# Control Mechanisms for Robust Data Security

Chandan Kumar Barman[#1], Pankaj Gupta[*2]

[#] *Superintending Engineer(IT), DGH, Ministry of Petroleum and Natural Gas, India*
** Assistant Professor, Birla Institute of Technology, Mesra, Noida Center, UP, India*

***Abstract***— Data undoubtedly is at the core of IT value chain in any organization. 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. In this paper we have defined different facets of data security vulnerabilities that are common to any data-store or data aware application. Later. we have defined and highlighted various control mechanisms required to be put in place to mitigate these data security vulnerabilities.

The three controls namely the procedural control, technical control and physical control as discussed below may be referred and deployed by any organization for robust data security.

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

## 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 safeguards. 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].

## II. RISE AND HARNESSING OF DIVERSE 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.

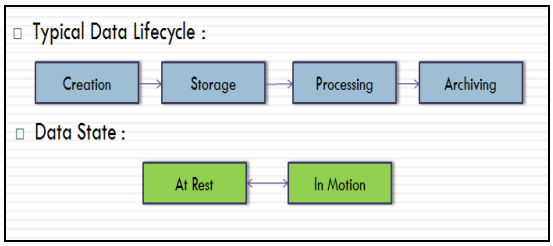



FIG 1: DATA LIFECYCLE AND STATE

As we can see, data are created, stored, retrieved/processed and eventually archived. Data may be at rest or in motion. For a full proof data security framework we have to ensure that the CIA principles are conformed at each phase of data lifecycle in either of the data 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 in details the practiced control mechanisms in the industry to mitigate those security risks.

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, 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.

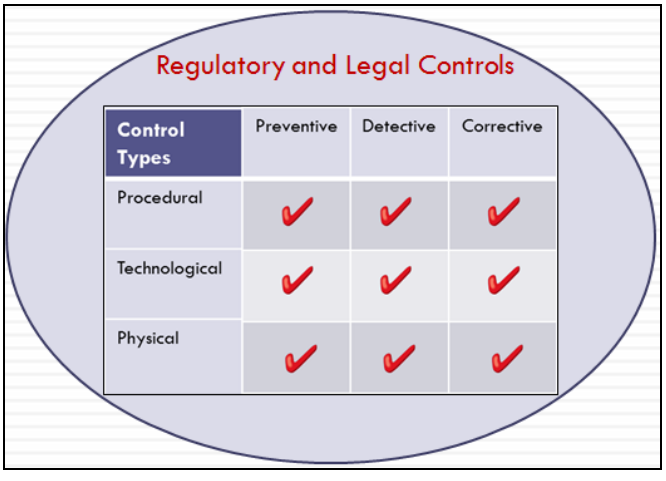
## IV. DIFERENT SECURITY CONTROLS

Data Security control refers to mechanism, technology and procedures deployed to ensure data protection and 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.

We find different criteria based on which security controls are defined. One is time based controls. Where controls are defined and deployed on the time scale prior and subsequent to data vulnerability exploitation. They are namely i) Preventive controls, that are deployed before the incident ii) Detective controls are intended to identify and characterize an incident in progress iii) Corrective controls are intended to limit the extent of any damage caused by the incident e.g. by restoring and recovering the organization to normal working status as efficiently as possible after a disaster (fire incident).

Another categorization of security controls is based on the nature of the controls. They are i) Physical Control like access control, CCTV monitoring etc. ii) Procedural Control like awareness, training, routine monitoring of data center etc. iii) Technical Control or logical controls that refer to all software and associated technology aided tools and mechanisms to impose data security. Finally we can talk of regulatory controls. Regulatory and compliance controls refer to the safety and security mechanism to be followed by data custodian as prescribed by statutory, legal or regulatory bodies.

For a prudent data security and governance framework we should develop and deploy appropriate combinations of control mechanisms based on time and nature of the control as depicted on the matrix below. Further the data custodian or data governance mechanisms should ensure that the control mechanisms being deployed conform to the legal and regulatory requirements.




The following table highlights few of the technological controls that are available to mitigate data security risks as discussed above.

TABLE I

| Technological Control |
|---|
| LDAP, Kerberos based access control for authentication requirements |
| OS, Data Store (RDBMS/NoSQL /Other DB) defined mechanism for authorization |
| SSH, VPN, PGP, SSL/TLS for security and confidentiality of data in motion. |
| AES or similar algorithm based file encryption for data at rest |
| Pseudonymize, encode, hash, randomize, tokenize, masking, redaction , format preserving encryption for sensitive and personal data de- identification or anonymization |
| Log management, IDS/IPS, Database Activity Monitoring for activity logging, monitoring and alerting |
| Web application firewall for safety against denial of service |
| Disaster recovery set up, Business Continuity Plan for system availability. |

Physical controls that are primarily required at the perimeter and at the data centers are another type of important control we should not ignore. Below is a list of typical physical controls that anybody can emulate in his IT infrastructure.

TABLE III

| Physical Controls |
|---|
| Data center location with minimal natural and man-made disaster risks. |
| Separate electricity from different power houses for maximum infrastructure uptime |
| Automatic role based access control at each data center access point |
| Exhaustive CCTV coverage both outside and inside data center |
| Optimum temperature and relative humidity backed by redundant precision cooling. |
| Firefighting arrangement with adequate extinguishers and by total flooding agents. |
| Adequate UPS power with diesel generator backed power. |
| Disaster recovery center at a different seismic zone with suitable disaster recovery plan. |
| Well-equipped Network Operations Center to monitor data center activities. |

Procedural and administrative controls primarily tell how security policy, standards, and guidelines will actually be implemented in an operating environment of an organization. Procedural and administrative controls are guidelines or agreements that mandate IT users to act in certain ways with the goal of protecting information assets. Procedural and administrative controls are guided by laws and local regulations, organizational policies, guidelines, goals, methods of protecting intellectual property such as copyrights, patents, or trade secrets, and contracts or other documents that govern the relationship between two or more parties. Procedural controls are broadly specific to any organization. Nevertheless they will inherit many common characteristics. For example sensitive and personal data de- identification or anonymization is a HIPAA regulatory requirement while handling personal data in the US. Below is a list of few industry accepted bodies and societies whose guidelines, frameworks, recommended best practices may be followed and adopted for prudent procedural controls.

*A. Control Objectives for Information and Related Technology (COBIT):* This is a framework created by ISACA for information technology (IT) management and IT governance.

*B. ISO/IEC 2700:* This is a set of standards published by the International Organization for Standardization (ISO) and by the International Electrotechnical Commission (IEC). It provides best practice recommendations on information security management for use by those responsible for initiating, implementing or maintaining information security management systems (ISMS).

C. *The Information Technology Infrastructure Library (ITIL):* This is a set of concepts and techniques for managing information technology (IT) infrastructure, development, and operations.

D. *OCTAVE (Operationally Critical Threat, Asset, and Vulnerability Evaluation):* This is another suite of tools, techniques, and methods for risk-based information security strategic assessment and planning. The OCTAVE method is an approach used to assess an organization's information security needs. OCTAVE Allegro is the most recently developed method.

E. *Committee of Sponsoring Organizations of the Treadway Commission (COSO):* This is a U.S. private-sector initiative, formed in 1985. Its major objective is to identify the factors that cause fraudulent financial reporting and to make recommendations to reduce its incidence. COSO has established a common definition of internal controls, standards, and criteria against which companies and organizations can assess their control systems.

## V. Conclusions

In this paper we have highlighted the growing importance of diverse data classes like image, text, sound, json, e-mails etc. in addition to normal tabular structured data found in RDBMSes. Data is at the core of IT assets value chain in any organization. As such data need to be protected by most effective and prudent way. We have discussed the various control mechanism adopted by the industry for a prudent data governance regime. The reference security frameworks devised by various bodies as mentioned

above may be referred while devising organization specific data security guidelines. All the security controls discussed here may serve as broad reference guide while considering various data safeguarding alternatives.

## 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*