informatica

White Paper

Data Security Controls

Review and Use Case

This edition published May 2015

Table of Contents

Executive Summary

As data breaches continue to plague private and public organizations, security teams look to data security controls to prevent both outside intruders and malicious insiders from accessing sensitive, private, or mission-critical data in the organization's databases.

As with traditional cybersecurity controls (firewalls, VPNs, intrusion detection), data security controls come in various shapes and forms, including encryption, tokenization, masking, and archiving. Even seasoned security veterans can be confused about how to use each one most effectively, but each has a place in protecting valuable information assets. They work best together in an integrated protection scheme that will thwart unauthorized access of data from both external and internal threats. Instead of taking a technology-centric approach, this paper explores use cases for these data-centric security controls in simple business terms.

Use Cases for Data Security Controls

What are the common use cases for data security controls? The goal here is not an exhaustive list of data security control use cases, but the review of common use cases that every organization should consider as part of its data protection strategy and tactics. These use cases fall into three broad categories: data at rest, data in motion, and data in use.

1. Data at rest – stored on disk or removable media.
2. Data in motion – transmitted across internal or external networks
3. Data in use – used by applications in testing or shared for outsourcing

- Application data in use is used by a line of business, business intelligence, or mission-critical application and is presented on screen to a user.
- Test data in use is used in application test environments.
- Shared data in use is used in the following ways:
  - internally – for centralized analytics or reporting
  - externally – for outsourcing or industry research
  - with industry or government analytic/research projects

Review of Data Security Controls Solutions

Generally, data security controls focus on protecting the data itself, as opposed to protecting access to a network or application. Persistent breaches and increasingly rigorous privacy laws require organizations to improve their protection of sensitive and private data. Outside breaches and insider threats, either malicious or accidental, circumvent cyber security; data security controls become the last barrier to prevent unauthorized access and disclosure.

Each technology that provides data security has its strengths and challenges and is at a different level of adoption and maturity. Encryption, for example, is widely accepted, while masking and tokenization are emerging with strong endorsement from security analysts and market leaders for critical use cases.

	KEY STRENGTHS	CHALLENGES	NOTABLE
ENCRYPTION	Well suited to protecting stored or transmitted data (data at rest and data in motion).	Management of encryption keys, performance, deployment, and implementation.	Encryption is recommended as a baseline for data privacy and security in many frameworks and regulations.
TOKENIZATION	Sensitive data, such as a credit card numbers, are removed and replaced with a token, typically of the last four digits of the number. The original data is stored in a highly secure data server/vault.	Requires modifications to applications that use sensitive data, does not scale to support enterprise data security/privacy, and does not support environments that need ongoing access to original data.	Tokenization has become a favorite solution for securing payment card data; removing credit card numbers significantly improves compliance and security.
DATA MASKING-PERSISTENT	Allows organizations to protect (de-identify or desensitize) data for application testing, analytical or reporting repositories, and outsourcing. Persistent masking permanently changes data, typically in a copy or subset of the production database.	Not well known or understood by security professionals.	Persistent data masking has made strong inroads in protecting development and test environments and is emerging as the best alternative for outsourcing and big data
DATA MASKING-DYNAMIC	Masks data when it is read but never changes the original data.	Not well known or understood by security professionals.	Dynamic data is gaining acceptance with early adopters through its ability to limit access to data per privacy regulations for both applications and BI.
DATA ARCHIVING	Reduces risk by removing unused sensitive data; meets regulatory guidelines that recommend inactive sensitive data retirement.	Archived data must be carefully selected and protected. Archiving is perceived as a part of data management.	Archiving has been identified as an important data security control with recommendations in privacy and security regulations for data retirement.

Which Data Security Control for Which Use Case

While not absolute, data security controls can easily be aligned to data security use cases that exist in most organizations. As part of a comprehensive data security architecture, these controls provide strong protection against external or internal threats.

Data at Rest

For data at rest (such as data stored on a HDD or USB device), encryption is widely accepted as the de facto standard. With encryption, data is stored in cypher text, scrambled beyond human recognition. If this data is accessed via editors or system management tools, it will remain encrypted. Stored data is decrypted when accessed by tools or applications that have the appropriate encryption tool and encryption key.

Data in Use

Application Data

Application data is data that is in use by an application, presented to the user in a form or graphical user interface. The data is typically presented in the clear, with limited control over who views it. In some cases, the data originates from unencrypted databases; encryption does little to help in this scenario because the application will use an encryption tool and key to convert the encrypted data to clear text.

Not every user needs to see all information associated with customer records or transactions; this is influenced by privacy laws, industry regulations, and security policies. Depending on role, a user may not have the need to know identity, financial, or health information associated with the record.

Data masking is the best solution for this scenario. Data masking (specifically dynamic data masking) allows organizations to control access to live data based on role, location, and time. Dynamic masking can be configured to obfuscate, block, randomize, or encrypt sensitive data at the record level. While format-preserving encryption may be considered for this use case, its inherent complexities (key management and application modifications), inflexible deployment, limited techniques for maintaining data context, and cost do not make it a desirable choice. Dynamic data masking requires no changes to applications or databases and can be installed without operational disruption.

Data for Application Testing

Application test teams rely on copies of production data to test application updates and new capabilities. It is common for testers to make 8-10 copies of a production database per test cycle. These copies are particularly vulnerable because they do not have the protections of production databases but contain all the original sensitive data.

Persistent data masking is the solution of choice for securing data in application testing. It provides several key capabilities to allow test and development teams to de-identify and desensitize production data while keeping the data in its context. For example, a Social Security number retains its format, but not the actual value. Further, masking is performed while retaining relationships of parent-child tables. Many other capabilities allow testers to mask data while retaining its context for realistic and reliable application testing.

Shared Data (Analytics/Reporting, Outsourcing, Industry Research)

Many outsourcing initiatives require production data to support customer service, billing, and support services. Industries and governments have launched data- sharing programs for the purpose of trend analysis on markets and constituents. Trading partners in supply chains share data to help drive better B2B management and efficiency.

Sharing data without compromising privacy requires the data to be de-identified and desensitized to some degree. As with data for application testing, de-identified data must retain its format and context so that reporting and analytical applications operate properly on the data. Persistent data masking provides the ability to de-identify and desensitize data with simple-to-complex masking scenarios while preserving the integrity of data fields.

Payment Card Data

For payment card data, the context is the security of credit card numbers. As a prime target for hackers and malicious insiders, these numbers present one of the highest-value data assets that can be stolen or misused. While encryption has been a fundamental requirement of credit card data, tokenization has gained popularity and endorsement over encryption, because encryption can be broken when focused on small samples. With

tokenization, payment card data is removed from databases and stored in secure vaults. Tokens are issued in place to facilitate payment card authorization and accounting. Removing the credit data from finance and customer services environments significantly reduces the probability of disclosure from these internal sources.

Obsolete or Inactive Data

The safeguarding of legacy data includes the retirement of sensitive/private information, such as health, identity, or credit information that has become inactive. Although this information may not currently be useful to the holding organization, it must be kept confidential to reduce the organization's overall risk and purged in an active, ongoing way to comply with privacy regulations and data security guidelines.

As an additional safeguard, sensitive data can be permanently masked during archiving to reduce the potential scope of future e-discovery requests.

Summary of use cases and suggested data security controls:

	SUGGESTED CONTROL	ALTERNATIVE CONTROL	IMPORTANT NOTES
DATA AT REST			
(disk or removable media)	Encryption	For environments that do not need access to all original sensitive data, permanent masking is a superior option.	Security practitioners should be current on vulnerabilities and what comprises encryption algorithms.
DATA IN USE			
APPLICATION/BI DATA	Dynamic masking	Tokenization	Tokenization works well for high-value assets such as credit card numbers; for broad-scale deployment, dynamic data masking is preferred for its superior scalability, administration, and management.
TEST DATA (Data for application testing)	Persistent masking	Custom scripts	Organizations should manage test data environments carefully, because hackers know this can be a vulnerable and poorly protected area for sensitive data.
SHARED DATA (Data provided for reporting or analytics; internal enterprise data warehouses; or external uses such as outsourcing and B2B collaboration)	Persistent masking	Custom scripts	For auditability and consistent implementation of policies, masking is strongly recommended to de-identify and desensitize data for general-purpose use.
PAYMENT CARD DATA	Tokenization	Encryption	While encryption is an important component of securing data at rest, tokenization offers clear advantages because it removes payment card data.
OBSOLETE/INACTIVE DATA	Archiving	Deletion	While it may be tempting to simply delete unused or inactive sensitive data, data retention laws and guidelines dictate that archiving is a best practice.

About Informatica

Informatica Corporation (Nasdaq:INFA) is the world's number one independent provider of data integration software. Organizations around the world rely on Informatica to realize their information potential and drive top business imperatives. Informatica Vibe, the industry's first and only embeddable virtual data machine (VDM), powers the unique "Map Once. Deploy Anywhere." capabilities of the Informatica Platform. Worldwide, over 5,500 enterprises depend on Informatica to fully leverage their information assets from devices to mobile to social to big data residing on-premise, in the Cloud and across social networks. For more information, call +1 650-385-5000 (1-800-653-3871 in the U.S.), or visit www.informatica.com.

Conclusion

Data security controls are essential to stem the tide of data breaches by securing data at the source. While each type of data security control has its own value in specific use cases, the best way to approach data security is to apply all of these controls where appropriate. This creates overlapping coverage, ensuring all appropriate protections are available against attacks that have breached cyber security controls or have been enabled by insider threats.

Put potential to work.™

Worldwide Headquarters, 2100 Seaport Blvd, Redwood City, CA 94063, USA Phone: 650.385.5000 Fax: 650.385.5500
Toll-free in the US: 1.800.653.3871 informatica.com linkedin.com/company/informatica twitter.com/InformaticaCorp

IN09_0515_2907