

# Approaches for Encryption of Data-at-Rest in the Enterprise

*A Detailed Review*

---

**Abstract**

This white paper discusses the motivations for and approaches to encrypting data-at-rest in the enterprise. Justification for deployment and tradeoffs between different methods are also discussed.

January 2008

---

---

Copyright © 2008 EMC Corporation. All rights reserved.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED “AS IS.” EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on EMC.com

All other trademarks used herein are the property of their respective owners.

Part Number H4173

---

## Table of Contents

<b>Executive summary .....</b>	<b>4</b>
<b>Introduction .....</b>	<b>4</b>
Audience .....	4
Terminology .....	4
<b>Encryption overview .....</b>	<b>5</b>
The process .....	5
Encryption algorithms .....	6
Key management .....	6
Configuration management .....	7
<b>Application of encryption .....</b>	<b>7</b>
Risk assessment and management .....	8
Modeling threats .....	9
Use cases for protecting data-at-rest .....	10
Use considerations .....	11
Deployment options .....	11
Application level .....	11
Host level .....	13
Network level .....	14
Device level .....	16
<b>Conclusion .....</b>	<b>19</b>
<b>Appendix A: Encryption modes .....</b>	<b>21</b>
<b>Appendix B: Standards .....</b>	<b>22</b>
FIPS 140-2 .....	22
FIPS 197 .....	23
Modes of operation for encryption .....	23
IEEE P1619 .....	23
<b>Appendix C: Encryption types .....</b>	<b>23</b>
<b>Appendix D: U.S. information classification levels .....</b>	<b>23</b>

---

## Executive summary

An evolution in the nature of security risks has raised the imperative for information-centric security that manages the relationship between data – increasingly the lifeblood of organizations – and the people who use this data. Information-centric security includes three critical measures: securing enterprise data by protecting confidentiality and integrity regardless of location; securing data access by employees, partners and customers; and managing security information to comply with security policies and regulations. To answer this demand, EMC is integrating technology from its security division, RSA, into its portfolio of storage products.

The following products from EMC allow organizations to protect information at various points in the enterprise depending on the threats involved and the associated risks.

- Backup: NetWorker<sup>®</sup>, Retrospect<sup>®</sup>, Avamar<sup>®</sup>
- Archiving: EmailXtender<sup>®</sup>
- Unstructured content: EMC<sup>®</sup> IRM, RSA File Security Manager, Documentum<sup>®</sup> TCS
- Database: RSA Database Security Manager
- Application development: RSA BSAFE, RSA Key Manager
- Host-based: PowerPath<sup>®</sup>
- Network level: Cisco/Connectrix<sup>®</sup> MDS
- Key management: RSA Key Manager
- Data erasure: EMC Certified Data Erasure
- eDiscovery: RSA DLP Risk Advisor, EmailXtender

To better understand the risks with respect to information EMC offers the following security services:

- Assessment Service for Storage Security
- RSA Classification for Information Security

## Introduction

Protecting data-at-rest in particular is a critical aspect of information centric security. There are many measures that can be taken to secure data-at-rest, including access controls, logical separation, physical security, and, potentially, encryption. This white paper focuses on the aspects of implementing *encryption of data-at-rest* (including online storage systems), justifications for its deployment, and the tradeoffs involved among various implementations. To investigate the topic of encryption, we must examine the threats to information in the enterprise and see how encryption can help to mitigate these treats. Management of the encryption process is critical to the success of any implementation, as will be discussed.

## Audience

This white paper is intended to provide an introduction to encryption technology, uses, and offerings. No prior knowledge of encryption or key management is required.

## Terminology

- **Accounting** – Measuring and documenting resources used during access.
- **Authentication** - Verify (that is, establish the truth of) an identity claimed by or for a system entity.
- **Authorization** – An "authorization" is a right or a permission that is granted to a system entity to access a system resource.

- 
- **Confidentiality** - The property that information is not made available or disclosed to unauthorized individuals, entities, or processes (that is, to any unauthorized system entity).
  - **Data-at-rest** encryption– Data that is stored on the storage array or tape will be encrypted. Encryption can be performed by the host, intermediate box, by the storage array, or tape device itself.
  - **Data-in-flight** encryption– Data is encrypted during the transmission of the data. The data stored on the storage array may be plaintext or ciphertext.
  - **Integrity** - The property that data has not been changed, destroyed, or lost in an unauthorized or accidental manner. This can be accomplished by using a Message Authentication Code (MAC). A MAC is a short cryptographic message used to authenticate the message such as MD5 or SHA.
  - **Nonrepudiation** – A security service that provides protection against false denial of involvement in a communication.
  - **Privacy** – The right of an entity (normally a person), acting in its own behalf, to determine the degree to which it will interact with its environment, including the degree to which the entity is willing to share information about itself with others.
  - **Re-key** – Change the value of a cryptographic key that is being used in an application of a cryptographic system.

## Encryption overview

Encryption is used to prevent disclosure of either stored or transmitted data by converting data to an unintelligible form called *ciphertext*. Decryption of the ciphertext converts the data back into its original form, called *plaintext*.

### The process

Encryption simplifies the problem of securely sharing information by securely sharing a small key used to encrypt the information.

In a two-party system a process similar to these steps would be followed<sup>1</sup>.

1. Alice and Bob agree on an encryption algorithm to be used.
2. Alice and Bob agree on a key to be used for encryption/decryption.
3. Alice takes her plaintext message and encrypts it using the algorithm and key.
4. Alice sends the ciphertext message to Bob.
5. Bob decrypts the ciphertext message with the same algorithm and key as the original encryption process.
6. Any change in the key or encryption algorithm has to be agreed upon between Alice and Bob. The process of converting to a new key or algorithm requires decrypting the ciphertext using the original key and algorithm and re-encrypting with the new key and algorithm. It is important that the key management system used securely preserves the old key for as long as the data retention policy for that data prescribes. Premature destruction of the key will result in loss of data.

The secure exchange of data in a two-party system is typically accomplished using a public/private key mechanism. Protecting data-at-rest, however, is best handled with a symmetric (private) key as the data is accessed from fixed and/or known locations. Typically one host would use the same algorithm and key to

---

<sup>1</sup> Ferguson, Niels, *Practical Cryptography*, Wiley Publishing, 2003

---

encrypt the data when writing to disk/tape and to decrypt the data when reading from disk/tape. In the case of multipathing or situations in which multiple applications from different nodes will access the data, centralized key management is essential.

Throughout the remainder of the paper only symmetric-key encryption will be discussed and will be referred to simply as “encryption.” Symmetric-key encryption, as noted, refers to the process by which data is encrypted and decrypted with the same key. This method of encryption is more suited to the performance demands of data path operations. Asymmetric-key encryption refers to the process where encryption is performed with one key and decryption is performed with another key, often referred to as a public/private key pair. Asymmetric-key encryption is not well suited to encrypting bulk data-at-rest, due to performance constraints and manageability.

## ***Encryption algorithms***

The algorithm used can be any one of a variety of well-known cryptosystems described in the industry. The U.S. Federal Information Processing Standards (FIPS) document the Advanced Encryption Standard (AES<sup>2</sup>) and specify it as the industry-standard algorithm in the United States. AES is the most common algorithm implemented in the current encryption methods described as follows. Triple-DES (Data Encryption Standard) is still a certified algorithm by the National Institute of Standards and Technology (NIST) and may be used but is not recommended<sup>3</sup>.

Encryption algorithms typically operate on block lengths of 64 to 128 bytes. To encrypt longer messages an encryption mode of operation may be used, such as:

- CBC – Cipher Block Chaining
- CTR – Counter
- XTS – Tweakable Narrow Block
- GCM – Galois/Counter Mode

The CBC, CTR, and GCM modes of operation used for encryption require the use of an Initialization Vector (IV), or nonce. The IV is a seed block used to start and provide randomization to the encryption process. The same IV and key combination must not be used more than once. XTS is the only one of the four that does not require an IV but instead has a second key called the tweak key.

In the event the length of the message to be encrypted is not a multiple of the block size it may be required to pad the final block.

## ***Key management***

The protection potentially afforded by encryption is only as good as the management, generation, and protection of the keys used in the encryption process. Keys must be available and organized in such a fashion that they can be easily retrieved, but at the same time, access to keys must be tightly controlled and limited only to authorized users. This attention to key management must persist for the lifetime of the data, not just the lifetime of the system that generates or encrypts the data.

Generation of keys should follow some simple guidelines:

- The key generated must be random, for example, as specified by FIPS 186-2<sup>4</sup>. There can be no predictability to the key used for encryption — pseudo-random number generators are not acceptable for key generation.
- Key length for AES can be 128, 192, or 256 bits.

---

<sup>2</sup> FIPS 197 (<http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf>)

<sup>3</sup> FIPS 46-3 (<http://csrc.nist.gov/publications/fips/fips46-3/fips46-3.pdf>)

<sup>4</sup> <http://csrc.nist.gov/publications/fips/fips186-2/fips186-2-change1.pdf>

---

Once the keys are generated, their protection is crucial to guaranteeing confidentiality. This requires:

- Secure access to the key management solution. The key management solution must provide a method to guarantee that unauthorized access to keys is restricted. This access restriction should also extend to the facility for generating and managing keys. This can be accomplished via a number of mechanisms including secure web, smart cards, or split key arrangements. The key management solution must also protect against physical tampering as outlined in FIPS 140-2<sup>5</sup>.
- Backup and recovery facilities for configuration and key information. This information itself must be encrypted and stored to a secure backup medium (for example, a smart card). The keys used for encryption must never be visible in plaintext outside of the key management solution and, under most circumstances, should not be visible at all. For additional security, the recovery of the configuration/keys should be performed by a group of security administrators. This eliminates the potential for misuse due to corruption of a single administrator and utilizes a group key recovery model where M of N (that is, 2 of 3, 3 of 5, and so on) or a quorum of administrators is needed to reconstruct an encrypted configuration.
- The ability to apply high availability and business continuity practices and protocols to key stores.
- The ability to store keys and identify where they have been used for the lifetime of the data. This covers data that is written to tape and that may be read up to 30 years later.
- Integrity checking of keys. This is particularly important if there are no integrity checks on the data.
- Comprehensive logging and regular auditing of how and when the keys are used.

Key management can be distributed or centralized. A common implementation of these requirements is a key management station that can reside either online with the encryption engine or out-of-band via TCP/IP. The key management station provides a centralized location where keys can be managed and stored securely and meet the stringent standards of FIPS 140-2. At this point in time, there are very few certified, standalone key management systems.

The RSA Key Manager is designed to address all of these concerns to help reduce complexity in encryption deployments by centralizing the provisioning and lifecycle management of encryption keys for all enterprise encryption.

## ***Configuration management***

In configuring any of the methods for encrypting data described below, there are several common steps that need to be executed. The unit to be encrypted needs to be identified (for example, record, file, file system, volume, tape) and an associated key needs to be generated. This configuration information needs to be recorded, securely transmitted to the encryption engine, and securely stored for the lifetime of the encrypted data.

To ensure access to the encrypted data, the configuration must account for all paths available to the data and identify which applications, hosts, or appliances will access the data through those paths. Each needs access to the algorithm and key to be able to read/write uniformly from each path. In addition, replicas (for example, snaps, clones, and mirrors) need to be identified and associated with the original source data to ensure that they can also be correctly decoded when read.

## **Application of encryption**

Encryption is only one tool that can be applied as part of a comprehensive information security strategy, and as such, should be applied selectively, only where it makes sense. Determining exactly where and how this takes place begins with an assessment of risks to the data, the suitability of encryption to address the risk, and then if appropriate, the options for deployment of the technology.

---

<sup>5</sup> <http://csrc.nist.gov/cryptval/140-2.htm>

---

## ***Risk assessment and management***

Risk assessment is a calculation that requires three key pieces of information: the number and nature of threats, the likelihood of a threat being realized in the form of an attack, and the impact to the business in the event the attack succeeds. Let's consider these in the context of a decision of whether it is appropriate to deploy encryption technology.

As administrators manage the flow of data from application to storage, they need to understand the nature of possible threats to the data and the likelihood of occurrence. These threats may take the form of:

- Unauthorized disclosure
- Destruction
- Denial of service
- Unauthorized access
- Unauthorized modification
- Masquerade<sup>6</sup>
- Replay<sup>7</sup>
- Man-in-the-middle attacks<sup>8</sup>

These threats may occur at any point from where the information is generated to where it is stored. For each of these threats, an evaluation must be made as to the likelihood of attacks occurring and succeeding in light of existing protection measures. If any attack is determined to be likely, the value of the information subject to threat must be also considered. If the value to the business of the data being threatened is low, it ultimately may not warrant additional protection.

For those risks deemed to be significant, another calculation is required: are the tradeoffs of the proposed solution (in this case, encryption) worth making in context of the level of threat to the data. Considerations should include:

- Cost to deploy
- Level of threat
- Severity of vulnerability
- Consequences
- Detection time
- Response time
- Recovery time
- New risks introduced by encryption, such as premature loss of keys

In this case, by restricting access to the information via authentication and authorization, the administrator can identify who has rights to use the information as well as who has attempted to use the information. Access privileges can be granted at various points in the information flow: at the application, operating system, network, and storage platform layers. If these measures are deemed insufficient, encryption might provide another layer of defense.

EMC offers an Assessment Service for Storage Security as a service offering designed to help customers identify opportunities for improving the security of their storage environments. The service involves the review of management, technical, and operational controls in order to propose recommendations for risk mitigation. A TS Kit for the offer is available on Powerlink<sup>®</sup>, EMC's password-protected customer- and

---

<sup>6</sup> An attack in which a third party tries to mislead participants in a privileged conversation using forged information.

<sup>7</sup> A form of network attack in which a valid transmission is maliciously or fraudulently repeated or delayed.

<sup>8</sup> An attack in which an attacker is able to read, insert, and modify messages between two parties without either party knowing that the link between them has been compromised.

partner-only extranet, at this path: Services > Plan > Service Offerings > Service Overview: Assessment Service for Storage Security.

The RSA Classification for Information Security service, available on Powerlink, provides a comprehensive review of specific customer data assets from a business value, regulatory, and practical security perspective: Services > Build > Service Offerings.> Service Overview: Classification for Information Security.

## Modeling threats

To make this process more specific to the problem at hand, Figure 1 illustrates some of the risks to data in the enterprise. By understanding the attacks that can occur, administrators can determine where encryption may help to protect data and where it would not be applicable.

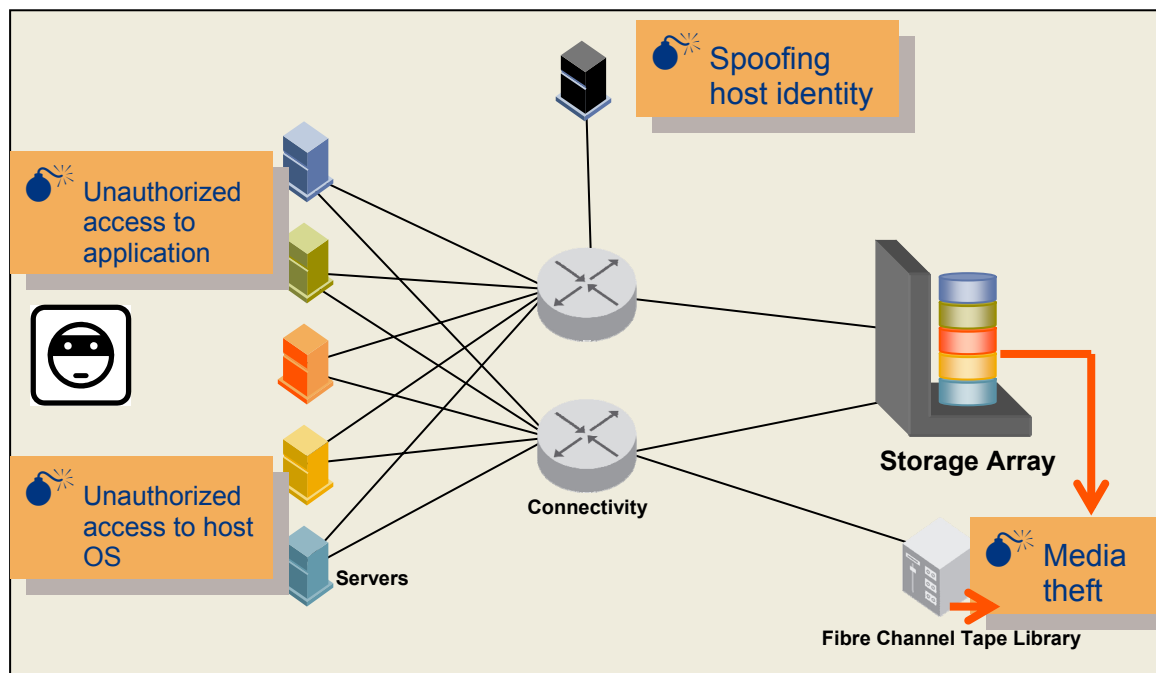


Figure 1. Threats to plaintext customer data

Figure 1 shows the following:

- Encrypting the information at the application level protects against unauthorized viewing of information at the operating system (user) and network levels, as well as protects against media theft. However, encryption at this level will not protect against unauthorized access at the application level (as the information is decrypted at that point) nor root access from the operating system unless strong application access controls are in place.
- Encrypting the information at the host or operating system level protects against unauthorized viewing of information at the network level as well as protects against media theft. Encryption at this level will not protect against unauthorized access at the application or operating system level as the information is decrypted at that point. Access control technology would be required to provide additional security at the operating system and application levels.
- Encrypting the information in the network protects against unauthorized viewing of information from the encryption device to the storage device in the network as well as protects against media theft. Encryption at this level will not protect against unauthorized access at the application or operating system level or in the network up to the encryption device as the information is decrypted at that point.

- 
- Encrypting the information at the device level protects against media theft. Encryption at this level will not protect against unauthorized access at the application or operating system level or in the network as all data external to the device is unencrypted.

## ***Use cases for protecting data-at-rest***

The following are some specific use cases that warrant deployment of encryption of data-at-rest.

The primary use case is protecting data that leaves administrators' direct control. Some examples of this situation include:

- Backup to tape
  - Tapes that are sent offsite
- Removal of disk for repair
  - Key-based data erasure for removed disk or array for return
  - Data sent to a disaster recovery or remote site
- Protection of data between and in disaster recovery sites
  - Consolidating data from many geographies to a single data center while still following each country's security laws
  - Using Type 1 encryption<sup>9</sup> to share data between multiple secure sites
  - Data in harm's way (used for military applications such as planes, Humvees, embassies)
- Data extracts sent to service providers and partners
  - Outsourcing scenarios where sensitive data resides in vendor systems

A second use case is protecting data from unauthorized access in the data center when existing access controls are deemed to be insufficient. Some examples of this situation are:

- Shared/consolidated storage used by numerous groups
  - Sharing a single data center/array for multiple levels of security
  - Sharing a platform between an intranet and internet for consolidation
- Protecting data from insider theft (employees, administrators, contractors, janitors)
- Protection of application/executables from alteration

In addition, data encryption is mandated or recommended by a number of regulations. Deploying encryption will enable or aid in compliance. Selected examples of these regulations include:

- **Sarbanes-Oxley Act** – U.S. regulation with respect to disclosure of financial and accounting information
- **CA 1798 (formerly SB-1386)** – California state legislation requiring public disclosure when unencrypted personal information is compromised
- **HIPAA** – U.S. health care regulation that recommends encryption for security of personal information
- **Personal Information Protection Act** – Japanese regulation on information privacy
- **Gramm-Leach-Bliley Act** – U.S. finance industry regulation requiring public disclosure of personal data breaches
- **EU Data Protection Directive** – European Union directive on privacy and electronic communications
- **National Data Privacy laws** – Becoming pervasive in many nations, including Spain, Switzerland, Australia, Canada, and Italy

---

<sup>9</sup> “Appendix C: Encryption types” provides further definition.

---

## ***Use considerations***

There are additional factors to consider when using encryption.

- Data de-duplication at the disk level may be affected. Any good encryption algorithm will generate different ciphertext for the same plaintext in different circumstances. As a result, algorithms for capacity reduction by analyzing the disk for duplicate blocks will not work on encrypted data.
- Encrypted data is not compressible. Lossless compression algorithms could potentially expand as often as they compress encrypted data if applied. This will impact any WAN connectivity needing to transmit encrypted traffic.
- There is overhead in converting current plaintext data to ciphertext. This is done as a data migration project – even when it is done in place. Host resources, impact to CPU utilization, and running applications must be considered.
- An additional benefit to encrypting data at any level described is the ability to provide data shredding with the destruction of the key. This is especially efficient when there are multiple, distributed copies of the data encrypted with the same key. In order for the data to be considered shredded, all management copies of the key need to be destroyed for all security administrators, smart cards, backups, key management stations, and so on. Key destruction must follow similar guidelines as to the data erasure outlined in NIST SP 800-88.

## ***Deployment options***

As we have seen, the use cases discuss why encryption would be used and the threats being protected against determine where encryption should be deployed. The following sections discuss in further detail deployments at each layer of the infrastructure.

### **Application level**

Perhaps the greatest control over information can be exercised where it originates, from the application. The application has the best opportunity to classify the information and manage who can access it, during what times and for what purpose. If the administrator has concerns/risks over the information at all levels in the infrastructure, it makes sense to begin with security at the application level and work down. In this case application-based encryption should be an option. Adding encryption at the application level allows for granular, specific information to be secured as it leaves the application. For example, a database could encrypt specific rows/columns of sensitive information (for example, Social Security numbers or credit card numbers) while leaving less sensitive information unencrypted. Attempts to snoop writes to disk or to read data directly from disk without the application decrypting it would yield useless information.

Encryption at the application level provides security from access at the operating system level as well as from other applications on the server as shown in Figure 2. The application would still need to provide user authentication and authorization to guarantee that only those with a need to know can access the application and the data. If the application lacks these strong access controls, application-based encryption will provide no additional security benefit. End-user activities with data after it is converted to clear text are potentially the highest security risk for organizations.

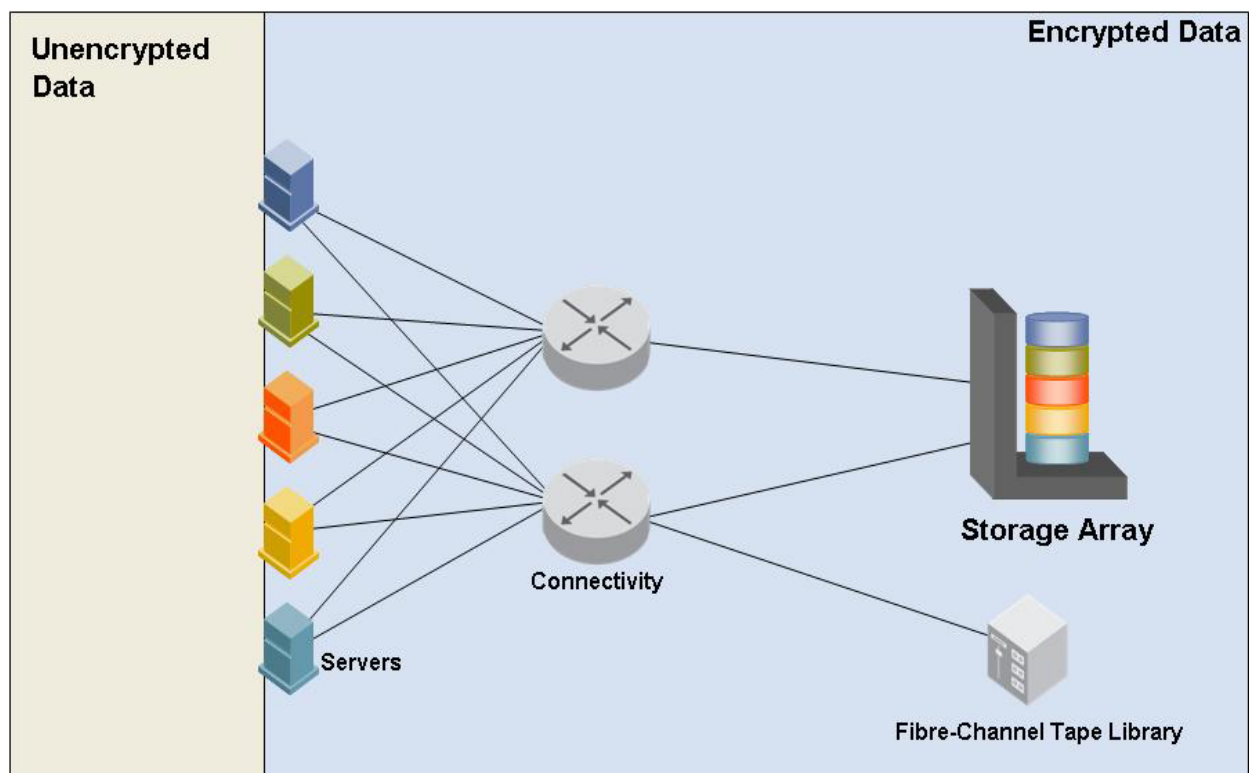
There are some drawbacks to encrypting at the application level. First, encryption is done on a per-application basis. If there are multiple applications needing encryption each would have to handle the task separately, creating additional management complexities to ensure that all confidential data is protected. Second, application-level encryption solutions are typically software-based. Encryption is a CPU-intensive process and will compete with normal operating resources on the server. In addition, the encryption keys will be stored in dynamic, non-volatile memory on the server. If a hacker were to break into the server and find the keys, the information can be decrypted. Externalizing the encryption engine or key manager may address these issues, at the expense of additional solution cost. An external key manager also enables clustered applications to share key information across nodes and geography (provided that each node can

supply a secure channel from the server to the key manager.) If FIPS 140-2 compliance<sup>10</sup> is a requirement for the encryption solution, an external appliance is typically used.

Application-based encryption also presents challenges in the area of re-keying. Any effort to re-key the data (to protect the integrity of the keys) will have to be done by the application. The application will need to read and decrypt the data using the old key and re-encrypt and rewrite to disk using the new key. The application will also have to manage old and new key operations until all the previously encrypted information is re-encrypted with the new key. This most likely will be done while the application is handling normal transactions, again presenting resource contention issues.

Another challenge occurs with the introduction of eDiscovery solutions in the enterprise. Encryption at the application level will expose only encrypted information to other applications (including backup) and devices in the stack. Any attempt to perform analysis on the data will be useless as patterns and associations will be lost through the randomization process of encryption. To accomplish any analysis the eDiscovery applications will need to be associated and linked to the application performing encryption to allow for a decryption of data at a level outside of the application, and a possible security risk could be introduced.

Application-based encryption must also account for variable record lengths. Encryption schemes must pad data up to their block size to generate valid signatures. Depending on the implementation, this may require some changes to application source code.



**Figure 2. Coverage for application-based encryption**

Application-based encryption doesn't take into account the impact on replicated data. Any locally replicated information at the storage layer, that is, a clone, does not have visibility into the application and the keys and the application does not have visibility into the replication process. Key management can

<sup>10</sup> "Appendix B: Standards" provides further definition.

---

become more complex. In addition, compression in the WAN is impossible for remote replication of the encrypted information causing WAN capacity issues.

EMC helps address information protection at the application level, both in providing application development support with the RSA BSAFE tools and the RSA Key Manager and delivering application encryption with the following solutions:

- Backup: NetWorker, Retrospect, and Avamar feature native encryption.
- Archiving: EmailXtender encrypts all user messages in local cache.
- Unstructured content: Documentum Trusted Content Services provides file store encryption to secure content in repositories, and Documentum Information Rights Management encrypts documents to control viewing, printing, copying, and other activities once documents have left the repository. RSA File Security Manager, meanwhile, encrypts laptop, desktop, and server files.
- Database: RSA Database Security Manager encrypts database objects within IBM, Oracle, Microsoft, Sybase, and Teradata environments.

## Host level

Encrypting at the host level provides very similar benefits and tradeoffs to application-based encryption. At the host level, there are still opportunities to classify the data, but on a less granular basis — encryption can be performed at the file level for all applications running on the host (as shown in Figure 3). However, there are options for a host-based adapter or software to provide encryption of any data leaving the host as files, blocks, or objects. One example for host-based encryption operating at the logical unit level, blocks, is EMC PowerPath. As with application-level implementations, the operating system must still provide user authentication and authorization to prevent against host-level attacks. If these strong access controls are absent, host-level encryption will provide no additional security benefit (aside from protection against loss or theft of media.) If implemented correctly and integrated with the encryption solution, they can provide some process authorization granularity, managing which users should be allowed to view plaintext data.

At the host level, encryption can be done in software, using CPU resources to perform the actual encryption and storing the keys in memory, or offloaded to specialized hardware. Offload involves use of an HBA or an accelerator card resident in the host to perform the actual encryption of the data. In the case of the HBA the encryption can be performed in-band and is dedicated to the particular transport connection from the host, that is, Fibre Channel. For an accelerator card approach, the encryption is done as a look-aside operation independent of the transport. This provides flexibility for host connectivity but increases the memory and I/O bus load in the system. In either case the host software would control the connection to the key manager and management of the keys.

There may be a need in the enterprise for the host-based encryption solution to support multiple operating systems, allowing for interoperability across systems or consistency in the management domain, something to consider when evaluating solutions. In addition, when encryption is implemented at the host level there is the flexibility of being storage- and array-independent, allowing for support of legacy storage with no new hardware needed. Host-based encryption does present a challenge when coupled with storage-based functionality, that is, replication. If replication is employed underneath the host encryption level the host implementation must have the ability to track replicas and associate encryption keys, eliminating the need for users to manually manage the replication and encryption technology. As host encryption supplies encrypted data to the array, remote replication would transmit encrypted, uncompressible data. This would severely impact WAN performance.

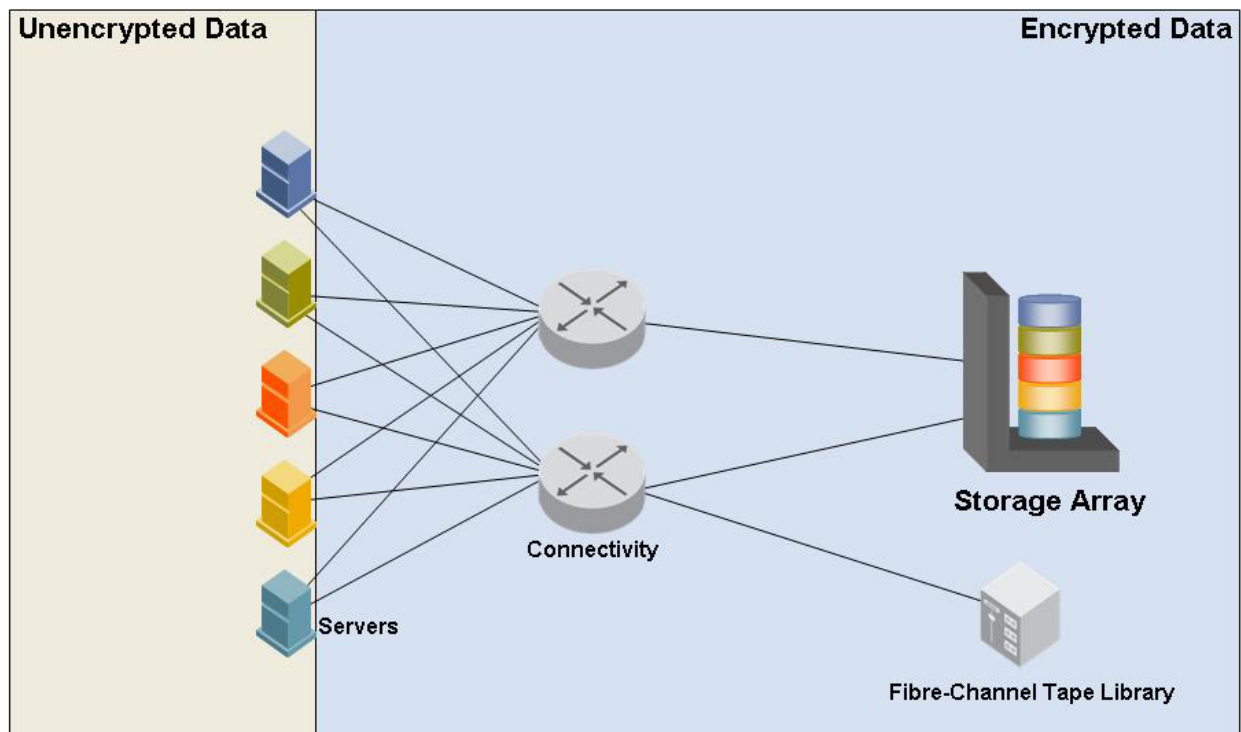
PowerPath provides mechanisms to deal with both the cross operating system interoperability and replication issues. As PowerPath provides consistent functionality through releases on Solaris, Windows, Linux, AIX, and HP-UX, there is a single implementation of encryption and key management, independent of the operating environment. Also PowerPath has developed a unique approach to managing replicas with

---

encryption, allowing for coordinated key management between source and replicated volumes independent of user intervention.

As with application-based encryption, eDiscovery solutions in the enterprise pose additional complexities. Encryption at the host level will expose only encrypted information to other hosts and devices in the stack, introducing the same challenges with analysis as those described in the prior section.

As encryption is performed at the host level, the data can be of variable record length. Similar to the application-based approach, the encryption solution can add information to the encryption payload to allow for a digital signature or cryptographic authentication. This would prevent a “man-in-the-middle” from substituting bad packets for the good encrypted packets from the host. PowerPath performs block-based encryption at the Logical Unit level with no added data to the payload.



**Figure 3. Coverage for host-based encryption**

EMC offers solutions to address information protection at the host level, both in providing application development support with the RSA BSAFE tools and the RSA Key Manager and delivering host encryption with

- Backup: NetWorker, Retrospect, and Avamar
- Unstructured content: RSA File Security Manager
- Host-based: PowerPath will feature block-based encryption on Logical Units

### Network level

If the threats in the enterprise are not at the server, operating system, or application level, but instead at the network or storage level, then a network-based appliance approach for encryption may work best. This approach is operating system-independent, and can be applied to file, block, tape, Fibre Channel, iSCSI, or NAS data. Encryption and key management are handled entirely in hardware and run at wire speed for the connection. The appliance presents an “unencrypted side” and “encrypted side” to the network. Encryption

---

can be designated on a per block, file or tape basis and the keys maintained for the life of the data. Appliances available today are typically FIPS 140-2 level 3 validated<sup>11</sup>.

There are two implementations for a network-based appliance design: “store-and-forward” or “transparent.” The store-and-forward design appears as storage to the server and a server to the storage, and supports iSCSI, Fibre Channel, SAN, NAS, and tape. An I/O operation comes to the appliance, is terminated, the data encrypted, and then forwarded to the destination storage device. This approach adds latency and as a result, some form of “cut-through”<sup>12</sup> ideally needs to be offered to minimize the impact of the device for non-encrypted traffic. In addition, to appear as both server and storage, the store-and-forward appliance either needs to spoof the identities of the attached devices or rely on robust security practices to counteract the attempts to circumvent the appliance. While there may be a latency penalty for encrypting data through the appliance, the store-and-forward-based design has the benefit of allowing the attached storage devices to be re-keyed in the background. This is performed with no disruption to host operations as all I/O operations to the storage are handled independently of the host. There may still be some performance impact to the re-keying process, depending on the I/O load on the encryptor.

The transparent approach provides a flow-through model for the data being encrypted, supporting Fibre Channel SAN and tape. The appliance inspects SCSI headers as data flows through the appliance and encrypts only the data payloads that match preset source/destination criteria in the appliance configuration. The latency associated with this approach is minimal. The transparent design does, however, have a drawback when the encrypted data needs to be re-keyed. Unlike the store-and-forward design, the device is essentially transparent in the data flow, requiring the host to perform the reads and writes required in re-keying of the encrypted data. This process can be done by a separate host agent and could be performed while normal operations are in process.

For block-based implementations, the size of the encrypted data cannot increase. This means no additional information can be added to the encrypted payload (for example, a digital signature). This is not true for file or tape-based encryption where the record information may be variable. As noted in the discussion on standards, the IEEE is working to provide standards for encrypting block data-at-rest, in IEEE P1619<sup>13</sup>.

There may be a need in the enterprise for the encryption to support multiple operating systems, allowing for interoperability across systems or consistency in the management domain. In addition, when encryption is implemented at the network level, there is the flexibility of being storage- and array-independent, allowing for support of legacy storage — at the cost of adding new hardware. Hardware in this case is added in increments of ports, typically two at a time, adding to the power, package, and cooling issues currently facing enterprises today. In addition, adding appliances in these increments can add complications in managing additional devices in the enterprise. Network-level encryption does present a challenge when coupled with storage-based functionality such as replication. If replication is employed underneath encryption at the network level, the implementation must have the ability to track replicas and associate encryption keys, eliminating the need for users to manually manage the replication and encryption technology. As network-level encryption supplies encrypted data to the array, remote replication would transmit encrypted, uncompressible data. This would severely impact WAN performance.

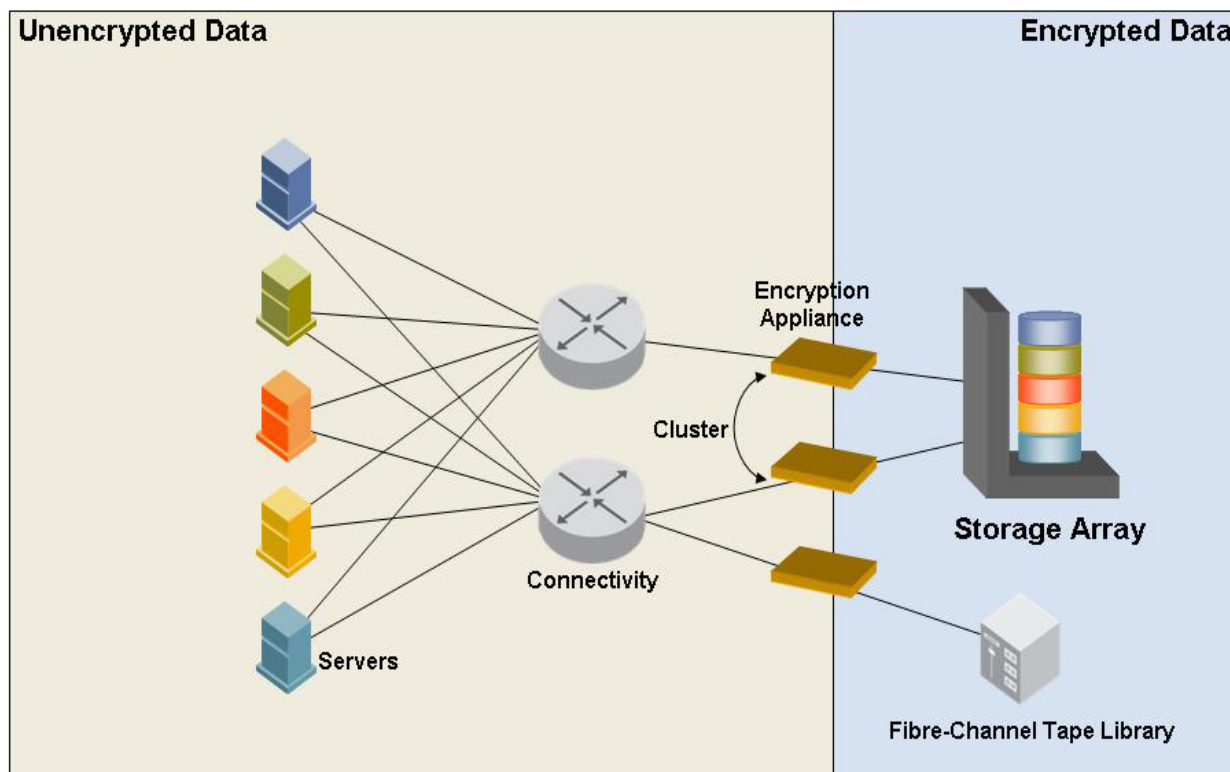
There are also implementations moving to use data integrity features as part of the protocols. Encryption in the network level would encrypt both the data and the data integrity, resulting in mismatches at this level of checking performed at the arrays.

---

<sup>11</sup> “Appendix B: Standards” provides further definition.

<sup>12</sup> A technique whereby a network device forwards frames or packets as soon as the destination address is processed (before the whole frame has been received).

<sup>13</sup> “Appendix B: Standards” provides further definition.



**Figure 4. Coverage for network-based encryption**

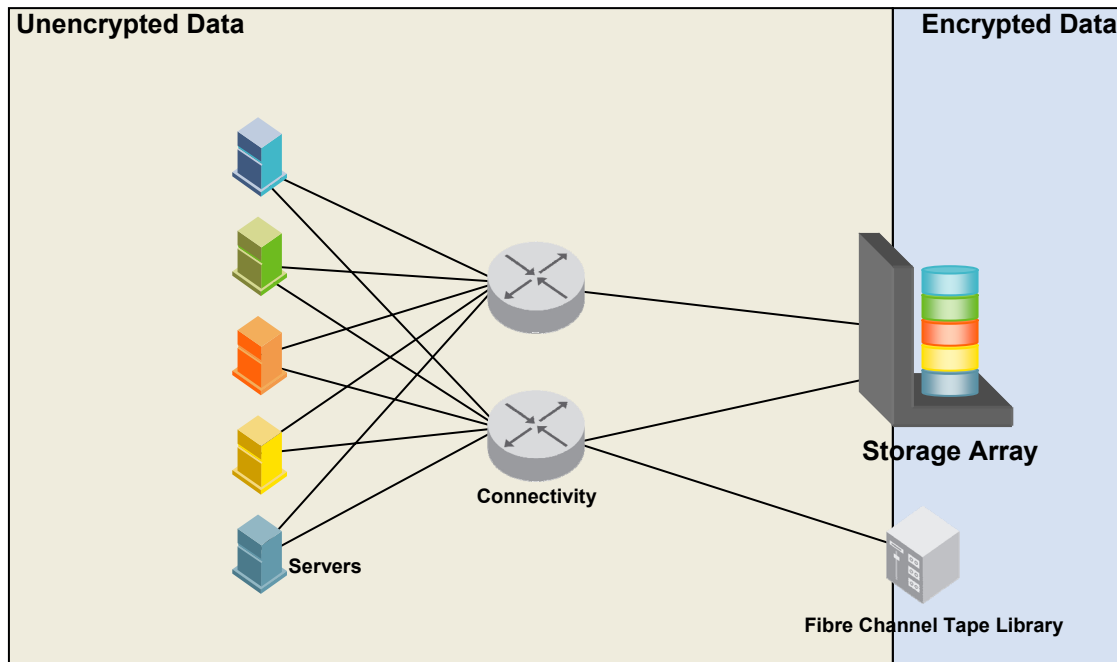
Network-level encryption doesn't take into account the impact on replicated data. Any locally replicated information at the storage layer, that is, a clone, does not have visibility into the network device management and the keys and the network device does not have visibility into the replication process. Key management can become more complex and more manual. In addition, compression in the WAN is impossible for remote replication of the encrypted information causing WAN capacity issues.

EMC offers solutions to address information protection in the network through partners:

- Network level: Cisco Storage Media Encryption (SME), or Connectrix MDS, provides encryption of data-at-rest as a service with its switches.

### Device level

Encryption at the device level — array, disk, or tape — is a sufficient method of protecting sensitive data residing on storage media, which is a primary security risk many organizations are seeking to address. All data written to the device would be encrypted and stored as such, and then decrypted when read from the device. Encryption at this level would be application- and host-independent, and can be transport-independent as well. When addressing media theft the granularity for encryption, and keys, can be at the disk or tape level. As demonstrated in Figure 5, exposure for unencrypted data is increased as compared to the previous implementation examples.



**Figure 5. Coverage for device-based encryption**

### Array-level encryption

There are a number of design points for encryption in the array, that is, at the disk or controller level. Design considerations for encryption include the interfaces to the array, software support, performance, FIPS validation, key management, and encryption object granularity to name a few. The intent is to have the encryption implementation transparent to the hosts attached while protecting the removable media. The connected hosts may not be knowledgeable of the encryption implementation but may be with respect to management and performance. All aspects of the design must be considered.

One possible approach is to implement the encryption in the disk drive, at the back end of the array. Some points to consider:

- As encryption is on a per-drive basis, the computes required are included in the drive enclosure, allowing for a scalable solution, adding encryption with every unit. The downside to this is cost to the functionality that is added with every unit. So while performance scales, so can cost.
- Customers might be unable to verify that encryption is enabled and functioning on the array, because data is always plaintext when it is external to the disk drive.
- Any approach to encryption at this level would also require interoperability of the encryption implementation across drive vendors to maintain flexibility and customer choice.
- Bulk drive encryption would not provide key granularity at the LUN/device level, which in many cases would eliminate the possibility of erasing specific confidential projects via key deletion.
- Lastly, as driven by the Trusted Computing Group, encryption at this level may follow a different path for validation, an alternative to FIPS 140 yet to be developed. Without a standard to evaluate it is impossible to understand the disk drive encryption validation proposal.

Another approach might be to implement encryption in the I/O controller connected to the disk drives. Some points to consider for this potential implementation are:

- Encryption is on the interface level and is required to support full wire speed versus interface speed in the drive approach.
- The cost model would be based on a single controller versus 10s of drives connected to a single controller.

- 
- The controller approach has the ability to perform encryption at the I/O level, allowing the granularity for key management to be at the LUN or disk level. This approach allows for future support of LUN-based erasure and logical data management.
  - The controller approach is drive-independent, not relying on any specific vendor or interface, allowing for all standard tools and failure analysis to be performed.
  - In supporting encryption at the controller level the crypto boundary can be well defined, allowing for FIPS 140-2 validation.
  - Encryption and key control would be separate from the disk drive containing the encrypted data. This allows the customer to validate the encryption functionality is working and not be concerned with keys leaving with a removed disk drive.

An alternative to the encryption option for the protection against media theft is EMC Certified Data Erasure. It addresses the same primary use case: protection of disk media containing sensitive data and is available today, as erasure services (for removed drives) and software (for in-frame erasure). Erasure overwrites data multiple times, in accordance with the Department of Defense specification 5220.22-M, removing the data from the media. One consideration is that a minority of drives are not erasable for mechanical reasons. Customers can keep these drives in a secure onsite location through the EMC Disk Retention Service.

### **Tape encryption**

As part of normal operations, data is frequently written from storage devices to tape for backup/data protection or third-party use. Data on tape cartridges becomes susceptible to theft or loss due to the size of the tape cartridge and quantity of the number of tapes to track during normal backup operations. To best protect the data on tape against unintended/unauthorized viewing, it can be encrypted. There are several approaches to encrypting tape as part of the backup operation:

- Reading encrypted data from application/disk and writing as encrypted data to tape
- Reading unencrypted data from application/disk and encrypting as part of the backup application
- Encrypting any/all data sent to tape via an encryption appliance in the network
- Encrypting any/all data written to the tape via an encrypting tape library or tape device

Tape encryption also presents key management challenges. Tapes may be stored for an extended period of time before an attempt is made to recover information. During the normal process of managing encrypted data, the application may have re-keyed the data on disk, updating all data on the disk to use a new key. This process would present the application with active data using one key and data on tape using an older key. The application must be therefore be able to manage keys for the lifetime of the data, regardless of where the data is stored.

### **Tape encryption deployment options**

- **Application level** - Backup is typically another operation running on a host as a peer to the encrypting application. Any peer application or process will read data from the storage array as encrypted data. This allows the backup process to write already encrypted data to tape without having to perform the encryption itself. It will, however, prevent data compression during the backup process, as encrypted data is not compressible. As typical compression ratios reduce data volumes anywhere from 2:1 to 4:1, this will impact performance of the backup process if a large amount of bulk data is encrypted. Applications providing encryption can also provide access for authorized peer applications to read data in encrypted or unencrypted form. This would allow a backup application to read data in unencrypted form and allow for compression followed by encryption to be performed as part of the backup process.
- **Operating system/host** - Backup is another process on the host when using host-based encryption. The encryption process in the host operating system has the option of allowing the backup process to read data in encrypted or unencrypted form. If the authorization module determines that the backup process can read plaintext data, backup will receive decrypted data to be sent to tape. Encryption will also need to be performed by the backup application to allow for writing secure tapes. The backup

---

process could take advantage of compression in this data flow. If the backup process is not allowed to view decrypted data, it will read encrypted data from disk and write it as such to tape. As in the application-based approach, compression may not be able to be utilized on this encrypted data, creating potential performance issues. In addition, the encryption engine for the host will have to maintain the keys for the lifetime of the data to ensure that decryption can take place in the event a restore from tape is needed.

- **In the network** - If an encryption appliance is placed in the network, backup can be handled in one of two ways. If backup is volume-based, any data read from the storage array may already be encrypted. The backup application will read the encrypted data and write it directly to tape. In this scenario, there would be no benefit of compression in the data path. If the backup is file- or incremental-based, the backup process would read the data through the appliance, decrypting it in the process, and could then write the data to tape. To provide encryption, a tape encryption appliance would be positioned in front of the tape device, compressing and encrypting the data as it is written to tape. The tape encryption appliance would manage the keys for the lifetime of the tape.
- **At the tape library/drive** – Data can be encrypted at the tape drive level, independent of the backup process and application software. All encryption is performed at the device, or library, when data is written to the tape and decryption performed at the drive when data is read from the tape. The backup application deals with nothing but plaintext data. The tape drive or library can be the management interface to the key manager, requesting generation of keys for new tapes written and retrieval of keys for each tape read. Association of keys to tapes is managed at the key manager appliance. In some cases the key manager can be integrated to work co-operatively with a volume pool policy defined with backup application. Jobs directed to use tapes in a pool associated with this policy begin with a request by the drive or library for an encryption key only when the backup or restore job uses tapes in this volume pool.

## Conclusion

Encryption is a tool that can be used to protect the confidentiality of the information in the enterprise. To understand if and how an encryption solution should be deployed, administrators need to understand and assess the risks of unauthorized access and disclosure at each point of the information flow. They must also understand how deployment of encryption technology may add risk to other areas of the business, including complexity added to management, and risks to availability of encrypted data to authorized users. Data unavailability can come from something as simple as key management, which is perhaps the most important factor to consider in implementing an encryption solution. Encryption should be considered as part of a total security solution, but not the only solution – administrators need to take advantage of protection options at all levels of the information flow and architecture.

Two general issues that are present across encryption implementations are:

- The conversion of plaintext to ciphertext when encrypting data for the first time or ciphertext to ciphertext when encrypting with a new key. Both are done as a data migration project - even when it is done in place. Host resources, impact to CPU utilization, and running applications must be considered.
- The replication of encrypted data across the WAN. Encryption, if done correctly, produces random, uncompressible data that will impact the utilization of remote connectivity.

Table 1 summarizes the various deployment options.

**Table 1. Summary of encryption approaches**

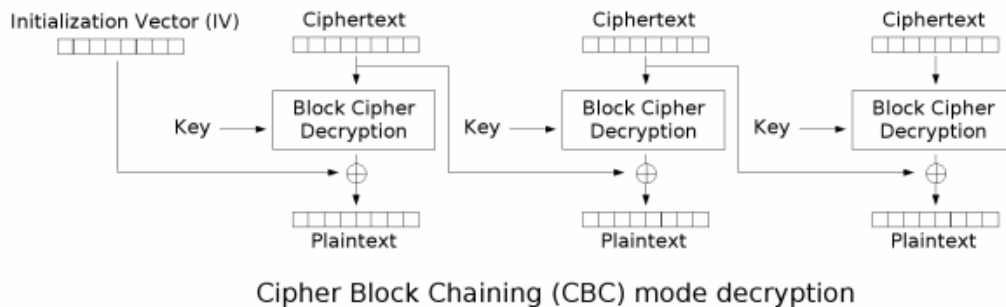
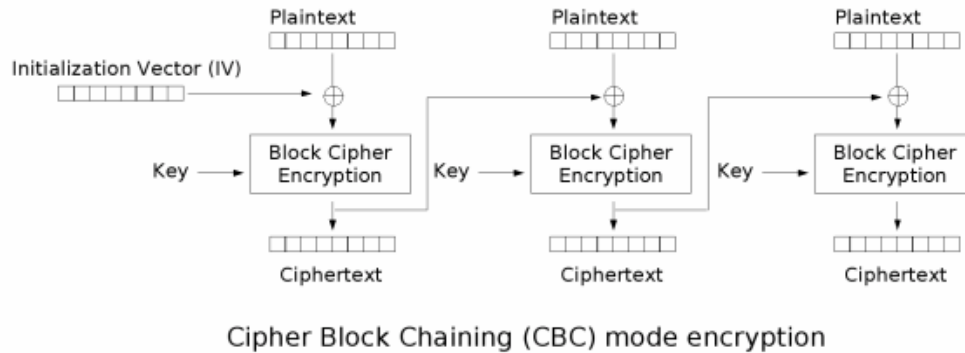
	<b>Encryption</b>	<b>Key management</b>	<b>Backup</b>	<b>Issues</b>	<b>Risks addressed</b>
<b>Application</b>	Typically done in software but can be done in hardware.	Typically stored in memory or file. Coordination of keys across applications presents challenges to sharing information. Needs external appliance to meet FIPS 140-2 Level 3	Peer process to the application and will back up encrypted data. No compression. Lifetime key management challenges.	Encryption can be host system intensive and is a per-application process. If more than one application is used on a host, sharing of information can be an issue. Storing keys for lifetime of data can also be an issue for application upgrades. Can impact eDiscovery.	Protects against operating system and network attacks as well as media theft.
<b>Host</b>	Typically done in software but can be done in hardware. Can be file- or block-based.	Typically stored in memory but can have external appliance.	Peer process will back up data and host will need to re-encrypt.	Encryption can be host system intensive. Per operating system approach needed if PowerPath not implemented. Storing keys for lifetime of data can also be an issue for OS upgrades (if external key management facility is not used.) Can impact eDiscovery.	Protects against network attacks as well as media theft.
<b>Network</b>	Typically done in hardware.	Managed for the lifetime of data in hardware.	Can perform block-based encryption to disk or tape or file-based encryption. Can also incorporate compression for tape backup or coordination for replication.	A single aggregation point in the network for encryption can be a performance bottleneck.	Protects against some network attacks as well as media theft.
<b>Disk-based</b>	Typically done in hardware but can be done in software	Can be done per disk or LUN. Key management can be resident to array or leveraged from external appliance.	Always presents unencrypted data external to disk	Handles very focused use case. Largest exposure of encrypted data in enterprise.	Protects against media theft.

## Appendix A: Encryption modes

Encryption algorithms typically operate on block lengths of 64 to 128 bits. To encrypt longer messages an encryption mode of operation may be used, such as

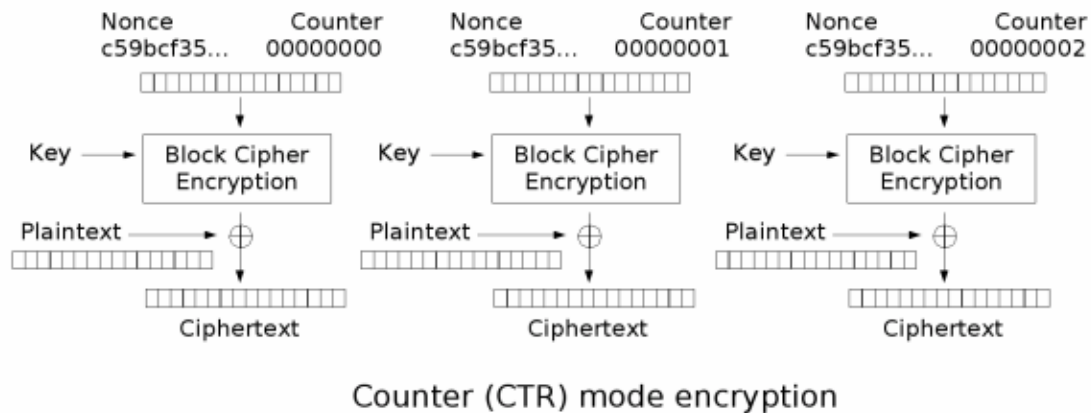
- CBC – Cipher Block Chaining
- CTR – Counter
- XTS – Tweakable Narrow Block
- GCM – Galois/Counter Mode

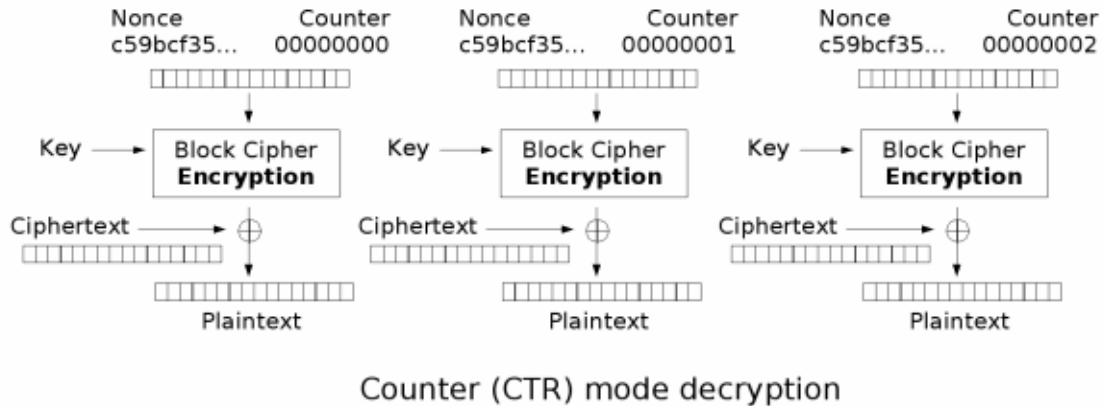
In CBC mode each block is XOR'd with the previous ciphertext block before it is encrypted, as shown in Figure 6. Decryption is a reverse of the process



**Figure 6. CBC mode encryption and decryption**

In CTR mode successive values of a counter are encrypted and that encrypted value is XOR'd with the plaintext to produce a ciphertext. The decryption is the reverse of this process, as shown in Figure 7.





**Figure 7. CTR mode encryption and decryption**

GCM mode provides both confidentiality and authentication. It has been developed as a means to deliver encryption at 10 Gb/s. While GCM is the preferred mode for current standards development, it is not yet on the NIST standards list. One key difference GCM brings is the authentication aspect, which will increase the encrypted payload size, changing the size of the data in disk.

All modes of operation that would be used for encryption require the use of an Initialization Vector (IV), or nonce. The IV is a seed block used to start and provide randomization to the encryption process. The same IV and key combination must not be used more than once.

In the event the length of the message to be encrypted is not a multiple of the block size it may be required to pad the final block.

## Appendix B: Standards

There are a number of standards governing the encryption implementation for data-at-rest. Standardization provides a minimum level of functionality that implementations must meet in relation to specific areas of security as well as some elements of interoperability.

### FIPS 140-2

FIPS Publication 140-2 describes Security Requirements for Cryptographic Modules<sup>14</sup>. Any implementation of encryption or other crypto protection must meet FIPS 140-2 requirements before U.S. federal government entities are allowed to implement it. The standard outlines multiple levels of protection, each with increasingly stringent documentation requirements:

- **Level 1:** The lowest level of security. No physical security mechanisms are required in the module beyond the requirement for production-grade equipment. This is an appropriate level for software-only solutions. Although it does not explicitly require tamper-evident physical security, tamper evidence can be built in for improved security.
- **Level 2:** Tamper evident physical security or pick resistant locks. Level 2 provides for role-based authentication. It allows software cryptography in multi-user timeshared systems when used in conjunction with a C2 or equivalent trusted operating system.
- **Level 3:** Tamper resistant physical security. Level 3 provides for identity-based authentication.
- **Level 4:** Physical security provides an envelope of protection around the cryptographic module. Also protects against fluctuations in the production environment.

In the enterprise, FIPS 140-2 Level 3 for key management is the most common requirement.

<sup>14</sup> <http://csrc.nist.gov/publications/fips/fips140-2/fips1402.pdf>

---

## **FIPS 197**

The Advanced Encryption Standard (AES) specifies a FIPS-approved cryptographic algorithm that can be used to protect electronic data<sup>15</sup>. The AES algorithm is a symmetric block cipher that can encrypt and decrypt information. The AES algorithm is capable of using cryptographic keys of 128, 192, and 256 bits to encrypt and decrypt data in blocks of 128 bits. AES is the preferred NIST algorithm for encryption.

### **Modes of operation for encryption**

In Special Publication 800-38A, NIST specifies five modes of operation for use with AES algorithm. The modes in SP 800-38A are the Electronic Codebook (ECB) mode, the Cipher Block Chaining (CBC) mode, the Cipher Feedback (CFB) mode, and the Output Feedback (OFB) mode. These are updates to the modes that are specified in FIPS Pub. 81; in addition, SP 800-38A specifies the Counter (CTR) mode.

## **IEEE P1619**

The IEEE P1619 working group is focused on defining standard architectures for encrypted shared storage media<sup>16</sup>. These standards will theoretically allow for interoperability among vendors of encrypted storage. The areas of focus include:

- Standards for cryptographic algorithms, modes, and methods for encrypting data before it is sent to a storage (disk or tape) device
- Formats for specifying keys and related configuration parameters required in backup and recovery of encrypted data
- Common Criteria protection profiles for the standards defined

To date, the P1619 working group has created P1619.1 to specify tape encryption, using GCM. The original P1619 document has only specified XTS as an encryption method for disk as well as defined the secure key exchange mechanism.

## **Appendix C: Encryption types**

There are four types of encryption products in use:

- Type 1 is an algorithm or device or assembly of algorithm and device that is specifically approved by the director of the NSA/CSS for the protection of classified information.
- Type 2 is a cryptographic algorithm or device approved by NSA for protecting sensitive unclassified information (as specified in section 2315 of Title 10, United States Code, or section 3502(2) of Title 44, United States Code).
- Type 3 is a cryptographic algorithm or device approved as a Federal Information Processing Standard.
- Other includes algorithms and implementations that are not reviewed by NSA/NIST or FIPS.

## **Appendix D: U.S. information classification levels**

Due to national security concerns some of the encryption devices and algorithms are subject to export restrictions, mostly just the type 1 and 2 devices mentioned previously. The U.S. government classifies information according to the degree to which the unauthorized disclosure would damage national security:

- **Top Secret** – This is the highest security level, and is defined as information that would cause "exceptionally grave damage" to national security if disclosed to the public. Despite public mystique, relatively little information is classified as "top secret" (when compared to the other levels of classification). Only that which is exceptionally sensitive (weapon design, presidential security information, nuclear-related projects, various intelligence information) is classified at the Top Secret level.

---

<sup>15</sup> <http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf>

<sup>16</sup> <http://siswg.org>

- 
- **Secret** – This is the second highest classification. Information is classified secret when its release would cause "serious damage" to national security. By far, most information that is "classified" is held at the secret sensitivity.
  - **Confidential** – This is the lowest classification level and is defined as information that would "damage" national security if disclosed.
  - **Unclassified** – Not technically a "classification," this is the default, and refers to information that can be released to individuals without a clearance. Information that is unclassified is sometimes "restricted" in its dissemination, although such restrictions are generally meaningless.
  - **FOUO** (for official use only)
  - **Sensitive but unclassified**