without revealing thedata in their own individual datasets. Fundamental work in this area includes: (i) techniques allowing twoparties to build a decision tree without learning anything about each other's datasets except for what canbe learned by the final decision tree; (ii) specialized collaborative privacy-preserving techniques forassociation rules, clustering, k-nearest neighbor classification. These techniques are however stillvery inefficient. Novel approaches based on cloud computing and new cryptographic primitives should beinvestigated.

Privacy-preserving through biometrics authenticationrequire recording biometrics templates of enrolled users and then using these templates for matching withthe templates provided by users at authentication time. Templates of user biometrics represent sensitiveinformation that needs to be strongly protected. In distributed environments in which users have to interactwith many different service providers, the protection of biometric templates becomes even more complex.A recent approach addresses such an issue by using a combination of perceptual hashing techniques,classification techniques, and zero-knowledge proof of knowledge (ZKPK) protocols. Under suchapproach, the biometric template of a user is processed to extract from it a string of bits which is then furtherprocessed by classification and some other transformation. The resulting bit string is then used, togetherwith a random number, to generate a cryptographic commitment. This commitment represents anidentification token that does not reveal anything about the original input biometrics. The commitment isthen used in the ZKPK protocol to authenticate the user. This approach has been used for secureuse on mobile phones. Much work remains, however, to be done in order to reduce the false rejectionrates. Also different approaches to authentication and identification techniques need to be investigatedbased on recent homomorphic encryption techniques.

Communication [2] in the IoT and Big Data should be protected by providing the security services discussed above. Using standardized Internet security mechanisms we can provide communication security at different layers of the IP stack, each solution has its own pros and cons. The communication security can be provided E2E between source and destination, or on a per-hop basis between two neighboring devices.

Though the communication security that protects the messages with confidentiality and integrity services, a number of attacks are possible against networks mainly to breach availability security services. The attacks happens, are aimed to disrupt networks by interrupting, for example, the routing topology or by launching DoS attacks. Intrusion Detection Systems (IDS) are required to detect impostors and malicious activities in the network, and firewalls are necessary to block unauthorized access to networks. It is important to not only protect communication and networks but to also safeguard the stored sensitive data in an IoT device. Most of the IoT devices are tiny wirelessly connected resource-constrained nodes, and practically it is neither possible to physically guard each device nor to protect them with hardware-based tamper-resistant technologies such as with the use of smart cards or Trusted Platform Modules.

VI. COMBINED SECURE STORAGE AND COMMUNICATION FOR THE INTERNET OF THINGS

IoT security is the area of endeavor concerned with safeguarding connected devices and networks in the internet of things (IoT). The Internet of Things involves the increasing prevalence of objects and entities – known, in this regard as things -- provided with unique identifiers and the ability to automatically transfer data over a network. Much of the increase in IoT communication comes from computing devices and embedded sensor systems used in machine-to-machine (M2M) communication, smart energy grids, home and building automation, vehicle to vehicle communication and wearable computing devices.

The future Internet of Things (IoT) may be based on the existing and established Internet Protocol (IP). Many IoT application scenarios will handle sensitive data. However, as security requirements for storage and communication are addressed separately, work such as key management or cryptographic processingis duplicated. Cryptographic processing [4] is one of the main resource, while providing communication security. These operations include encryption and decryption, key and hash generation, and sign and verify hashes. For secure communication on IoT cryptographic methods and data formats defined for data processing before storage. This requires us to store not only data but also all header information that is involved in the cryptographic processing. Data on nodes must be secured when stored and transported in order to implement a comprehensive security solution. As resource-constrained embedded systems are limited in resourcesit is necessary to find efficient solutions.

Secure communication [2] in IoT-type systems currently requires many levels of configuration and/or application-level proprietary algorithms, which discourages users from implementing protection and often encourages functionality to be prioritized over security. The lack of secured links exposes data to attacks and theft, and fraudsters and hackers are already beginning to show increasing interest in this area.

Generic Bootstrapping Architecture (GBA) [8] technology, based on the Authentication and Key Agreement (AKA) protocol used in network access authentication, can provide device authentication and support communication security at the transport layer. GBA is a key bootstrap method standardized by the 3GPP. The protocol enables the creation of service or application keys through authentication using 3GPP subscription credentials. The credentials are typically stored on a SIM card, which runs on an UICC. Alternatively, they can be provided as remotely managed credentials stored and managed on an embedded UICC (eUICC) such as the GSMA-specified eSIM.

One way to solve these issues is to leverage the existing 3GPP [7] network authentication framework that is an inherent part of cellular networks. Cellular networks use strong authentication and communication security [10], where the Universal Integrated Circuit Card (UICC) acts as the secure storage point of the secret keys on the device side. Building on this framework, GBA technology provides the means to implement AKA with GBA generating time-limited session keys during the SIM authentication. The generated keys can be used for creating, for example, a TLS-based protected communication channel. Furthermore, GBA can also be used over non-cellular connectivity options like Wi-Fi. For capillary network, GBA can also cover non-3GPP devices, in other words, those devices that do not have a UICC or cellular network access.

VII. CONCLUSION

Big Data is changing the way we perceive present world. The impact which big data and internet of things has created is great and will continue to create canripple through all facets of our life. Global Data is on the rise, by 2020, we would have quadrupled the data we generateevery day. This data would be generated through a wide array of sensors we are continuously incorporating in our lives. Data collection would be aided by what is today dubbed as the “Internet of Things”. Through the use of smart bulbs to smart cars, everyday devices are generating more data than ever before. These smart devices are incorporated not only with sensors to collect data all around them but they are also connected to the grid which contains other devices. Big Data is also changing things in the business world. Companies are using big data analysis to target marketing at very specific demographics. At the same time, if a business wants in on the enabling world of big data analytics, they’ll need to be aware of the security concerns first. IOT enabled devices would generate and transmit so much data that security issues as well as managing the life cycle of those data are other dimensions that need to be addressed. For ensuring secure communications through IoT new security techniques should be adapted rapidly and it should be a continuous process.

VIII. REFERENCES

[1] L. Okman, N. Gal-Oz, Y Gonen, E. Gudes and J Abramov, “Security Issues in NoSQL Databases” in TrustCom IEEE Conference on International Conference on Trust, Security and Privacy in Computing and Communications, pp 541-547, 2011.

[2] J. Feng, Y. Chen, and P. Liu, “Bridging the Missing Link of Cloud Data Storage Security in AWS,” the 7th IEEE Consumer Communications and Networking Conference - Security for CE Communications (CCNC ‘10), Las Vegas, Nevada, USA, January 9 - 12, 2010.

[3] Dona Sarkar, Asoke Nath, “Big Data – A Pilot Study on Scope and Challenges”, International Journal of AdvanceResearch in Computer Science and Management Studies (IJARCSMS,

ISSN: 2371-7782), Volume 2, Issue 12, Dec31, Page: 9-19(2014).

[4] "An Inside-Out Approach to Enterprise Security," Oracle/CSO Custom Solutions Group white paper, 2013

[5]. Raghav Toshniwal et al., "Big Data Security Issues and Challenges", International Journal of Innovative Research in Advanced Engineering (IJIRAE), Issue 2, Volume 2 (February 2015).

[6] Top 10 Big Data Security and Privacy Challenges, Cloud Security Alliance, 2012 https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Big_Data_Top_Ten_v1.pdf

[7]. Ericsson White paper "Bootstrapping security" 284 23-3284 Uen | February 2016

[8] 3GPP, "3GPP Specification detail; 3GPP TS 33.220 – Generic Authentication Architecture (GAA);Generic Bootstrapping Architecture (GBA)," accessed February 2016, available at:http://www.3gpp.org/DynaReport/33220.htm

[9] W. Jonker and M. Petkovic´ (Eds.), "Data Security challenges and research operations" SDM 2013, LNCS 8425, pp. 9–13, 2014. DOI: 10.1007/978-3-319-06811-4_2, Springer International Publishing Switzerland 2014

[10] The ESG White Paper, "The Big Data Security Analytics Era Is Here", January 2013