A Reviewed Paper on- Cryptographic RSA approach to Solve Big Data Security Issues.

Mrs. Bareen Shaikh

Asst. Professor (Computer Science Dept.)

MIT Arts Commerce Science College, Alandi(D), Pune

mail-id: bkshaikh@mitacsc.ac.in

Mrs. Shital Ghotekar

Asst. Professor (Computer Science Dept.)

MIT Arts Commerce Science College, Alandi(D), Pune

mail-id: svghotekar@mitacsc.ac.in

Mrs. Sangeeta Borde

Asst. Professor (Computer Science Dept.)

MIT Arts Commerce Science College, Alandi(D), Pune

mail-id: smborde@mitacsc.ac.in

Abstract:

RSA makes its big push into using big data to help businesses secure their information. All organizations prefer networking system for communicating with their consumers or clients at a faster rate compared to earlier years and achieve success in dynamic environment. The term Big Data refers to large-scale information management and analysis technologies that exceed the capability of traditional data processing technologies. Big Data is differentiated from traditional technologies in three ways: the amount of data (volume), the rate of data generation and transmission (velocity), and the types of structured and unstructured data (variety).Big Data analytics can be leveraged to improve information security and situational awareness. RSA technique which is one of the cryptography techniques is very secure and safe technique for transferring the confidential data. This review paper shows how RSA algorithm improves security aspect of big data technology and makes application of big data technology safer to operate by organizations.

Keywords: Big Data, Networking, Security, RSA, cryptography.

Introduction:

Big data describe data sets that are too large, too unrefined, too fast changing for analysis using relational or multidimensional database techniques. Analyzing big data can require dozens, hundred or even thousands of servers running massively parallel software. What truly distinguishes big data aside from volume and variety is the potential to analyze it to uncover new insights to optimize decision making. Big data refers to collection of massive data with processing and data retrieval. Big data have three important properties like volume, velocity and variety.

A. Volume:

As name indicates data is in large amount. Daily terabytes to zettabytes of data is collected from various resources.

B. Velocity :

Now a day's social sites are favorably used. Data comes at very high speed and with high frequency from social sites just like Gmail, Facebook, Twitter and WhatsApp.

C. Variety:

Data comes in the structured or in the unstructured form just like image, video, sounds etc.

RSA algorithm is based on the mathematical fact that it is easy to find and multiply large prime number together, but it is extremely difficult to factor their product.RSA is based on a public key system that was made by Mr.Ron Rivest, Mr.Adi Shamir, and Mr.Leonard Adleman in 1978. Three basic steps are required to complete the process of RSA operations which are; key generation, encryption and decryption.

First, messages are converted into numbers (integers), and then the numbers are manipulated according to the prescribed encryption scheme. Here is the description of the RSA cryptosystem. For the implementation of RSA, following steps are followed:

1. Choose two large prime number p and q.
2. Calculate n= p x q.
3. Chooses e with $(e, (p-1)(q-1)) = 1$ and computes d with $de \equiv 1 (\mathrm{mod}\ (p-1)(q-1))$.
4. Makes n and e public and keeps p, q, and d secret.
5. Sender encrypts m as $c \equiv m^e (\mathrm{mod}\ n)$ and sends c to Receiver
6. Bob decrypts by computing $m \equiv c^d (\mathrm{mod}\ n)$.

Objective:

The objective of this paper is to study how Big Data can improve information security by using RSA algorithm and also identifying the best practices in Big Data privacy and increasing awareness of the threat to private information. Big Data tools have the potential to provide a significant advance in actionable security intelligence by reducing the time for correlating, consolidating, and contextualizing diverse security event information, and also for correlating long-term historical data for forensic purposes.

Context of Big Data & RSA Security:

RSA security analytics is a security solution that helps security analysts detect and investigate threats that are often missed by other security tool. By combining big data security data collection, management, and analytics capabilities with full network and log based visibility and automated threat intelligence, security analyst can better detect, investigate, and understand threats they could often not easily see or understand before. Ultimately this improved visibility and speed helps organization reduce an attacker's free time in their computing environment.

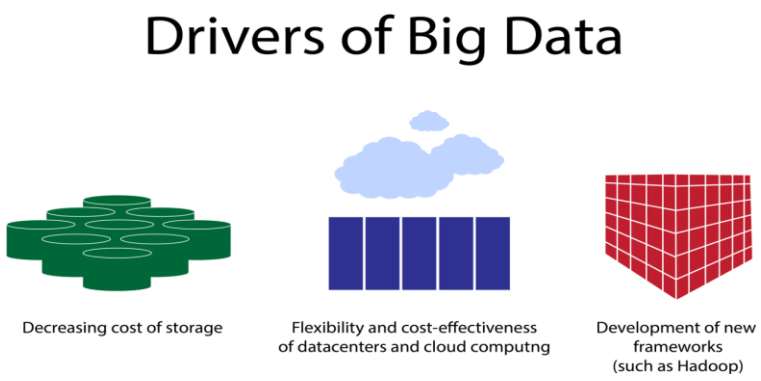
RSA has unveiled a new tool designed to let enterprises detect security threats more quickly than current technologies permit by combining big data management and analytics approaches with traditional network monitoring and threat detection. Big Data analytics – the process of analyzing and mining big data – can produce operational and business knowledge at an unprecedented scale and specificity. The need to analyze and leverage trend data collected by businesses is one of the main drivers for Big Data analysis tools. The technological advances in storage, processing, and analysis of Big Data include

(a) The rapidly decreasing cost of storage and CPU power in recent years.

(b) The flexibility and cost-effectiveness of datacenters and cloud computing for elastic computation and storage.

(c) The development of new frameworks such as Hadoop, which allow users to take advantage of these, distributed computing systems storing large quantities of data through flexible parallel processing.

These advances have created several differences between traditional analytics and Big Data analytics.



Security issues arrive in technologies such as memory management, transaction management, virtualization and networking for big data. Big data deals with storing the data, processing the data, retrieval of data in above technologies security issues. Algorithms and systems must be designed and implemented in order to identify actionable security information from large enterprise data sets and drive false positive rates down to manageable levels. Many challenges must be overcome to realize the true potential of Big Data analysis. Among these challenges are the legal, privacy, and technical issues regarding scalable data collection, transport, storage, analysis, and visualization.

The four important security issues of big data are authentication level, data level, network level and generic issues

A. Authentication level issues

There are many clusters and nodes present. Every node has a different priorities or rights. Nodes with administrative rights can access any data. But sometimes if any malicious node got administrative priority then it will steal or manipulate the critical user data. For faster execution with parallel processing, many nodes join clusters. In case of no authentication any malicious node can disturb the cluster. Logging plays an important role in big data. If logging is not provided then no activity is recorded which modify or deleted data. If new node joins the cluster then that will not be recognized because of logging absence. Sometimes users may also used malicious data if log is not provided.

B. Data level issues

In big data, data is very important part and also plays vital role. Data is nothing but some important and personal information about us by the government or social networking sites. Data level issues deals with data integrity and availability such as data protection and distributed data.
To improve efficiency, big data environments like Hadoop store the data as it is without encryption. If hacker access the machines, then it is impossible to stop him. In distributed data store, information is stored in many nodes with replicas for quick access. But if any replica or information from other node is deleted or manipulated by hacker then it will be difficult to recover that data.

C. Network level issues

There are many nodes present in clusters and computation or processing of data is done in these nodes. This processing of data can be done anywhere among the nodes in cluster. So it is difficult to find on which node data is processing. Because of this difficulty on which node security should be provided is going to be complicated. Two or more nodes can be communicate with each other or share their data/resources through network. Many times RPC (Remote Procedure Call) is used for communicating via network. But RPC is not securing until and unless it is encrypted.

D. General level issues

In big data environment many technologies are used for processing the data also some traditional security tools for security purposes. Traditional tools are developed over years ago. So these tools may not be performed well with new distributed form of big data. As big data uses many technologies for data storing, data processing and data retrieval, there may be some complexities occur because of these various technologies.

Big data promises for security

Big data new role in security comes at time when organization confronts unprecedented information risk arising from two conditions:

1. **Dissolving network boundaries:** As organization open and extend their data networks, they become more vulnerable to data misuse and theft. Corporate applications and data are also increasingly accessed through cloud services and mobile devices shattering network boundaries and introducing new information risks and threat vectors.
2. **Sophisticated Adversaries:** Oftentimes, cyber attack or fraud schemes perpetrated by advanced adversaries aren't detected until well after damage has been done.

In today's hyper-extended, cloud based, highly mobile business world, security approaches solely reliant on perimeter defenses or that require predetermined knowledge of the threat or direct control over all infrastructure elements being made obsolete. More agile approach based security operation are essential for meaningful security. The security based

innovation council advises organization to move to an intelligence driven security model, which relies on security related information from internal and external sources to deliver comprehensive picture of risk and security vulnerabilities.

Conclusion:

RSA approach reduces risk of compromise by using the latest analytics and detection techniques and threat intelligence to aid in the detection, investigation and response to security incidents. It reduce deployment risk and quicker time to value through proven validated architecture for collection analytics of data that produce actionable intelligence at enterprise scale. Its less reliance on data science expertise to leverage cutting edge analytic techniques.

References:

1. Cryptography and security – Second Edition by Atul Kahate.
2. "Approaches to Solve Big Data Security Issues and Comparative Study of Cryptographic Algorithms for Data Encryption". Vinit G. Savant
3. http://searchsecurity.techtarget.com/definition/RSA
4. http://www.computerworld.com/article/2495004/big-data/rsa-brings-big-data-analytics-to-security-threat-management.html
5. https://www.emc.com/collateral/industry-overview/big-data-fuels-intelligence-driven-security-io.pdf