*International Journal of Advanced Engineering Research and Science (IJAERS)* *Special Issue (NCCSIGMA-16)*
*https://dx.doi.org/10.22161/ijaers/si.9* *ISSN: 2349-6495(P) | 2456-1908(O)*

# A Comprehensive Approach for Data Protection

Jaya Saxena[1], V Subramanian[2]

[1]PPEG, National Remote Sensing Center, Balanagar, Hyderabad, India
Email: jayasaxena@nrsc.gov.in,
[2]PPEG, National Remote Sensing Center, Balanagar, Hyderabad, India
Email: subramanian_v@nrsc.gov.in

***Abstract—*** *It has been predicted that analyzing Big Data will become a key basis of competition, underpinning new waves of productivity growth, innovation, and consumer surplus. Since data is a critical component, it is essential to ensure privacy and protect data no matter where it resides and how it is consumed. Organizations are now defending for attacks occurred yesterday, however their adversaries look to exploit the vulnerabilities of tomorrow. Consequently, sophisticated intruders are bypassing perimeter defenses to perpetrate dynamic attacks that are highly targeted and difficult to detect. In this work we present a strategy focusing the few important questions to the most critical data vulnerabilities.*

***Keywords—Data protection, data access policy, masking, threats and risks, security***.

## I. INTRODUCTION

Today, in the form of Big data, daily more than 2.5 quintillion bytes of data are created from digital pictures, videos, posts to social media sites, intelligent sensors, purchase transaction records, cell phone GPS signals to name a few. There is a great interest both in commercial and in research communities around Big Data. It has been predicted that analyzing Big Data will become a key basis of competition, underpinning new waves of productivity-growth, innovation, and consumer surplus. Big data is not just the size of an individual data set, but rather the collection of data that is available to us online (e.g., government data, NGOs, local governments, journalists, etc). By putting these data together makes the big data a value to the wider public. Data activists and engineers world-wide easily discover data sets, merge them in a sensible fashion for a meaning full outcome. As an example, helping people in crisis response situations has huge potential when people have used Google Fusion Table to create maps with critical information for people after the Japan earthquake in 2011[1]. When it comes to natural resources, we are leveraging big data to optimize the placement of turbines in a wind farm so that we get the most power with the least environmental impact. We can also look at man-made phenomena for example, understanding traffic patterns and using the insight to do better planning or provide incentives that can reduce traffic during peak hours. Many other examples can be given of how Big Data is being used to improve the planet.

Since data is a critical component it is essential to ensure privacy and protect data no matter where it resides and how it is consumed. Different types of information have different protection requirements; therefore, organizations must take a holistic approach to safeguard. Strategies, typically compliance-based and perimeter-oriented, have not kept pace with the sophisticated approaches intruders are evolving. Most organizations are now defending for attacks occurred yesterday, however their adversaries look to exploit the vulnerabilities of tomorrow.

Consequently, sophisticated intruders are bypassing perimeter defenses to perpetrate dynamic attacks that are highly targeted and difficult to detect, STUXNET[2] is a good example. Many use well-researched phishing exploits targeted groups. Similarly, the attack surface includes partners, suppliers, customers, and others, has expanded with greater volume of data flows and through multiple channels. A study in 2012 conducted by Symantec, calculated that global cybercrime cost USD114 billion annually and claimed more than one million victims per day [3]. In an another study conducted by the Ponemon Institute, the average organizational cost of a data breach in 2011 was USD5.5 million [4], Target, Sony data breaches are few know examples.

Further, safeguarding all data at an equally high level is no longer practical as new attack vectors including cyber security threats (worms, trojans, rootkits, rogues and spyware) and security complexities resulting from changing IT architectures challenge organizations to focus on data protection and requires more granularities. In this work, we present a strategy focusing the following few questions to the most critical data vulnerabilities:

1) Where does the classified and sensitive data reside across the enterprise?
2) How access to the enterprise databases can be protected, monitored and audited?

*https://dx.doi.org/10.22161/ijaers/si.9* *ISSN: 2349-6495(P) | 2456-1908(O)*

3) How data can be controlled from both authorized and unauthorized access?
4) How data in non-production environments can be protected, and still be usable for training, application development and testing?
5) What types of data encryption are appropriate for data at rest and in transit?

ORGANIZATION OF THE PAPER

The paper is organized as follows: Section II, contains Threats and Risks associated to the IT infrastructure. In Section III we present our approach to data protection, followed by concluding remarks in Section IV and references in the last Section.

## II. THREATS AND RISKS

As per Wiki[7], in IT area a threat is a possible danger that might exploit a vulnerability to breach security and thus cause possible harm. A threat can be either "intentional" (i.e., intelligent; e.g., an individual hacker or a criminal organization) or "accidental" (e.g., the possibility of a computer malfunctioning, or the possibility of a natural disaster such as an earthquake, a fire, or a tornado) or otherwise a circumstance, capability, action, or event. However, the nature of computer crime has changed over the years as the technology has changed and the opportunities for crime have increased by multiple folds. Although thrill-seeking adolescent hackers are still common, the field is increasingly dominated by professionals who steal information for sale and disgruntled employees who damage systems or steal information for revenge or profit. Survey after survey has shown that most damage is done by insiders, people with authorized access to a computer network. Many insiders have the access and knowledge to compromise or shut down entire systems and networks [8].

**Insider threats**

A high percentage of data breaches actually emanate from internal weaknesses. These breaches range from employees who may misuse payment card numbers and other sensitive information to those who save confidential data on laptops that are subsequently stolen. Furthermore, organizations are also accountable for protecting data no matter where the data resides, be it with business partners, consultants, contractors, vendors or other third parties. Few common sources of risk include:

- Excessive privileges and privileged user abuse. When users (or applications) are granted database privileges that exceed the requirements of their job function, these privileges may be used to gain access to confidential information.
- Unauthorized privilege elevation. Attackers may take advantage of vulnerabilities in database management software to convert low-level access privileges to high-level access privileges.
- SQL injection. SQL injection attacks involve a user who takes advantage of vulnerabilities in front-end web applications and stored procedures to send unauthorized database queries, often with elevated privileges. Using SQL injection, attackers could even gain unrestricted access to an entire database.
- Denial of service. Denial of service (DoS) may be invoked through many techniques. Common DoS techniques include buffer overflows, data corruption, and network flooding and resource consumption. The latter is unique to the database environment and frequently overlooked.
- Exposure of backup data. Some recent high-profile attacks have involved theft of database backup tapes and hard disks which were not encrypted.

According to one of the leading computer security body the OWASP top 10 threats are:

Threat #1: virus
Threat #2: spam
Threat #3: spoofing, phishing
Threat #4: spyware
Threat #5: keystroke logging (keylogging)
Threat #6: adware
Threat #7: botnet
Threat #8: worm
Threat #9: trojan horse
Threat #10: denial-of-service attack (dos attack)

These can impact performance by slowing down the computer or completely block thus can create instability by an active conduit for download and installation. It may also compromise the privacy by release of confidential, protected, or sensitive information, release of browser-tracking information, logged keystrokes, or other forms of data. Infections can allow programs to spread to other computers, mobile devices, or network file shares. These can lead to data loss, corruption, or other forms of operational impairment to infected hosts. In additional to legal issues surrounding violations of privacy laws, owners of infected hosts might find themselves liable for harm or loss caused by infected computer. Spam relay programs can be implemented, allowing the creator to hide the origin of spam messages. Trojan horse programs can replace common applications on the host computer, creating vulnerabilities and softening the host's defenses. Trojan horse programs can also be used to coordinate mass network scanning or network attack efforts, making it harder to detect the profiling scan or attack coming from tens of thousands of separate computers controlled by the creator of the program.

*International Journal of Advanced Engineering Research and Science (IJAERS)* *Special Issue (NCCSIGMA-16)*
*https://dx.doi.org/10.22161/ijaers/si.9* *ISSN: 2349-6495(P) | 2456-1908(O)*

## III. OUR APPROACH FOR DATA PROTECTION

Organizations may find it difficult to identify to protect sensitive data unless they know where it resides and how it is related across. Organizations need to define and document all data assets and relationships, no matter what is the source. It is important exercise to classify data, understand relationships and define service levels. The data discovery process analyzes data values and patterns to identify the relationships that link disparate data elements into logical units of information. Key issue is to safeguard sensitive data, both structured and unstructured. Structured data contained in databases must be protected from unauthorized access. Unstructured data in documents and forms requires privacy policies to remove sensitive information while allowing required business data transaction. Data in nonproduction, development, training and quality-assurance environments needs to be protected however, still usable during application development, testing and training processes.

- Structured data: This data is based on a data model, and is available in structured formats like databases or XML.
- Unstructured data: This data is in forms or documents which may be handwritten, typed or in file repositories, such as word processing documents, email messages, pictures, digital audio, video, GPS data and more.
- Online data: This is data used daily to support the business, including metadata, configuration data or log files.
- Offline data: This is data in backup tapes or on storage devices.

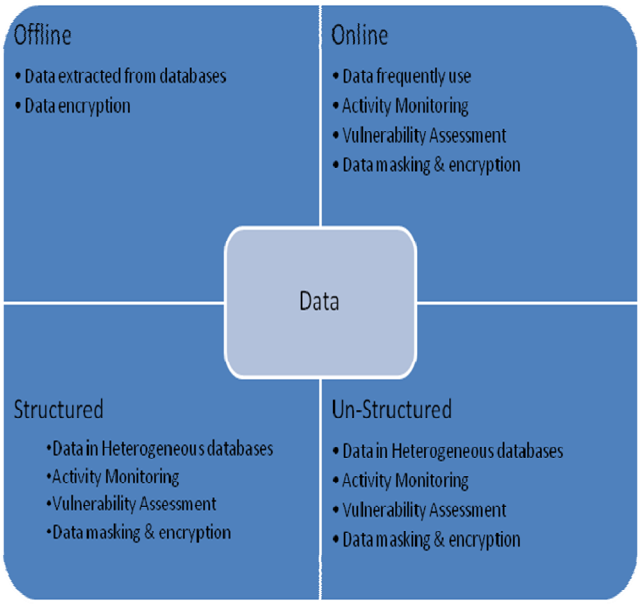


*Fig. 1. Data Classification*

The ability to understand which requirements are applicable is particularly important because data security is not a single process or single technology. Instead it is addressed through a defense in-depth approach using a variety of technologies and controls appropriate for specific situations and a complex legal and threat landscape as no single best practice exists.

Organizations must understand the particulars of their unique use cases to enable better decision making, more cost effective investments and more successful technology implementations. This is especially challenging because of continuing uncertainty and constrained resources. Organizations also struggle to satisfy data security requirements as data is integrated and leveraged across a dynamic supply chain of information. The data gets created, accessed, used and retired. As data moves through each phase, its value and the associated risks change so security protection requirements need to be managed appropriately.

Investment in data security controls should be evaluated within the context of the business goals, compliance mandates and risk levels. This requires a clear understanding of how the data is used and what compliance mandates are important to that particular transaction as it can be different depending on what that system does, who has access to it, what data is stored there and how it is used in the end. Once sensitive data is discovered, its vulnerability must be assessed. This assessment may include several things from checking admin-level access privileges to verifying the current configurations of databases are compliance with any known vulnerabilities.

De-identifying data in non-production environments is simply the process of systematically removing, masking or transforming data elements that could be used to identify an individual. Data de-identification enables developers, testers and trainers to use realistic data and produce valid results while protecting sensitive data. This is especially important for organizations outsourcing development or testing activities. Dynamic data masking prevents unauthorized users from accessing structured data in real time. Organizations can apply sophisticated, flexible data masking rules based on business rules and requirements. Dynamic data masking policies mask sensitive information in transit after it is fetched from the database and masked results are returned to the requesting web application. Organizations with call centre setup of operations typically use dynamic data masking to hide customer information from call center employees. There are many different types of encryption available. Organizations should consider a file level encryption approach because it provides a single, manageable and

*https://dx.doi.org/10.22161/ijaers/si.9* *ISSN: 2349-6495(P) | 2456-1908(O)*

scalable solution to encrypt enterprise data both structured and unstructured without sacrificing application performance or creating key management complexity. Data encryption is ideal for protecting online and offline data. Unfortunately, it is not enough to understand sensitive data and establish the right kinds of policies to protect the data. Organizations also have to continuously monitor data sources for suspicious behavior. Organizations should not rely on manual auditing procedures to detect suspicious behaviors. This approach prolongs audits and is resource intensive.

In most IT environments, privileged users such as DBAs, developers and outsourced personnel can have unfettered access to sensitive data, with little or no monitoring controls around their activities. These super users can easily bypass application or network-level security control points. In addition, detecting database changes is important from the perspective of placing controls around privileged users. These changes can be indicators that a database has been compromised. However, it is also important from an external security perspective.

Create fine-grained audit trails and reporting to prove and validate compliance, organizations must have a defined process for monitoring, recording and reporting database access and change activity on a periodic basis. A fine-grained audit trail identifies the "who, what, when, where, and how" of each transaction. Through continuous monitoring and reporting, detection of data access violations gives IT management and auditors the necessary information to show that the proper controls are in place and are being enforced. Audit trails provide details and analysis of behaviors and patterns that may be deemed suspicious versus legitimate or routine. Any behavior that is not identified as routine and valid access to the database must be examined and analyzed further. Building a centralized audit and reporting environment enables:

- A secure centralized repository containing a fine-grained audit trail of all database activities across the enterprise, as well as important file sharing activities.
- Customizable workflow automation to generate compliance reports on a scheduled basis, distribute them to oversight teams for electronic sign-offs, escalation and store the results of remediation activities in the repository.
- Continuous monitoring and analyzing of data to identify unauthorized or suspicious activities and execute a response action ranging from blocking the transaction in real time to generate an alert.

Activity monitoring provides privileged and non-privileged user related information. Application access monitoring that is independent of native database logging and audit functions are also very useful. It can function as a compensating control for privileged user by monitoring all administrator activity. Activity monitoring also improves security by detecting unusual access to database, data warehouse, file share and update activities from the application layer. Activity monitoring solutions should be able to detect malicious activity or inappropriate or unapproved privileged user access.

Data redaction can remove sensitive data from forms and documents based on job role or business purpose. Traditionally, protecting unstructured information in forms, documents and graphics has been performed manually by deleting electronic content and using a black marking pen on paper to delete or hide sensitive information. But this manual process can introduce errors, inadvertently omit information and leave behind hidden information within files that exposes sensitive data. Today's high volumes of electronic forms and documents make this manual process too burdensome for practical purposes, and increase an organization's risk of exposure. For example, physicians need to see sensitive information such as symptoms and medical data, whereas a billing clerk needs the patient's insurance data and billing address. The challenge is to provide the appropriate protection, while meeting business needs and ensuring that data is managed on a "need-to-know" basis.

Despite data encryption is an old technology, and many different approaches exist, encryption is explicitly required by many regulations across globe. It is challenging for an organization to identify the best encryption approach due to various prolific offerings. For encrypting structured data, consider a file-level approach. This will protect both structured data in the database management system (DBMS) and also unstructured files such as DBMS log or configuration files, and is transparent to the network, storage and applications. Explore for encryption offerings which provide a strong separation of duties, a unified policy and key management system to centralize and simplify data security management.

A security solution which addresses the entire database security and compliance life cycle back-end data store with workflow automation system should consist of vulnerabilities repository and configuration flaws list. It should ensure that configurations are locked down after recommended changes are implemented while providing 100% visibility and granularity into all data source transactions across all platforms and protocols with a secure, tamper-proof audit trail that supports separation of duties. Monitor and enforce policies for sensitive data access, privileged user actions, change control, application user activities and security exceptions. Create

*https://dx.doi.org/10.22161/ijaers/si.9* *ISSN: 2349-6495(P) | 2456-1908(O)*

a single, centralized audit repository for enterprise wide compliance reporting, performance optimization, investigations and forensics.

**Masking Solution**

A comprehensive set of data masking techniques that can support data privacy compliance part is required The masking capabilities will ensure that masked data, like names and street addresses, resembles the look and feel of the original information.

- Context-aware, prepackaged data masking routines make it easy to de-identify elements such as payment card numbers, street addresses and email addresses.
- Persistent masking capabilities propagate masked replacement values consistently across applications, databases, operating systems and hardware platforms.
- Static or dynamic data masking supports both production and non-production environments.

The solution should consist of a single, manageable and scalable solution to encrypt enterprise data without sacrificing application performance or creating key management complexity. it should consistent, transparent encryption method across complex enterprises which can be auditable with almost no or minimal application, database or system changes while having secure and centralized key management across distributed environments. The solution may be intelligent, easy-to-customize data security policies for strong, persistent data security with strong separation of duties.

After data has been located and locked down, organizations must prove compliance, and monitor systems on an ongoing basis. Monitoring of user activities, object creation, data repository configurations and entitlements help IT professionals and auditors trace users between applications and databases. These teams can set fine-grained policies for appropriate behavior and receive alerts if these policies are violated. Organizations need to quickly show compliance and empower auditors to verify compliance status. Audit reporting and sign-offs help facilitate the compliance process while keeping costs low and minimizing technical and business disruptions. Thus organizations should create continuous, fine-grained audit trails of all database activities.

## IV. CONCLUSION

We may like to give our conclude remarks that in order to be most effective, information security must be integrated into the SDLC from system inception. Early integration of security in the SDLC enables agencies to maximize return on investment in their security programs. It is observed that early identification and mitigation of security vulnerabilities and misconfigurations, results in lower cost of security control implementation and vulnerability mitigation.

Further identification of shared security services and reuse of security strategies while improving security posture through proven methods and techniques will facilitate for decision making through comprehensive risk management in a timely manner. In this work we attempted to answers few of the questions which provide the foundation for a holistic approach to data protection and scales as organizations embrace the new era of computing. The answers also help organizations focus in on key areas they may be neglecting with current approaches.

## REFERENCES

[1] "A Decade at Google", Alon Halevy, ACM SIGMOD, September 20, 2015 http://wp.sigmod.org/?p=1851

[2] "The Real Story of Stuxnet", David Kushner, IEEE SPECTRUM. Feb 2013, http://spectrum.ieee.org/telecom/security/the-real-story-of-stuxnet

[3] “Norton Study Calculates Cost of Global Cybercrime: $114 Billion Annually,” Sept. 7, 2011, www.symantec.com/about/news/release/article.jsp?prid=20110907_02

[4] “2011 Cost of Data Breach Study,” Ponemon Institute LLC, March 2012, www.symantec.com/content/en/us/about/media/pdfs/b-ponemon-2011-cost-of-data-breach-us.enus.pdf?om_ext_cid=biz_socmed_twitter_facebook_marketwire_linkedin_2012Mar_worldwide__CODB_US

[5] "Target cyber breach hits 40 million payment cards at holiday peak", By Jim Finkle and Dhanya Skariachan, Technology News,Dec 19, 2013 , http://www.reuters.com/article/us-target-breach-idUSBRE9BH1GX20131219

[6] "PlayStation Network hackers access data of 77 million users" Ben Quinn and Charles Arthur, The Guardian, 26 April 2011 https://www.theguardian.com/technology/2011/apr/26/playstation-network-hackers-data

[7] https://en.wikipedia.org/wiki/Threat_(computer)

[8] "A Survey of Insider Attack Detection Research", Malek Ben Salem, Shlomo Hershkop Salvatore J. Stolfo http://lasr.cs.ucla.edu/classes/239_1.fall10/papers/Insider_survey.pdf