A Public Key Cryptography Security System For Big Data

M. Koushikaa
3rd Sem Student
Department of Computer Science & Engineering
Knowledge Institute of Technology, Salem

S. Habipriya
3rd Sem Student
Department of Computer Science & Engineering
Knowledge Institute of Technology, Salem

Mr.S.S.Aravinth
Assistant Professor
Department of Computer Science & Engineering
Knowledge Institute of Technology, Salem

Mr.T. Karthikeyan
Assistant Professor
Department of Computer Science & Engineering
Knowledge Institute of Technology, Salem

Dr. V. Kumar
Professor & Head of the Department
Department of Computer Science & Engineering
Knowledge Institute of Technology, Salem

Abstract

Big data is the one which is emerging in IT industry. As we know, it is used to describe the massive data produced by various applications and systems. The database becomes so large and difficult to manage it. Normally the data is stored in petabyte, Exabyte and even zeta bytes. For decision making and accurate analysis these data's are much important for future usage. So, to protect these data against modification and theft, we need to go for big data security which is very essential. Many security challenges involved in big data such as lack of tools and storage for distributed data sets and processing analysis. As big data comes more vital to all industries the challenges also are increased so high. The collection of massive heterogeneous data and aggregative data are augmented every day. Hence it's very tough to maintain the secrecy of all forms of data. Medical health care centers, bio chemical researchers and enterprises are wanted to maintain the efficient security for their huge collections of data.
Keywords: Big data, Analytics, Security, Public key infrastructure and Applications.

I. INTRODUCTION

The term 'big data' is widely used everywhere in these days. The importance of big data has existing fast. It refers to a wide range of large data that sets almost impossible to manage using some traditional data management tools, due to their complexity. It also exist in finance and business processes where enormous amount of stock exchange parameters, banking appliances, online shopping.

Big data is also used in other areas of researches as astronomy, oceanography and engineering etc., even now the companies are introducing the technological solutions to data analytics.

II. BIG DATA TECHNOLOGIES AND TOOLS

The big data platform is underlying all of the proof point's solutions, which uses both proprietary and open source 'big data' technologies.Proof point's has the use of technologies including hadoop, cassandra, Map Reduce in the delivery of security enterprise and solutions of data protection.

A big data technique that enables proof point's to rapidly analyze and gain insight from massive amount of data.

The result of data protection is better, higher performance and superior scalability for organization.

A. Big data platform that benefits customers:

- Protection against targeted threads.
- High performance search.
- Accurate, real time reputation.

III. HADOOP

High-availability distributed object –oriented platform is the software framework which analyses the data and distributes applications on different servers. The most popular for big data processing is hadoop. It was designed originally without security in mind, and hadoop's security model has continued to evolve. It points out to continue as security professionals potential and big data security risks with hadoop, this has led to continued security modification to hadoop.

Fig. 1: Data Sensitive Parameters

To meet security requirements, the great demand for hadoop is resulting the ongoing changes to hadoop. Initially, there was no security model and also there was no data privacy and authenticate users or services. Later, it becomes the access control for security that becomes the popular platform for data analysis and processing.

E.g. writing a new task tracker and registering itself as a hadoop service.

It is used in maintaining scaling, error handling and security large scale of data which are structured and unstructured. A Map Reduce job that divides the given inputs of data set processed by map task in parallel. This is followed by the steps of reducing tasks and this reduced tasks use the output of maps to obtain the final result of the job.

IV. PUBLIC KEY INFRASTRUCTURE (PKI) FOR BIG DATA SECURITY

A PKI is the system which we can trust the third party user to identity inspection and assurance. This is done by normally, certificate authority and uses cryptography involving private and public keys.

A typical PKI system consists of

- Client software
- A certificate authority server
- May involve smart cards
- Operational procedures

Public key cryptography is advance in IT security. It enables the tradeoff between entities are confidential in open network. Apart from PKC also it enables the protection techniques that has no analogue in such traditional cryptography, importantly digital signature.

E.g. Software downloads and updates.

The PKC has the public and private keys of proper management which does not require confidential exchange of secret keys but PKC still in vital importance as the public key that is verifiably authentic and the private key remains private.

So called is public key infrastructure, which manages the key pairs.

A. How Public Key Infrastructure Works Using Authentication:

Firstly, the certificate authority (CA) will check the user and different CA'S have different identity validation process. CA acts as a trust and independent provider of DC. In this some of them may grant user as digital certificate (DC) which will contain only the name and email address. While the CA users can involve personal interviews, background checks etc. if the user is granted DC often they will undergo two components i.e., private and public keys.

E.g. for DC: The user wishes to send an email to his business associates, he wants sign digitally the email with his private key.

The email sent to his business associates. Then they will decrypt the message from the user's sending public key.

This example of DC provides secret information can be shared with user authentication, without exchange the secrets in advance. The PKI are also used in Medical Application Systems. The role of PKI is that assures this binding is called the registration authority (RA) which ensures that the public key is bound to as individual, which is assigned in a way that ensures non-reputation.

B. Uses of PKI:

Many users and from any of the several vendors which includes providing public keys and binding to user identities which are used for:

- Encryption or sender authentication of email message and also document e.g. using open PGP.
- Authentication of users to applications e.g. SMART CARD login system.
- Bootstrapping which enables secured communication with protocols, such as internet key exchange.

C. Why PKI:

PKI enables unsecure public network such as internet to securely and privately transfer their money and data through the use of public and private key pair cryptographic that is obtained and shared through a trusted authority. The PKI provides for digital certificates that can identity of an individual or an organization that can store and revoke their certificates when necessary.

A number of products are now offered to enable a company or a group of companies to implement PKI. The e-commerce and business to business companies are increased in demand for PKI solutions with related ideas are of virtual private network (VPN) and IP security (IPsec) standards.

A conventional DBMS data is stored and maintained by the DBA. It has access and flow control. By analyzing with statistics, data is coming from different external resources. It has to be ensure that the data should not be compressed while transmission from source to big data server.

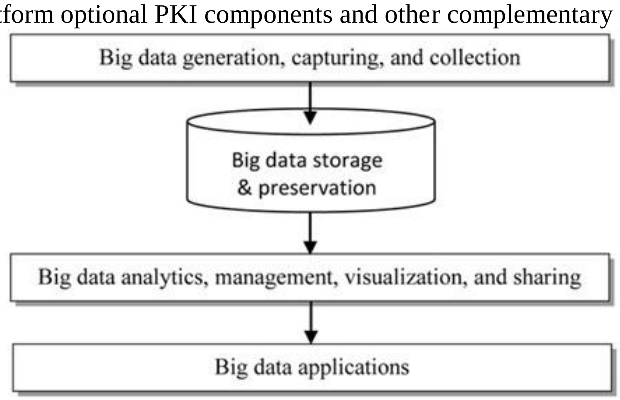
The PKI provides the trust that need to the relationship between the data source and big data server where the source data is encrypted by the private key along with DC so the server can check whether the data is modified or altered during the transmission. The PKI is also used in financial, e-commerce, healthcare sectors and already been used in NASA, DoD, other states are used for e-governance application for authenticate and integrity purpose.

D. An integrated security infrastructure for encryption, digital signature and CA:

A PKI maintains a trustworthy networking environment and establishes by providing key and certificate management services which enable encryption and digital signature capacity across many applications. It is very transparent in manner and easy to use.

The world's first commercially available PKI that entrusts first was released in 1994. By managing the full life cycles of digital certificate based identities and encryption that enables PKI entrust authority, digital signature and certificate authentication capabilities to be consistent and transparently applied across a broad range of applications and platforms.

Add more security management capabilities that include self-registration, self-recovery or inventory of digital identities and PIN authentication of using the platform optional PKI components and other complementary solutions.



E. Advantages of PKI approach:

The primary advantage of PKC is increased security and convenience and also the public key systems is that they provide a method of digital signatures. It is the trust based system that enables the customer to depend upon.

There is no limit to access the PKI and the users can maintain their own certificates and authentication involves exchange of data between client and server only. It enables that no third party can intervene the system.

V. CONCLUSION

In our fast-paced and connected world where big data is a king, it is critical to understand the importance of security as we processed the wide range of large data sets which are easily available. These data's are securely carried out for genuine analysis for the future. The PKI based authentication can be employed to ensure that big data is genuine. The scalable characteristics of PKI are most suitable for the big data environment. So PKI is used for Big Data security.

REFERENCES

[1] Michael Minelli, Michelle Chambers, and Ambiga Dhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses", Wiley,2013.
[2] P. J. Sadalage and M. Fowler, "NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence", Addison-Wesley Professional, 2012.
[3] Tom White, "Hadoop: The Definitive Guide", Third Edition, O'Reilley, 2012.
[4] Eric Sammer, "Hadoop Operations", O'Reilley, 2012.
[5] E. Capriolo, D. Wampler, and J. Rutherglen, "Programming Hive", O'Reilley, 2012.
[6] Lars George, "HBase: The Definitive Guide", O'Reilley, 2011.
[7] Eben Hewitt, "Cassandra: The Definitive Guide", O'Reilley, 2010.
[8] Alan Gates, "Programming Pig", O'Reilley, 2011.