Data Security Considerations from Regulatory Perspectives

Pankaj Gupta[1]
Assistant Professor
Birla Institute of Technology, Mesra, Noida Center

Chandan Kumar Barman[2]
Superintending Engineer (IT)
Directorate General of Hydrocarbons
Gov of India, Noida

Abstract— **The evolution of technology responsible for storing, managing and processing data has noticeably taken giant strides in recent times with the inception of technologies like Big Data, In Memory Computing etc. With wide scale business process automation initiatives taken by organizations of different sizes, more and more data are getting generated each passing day. The modern day data handling information systems are quite different from their traditional counterparts where RDBMS was the de-facto standard for data management. Today we need to deal with various structured, semi-structured and unstructured data classes like email, image, video, blogs, documents, live stream, xml/json data file etc.**

Security on the other hand till recently was considered to be a subject matter of network administrator where the primary goal was to protect the IT infrastructure perimeter. With increased adaptation and dependence on different data classes, data security has gained special interest in IT security landscape. However absence of well-defined data security guidelines plagues data governance in the industry. In this paper we have defined different facets of data security vulnerabilities that are required to be addressed as per well-known existing regulations and guidelines. Further we have cited no of industry recognized control references and framework relevant to specific data security requirement.

The compilation of technical controls for data security requirement as concluded by us may be used by any digital data custodian for data security conformity.

Keywords— ***Data Security, Data Governance, IT Security***

I. INTRODUCTION

IT, these days is no more a service discipline in any industry. Today we see industry where business processes are automated and backed by agile and rugged application software. Different classes of automation software like Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), Supplier Relationship Management (SRM) etc. run business operations of today's organization. Tactical and strategic management activities are aided by data warehousing and business analytics classes of software. In all, IT has been proved to be a business enabler for the industry as a whole. Organizations adopting and exploiting IT in a matured manner found to have clean edge over their counterparts operationally, tactically and strategically. Needless to say, IT assets comprising hardware, software, networks, data, information, process and procedures form the backbone of business operations of such organization. Any mishap or disruption in an organization's IT value chain can potentially cripple any modern day business operation. In other words IT itself manifests as a security risk for an IT dependent organization which if compromised can bog down the operations of the organization.

IT security was generally interpreted in a broader and subjective way. Till recently network firewall and anti-virus end point solutions were synonymous to IT security. However growing operational dependency on IT solutions has encouraged industry and statutory bodies to formally define IT security premise and associated guidelines. Delving deeper into any IT landscape in organizational perspective we can easily find that "information" and "data" are the most precious IT assets that are processed, harnessed and supported by the peripheral IT infrastructure. Traditional network and perimeter security mechanisms are normally intended to prevent external and unauthorized agents to enter into internal corporate network. However data stores holding data and information comprising traditional RDBMSes to recent Big Data environment present many new and special security vulnerabilities that are not addressed by network and perimeter security controls. The basic premise of information and data security revolves around three tenets. They are i) Confidentiality ii) Integrity and iii) Availability of data and information [1][2]. Confidentiality refers to preventing the disclosure of information to unauthorized individuals or systems. For example, a credit card transaction on the Internet requires the credit card number to be transmitted from the buyer to the merchant and from the merchant to a transaction processing network. Data integrity means maintaining and assuring the accuracy and consistency of data over its entire life-cycle. This means that data cannot be modified in an unauthorized or undetected manner. Availability refers to the uninterrupted accessibility of the information systems required to store, process, retrieve and transmit data and information. In all, data security should ensures that only authorized users (confidentiality) have access to accurate and complete information (integrity) when required (availability) [3].

Data security has gained special status as breach of data and information may lead to numerous legal disputes. Few countries have enacted data protection law and related guidelines. Many other countries have started drafting their own digital data privacy guidelines and law. Also there are

IJERT

many standard bodies and expert groups who have guided industries to adopt to data security guidelines and controls. In Indian context there is not any digital data security law as yet. However National Cyber Security Policy 2013 does talk of adopting data security standard and best practices on framework like ISO-27001 ISMS [4].

II. RISE AND HARNESSING OF DIFFERENT DATA CLASSES

Conventionally data-store is synonymous to RDBMS in the parlance of information system. A sophisticated schema based RDBMS engine used to sit on mainframe or high end database server. The wide scale success and adaptability of client server based business processing system reinforced the ruggedness and usefulness of RDBMS systems. Later, overwhelming success of web based 3 tier application systems on the same RDBMS infrastructure towards the end of last century further vindicated industry's confidence on RDBMS technology. However with larger automation initiatives, industry gradually started to adopt diverse data types which are simply not limited to RDBMS tenable structured relation based datasets. Social media has been fuelling the industry to adopt and use diverse data types for all possible value addition. Today we are required to store structured data like table, semi structured data like XML, JSON objects, unstructured data like email or conversation, multimedia data like audio or video in a seamless manner in the same data store. We can simply take up the example of a blogging web application. In a typical blog we can expect content comprising of text, table, audio, video and a thread of user discussions. A close look will readily reveal that traditional RDBMS is simply not capable of seamlessly integrating these diverse datasets without painful programming and tweaking. A new breed of database namely NoSQL (Not only SQL) database can handle such data classes with relative ease and efficacy [9]. RDBMS banks on consistency of data as per ACID principle whereas NoSQL databases are based on BASE (Basically Available Soft-state Eventually-consistent) principle where mostly we compromise consistency as the cost of database availability. A glaring testimony of NoSQL database success story is Facebook where clusters of Hbase NoSQL databases are used to store millions of posts, images videos etc. Further these abysmal datasets are analyzed in real time to generate instant deduction and knowledge [10]. Nevertheless, NoSQL is not a silver bullet for all kinds of applications. RDBMS still reigns for OLTP kinds of applications.

NoSQL database has given due recognition to all types of data classes. In fact much touted Big Data revolution is primarily riding on the success of NoSQL database technology. Wide scale adoptability of different data classes comprising text, mail, xml, json, image, sound, video etc. has simply increased the threat vulnerabilities of digital data.

III. DATA SECURITY VULNERABILITIES

Data security understandably has been gaining more significance owing to the advancement of Big Data and other advanced technology to handle diverse classes of data. Traditionally data security emphasized on securing structured data found in a RDBMS. Needless to say, data security philosophy is essentially same for other data classes as well based on the Confidentiality, Integrity and Availability tenets. Before highlighting the data security aspects, let us try to understand data lifecycle and state of data in a typical information processing setup.

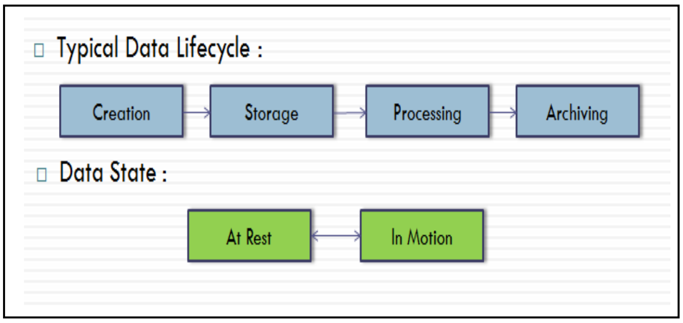

Fig 1. Data Lifecycle and State

Matured RDBMSes do give few security mechanisms like authorization, authentication, auditing, logging, encryption etc. with regard to data storage and archiving. However those mechanisms fall short of providing a complete and comprehensive data security controls as per the framework shown above. NoSQL databases are on its path of maturity and as such security has taken back seat [16]. A brief description of each possible generic data security weak link is discussed below [7][11][12]. These vulnerabilities are equally true for all kinds of data store. Later we will discuss the control mechanism to mitigate those security risks and associated regulatory requirements.

A. *Authentication*

Authentication is the process by which a data store verifies the credentials of a legitimate user. Unlike matured RDBMS like Oracle/SQL Server NoSQL database like MongoDB does not enable authentication by default [16]. That way it exposes a security vulnerability. For remote authentication if the authentication channel is not secure or protected across the network then there is a chance of user credentials getting leaked.

B. *Authorization*

Authorization refers to the privileged access of specific data object like table, image etc. depending on the access role of the specific user. For example an user handling the inventory module of a database should not see the employee salary of his company leading to confidentiality breach. Similarly much talked SQL injection attack is also a result of authorization breach.

C. *Data Integrity*

Data should not be changed either in motion or at rest. In case of data tampering , resulting from data corruption or any intrusion the same should be known to the owner of data or intended recipient.

D. *Plain Text or Original Data*

Data if allowed to remain in its plain or in original standard format may be exploited by known or unknown element with or without malicious intent. This is equally important for data at rest and for data in motion. For example,

IJERT

one's credit card data should not be kept as plain text in any database table or on any data store like XML or JSON object. Similarly a scientific drawing or a highly technical study report if left on its original format may be exploited leading to confidentiality breach.

E. Auditing, Logging and Monitoring

Detailed auditing of database/data store activities is an important aspect for detective data security. Imagine a scenario where a DBA with malicious intent accesses personal data. With proper procedures, technology and audit trail monitoring we can detect such breaches as can take corrective actions.

F. Denial of Service

If our application fails to deliver us our intended data or information for reasons which are not its core activities then it is said to be a denial of service attack. For example if a database listener resource is overwhelmed with abnormally high database connection request then its performance while servicing the existing connection will definitely degrade or stop leading to unavailability of service.

G. Business Continuity

Business continuity basically deals with the availability tenet of information security aspect. If proper data and information availability arrangements are not ensured for an emergency or disaster scenario then any organization's hard earned data and information may get wiped for any mishap , accident or disaster. Implementation of proper backup policy, restoration and recovery drill, offsite data archiving, high availability secondary site etc. are important measures for ensuring adequate business continuity.

IJERT

IV. SECURITY CONTROLS AND REGULATORY GUIDELINES

Data Security control refers to mechanism, technology or procedures deployed to ensure data security. There are different controls to mitigate different security risks. For example encryption could be a technical security control to safeguard plain text data. Similarly mandatory change of password after 2 weeks is a good example of procedural data security control.

Regulator refers an official or a government/statutory agency responsible for control and supervision of a particular activity or area of public interest. There are different regulators for various domain specific industries. For example we see telecom regulator, oil & gas regulator, health regulator etc. Data however are not restricted to any specific industry. Primarily domain specific regulator has laid down standards for data security for data of special interest to that domain. For example The Payment Card Industry Data Security Standard (PCI DSS) is a proprietary information security standard for organizations that handle cardholder information for the major debit, credit, prepaid, e-purse, ATM, and POS cards which is defined by the Payment Card Industry Security Standards Council. Similarly US health industry needs to adhere to HIPAA (The Health Insurance Portability and Accountability Act) guidelines and standard while dealing with patient data in US. Infamous Sarbanes-Oxley (SOX) Act of 2002 in the U.S. is a piece of legislation created for the purpose of protecting investors from accounting fraud which talks of using IT security controls to great extent. Data governance and regulatory aspects have reached a fair and matured level in advanced countries like US and Canada. Nevertheless, other countries and regions are also pursuing their data governance and security standard. Below we have summarized and tabulated different industry adopted security controls in light of the above mentioned data security aspects as required by respective known regulatory agency.

TABLE 1 : Controls and Regulations

Security Requirement	**Technological Control**	**Regulation**
Authentication	LDAP, Kerberos based access control	PCI-DSS, Token based authentication
Authorization	OS, Data Store (RDBMS/NoSQL /Other DB) defined mechanism	SOX
Confidential and secure Internal Data Communication	SSH, VPN, PGP, SSL/TLS	SOX
Confidential and secure External Data Communication	SSH, VPN, PGP, SSL/TLS	PCI-DSS
Confidentiality of Files	AES or similar algorithm based file encryption	PCI-DSS
Sensitive and personal data de- identification or anonymization	Pseudonymize, encode, hash, randomize, tokenize, masking, redaction , format preserving encryption	HIPAA, PCI-DSS
Logging monitoring, alerting	Log management, IDS/IPS, Database Activity Monitoring	PCI-DSS, SOX
Prevent Denial of Service	Web application firewall	PCI-DSS
Business Continuity	Disaster recovery set up, Business Continuity Plan	PCI-DSS

V. CONCLUSION AND FUTURE WORK

No security mechanism is completely full proof. Data security controls are also no different. However the best we can do is to strengthen all possible proactive control mechanisms. Studies by many industry bodies [14] show that internal threats are far more dangerous than threat from outside organization. Mandatory auditing and exhausting logging mechanisms for database activity monitoring can play a vital role to mitigate internal threat. For example how can we safeguard someone's credit card number or salary information from the DBA of that data store? Similarly how can we proactively detect and prevent a novel SQL injection attack? There are Database Activity Monitoring (DAM) products [15] available to implement such safeguard for matured RDBMS like Oracle. However, not much work has been done towards achieving such proactive data protection for NoSQL or such other non RDBMS databases.

These DAM applications primarily work on two principles [17]. The first one is signature based protection where an exhaustive signature database is maintained and updated for legitimate operations or activities on the database. Any deviation from the given signature is detected from the audit logs or from the database instance. Subsequently, appropriate alert mechanisms and protective measures are triggered. The other principle is 'anomaly detection', where behavioral pattern of database operations in the database is gradually built based on the allowed activities on the database. Data analytics principles [18] can be used for generating the allowable patterns. In recent time, NoSQL based databases have started putting sensitive structured and unstructured data in it. A lot of works can be done in devising appropriate DAM based security controls for these kinds of databases.

REFERENCES

Regulatory Guidelines and Resource from Standard Bodies:

[1] Risk Management Guide for Information Technology Systems, NIST, US Deptt. Of Commerce

[2] ISO/IEC 27000:2009

[3] ISACA, 2008, *www.isaca.org*

[4] National Cyber Security Policy 2013. http://deity.gov.in/content/national-cyber-security-policy-2013-1

[5] PCI-DSS, HIPAA, SOX guidelines.
https://www.pcisecuritystandards.org/security_standards/
http://www.hhs.gov/ocr/privacy/
www.soxlaw.com

Journal Papers

[6] Security Issues in NoSQL Databases, Lior Okman, Nurit Gal-Oz, Yaron Gonen, Ehud Gudes, Jenny Abramov *2011 International Joint Conference of IEEE TrustCom-11/IEEE ICESS-11/FCST-11.*

[7] Database Security: A Historical Perspective, *University of Minnesota CS 8701, Fall 2008*

[8] Survey on Data Mining Techniques to Enhance Intrusion Detection, Deepthy K Denatious & Anita John *2012 International Conference on Computer Communication and Informatics (ICCCI -2012), Jan. 10 – 12, 2012, Coimbatore, INDIA, 978-1-4577-1583-9/ 12/ © 2012 IEEE*

Web Resource:

[9] Introduction to NoSQL *w3resource.com*

[10] Storage Infrastructure Behind Facebook Messages, Kannan Muthukkaruppan, Software Engineer, facebook.com *Big Data Experiences & Scars, HPTS 2011*

[11] Understanding Holistic Database Security, Whitepaper, IBM.COM

[12] Oracle Security Solutions, Oracle.com

[13] www.iri.com

[14] Symantac Internet Security Threat Report 2013, Symantec.com

[15] McAfee Real-Time Database Monitoring, Auditing, and Intrusion Prevention

Books:

[16] Mongodb Documentation. MongoDB Documentation Project *mongodb.com*

Industry White Paper:

[17] The Need for Real-Time Database Monitoring, Auditing and Intrusion Prevention. *Analytics.InformationWeek.com*

[18] Securing Unstructured Data, *Analytics.InformationWeek.com*

Conference Proceeds:

[19] Data Mining for Intrusion Detection, Department of Computer Science University of Minnesota *Tutorial on the Pacific-Asia Conference on Knowledge Discovery in Databases 2003*

IJERT