Addressing the Unstructured Data Protection Challenge

Security that Stays with the Data

Executive Summary

The concept of data-centric security is getting a lot of attention from the media and corporate executives. While most professionals agree that protecting sensitive customer and company data is a top priority, few can agree on precisely what data-centric security means and how best to implement it.

Part of the problem is that data is ubiquitous, comes in many forms and is stored in many places. Critical data, like a customer contact list, can simultaneously exist in a highly structured database on a data center mainframe, in a software-as-a-service CRM program on the web and as a spreadsheet attached to an email. The data may be transmitted from an internal desktop to a mobile smart device in seconds.

Technologies to protect data are myriad: information and data leak protection gateways, network traffic monitoring and encryption and rights management, to name a few. While some vendors and consultants may claim that one technology is superior to another, the reality is more complicated and dynamic. There are no short-cuts. Comprehensive data-centric protection requires a phased, prioritized approach that applies processes and technologies at multiple layers.

This document provides an overview analysis of the many facets of data-centric protection and explains how organizations can approach the problem strategically. Next we'll concentrate on one key aspect of data-centric security: that of unstructured data and detail selection criteria most companies should consider when choosing an enterprise solution for the unstructured data-centric security problem.

Contents

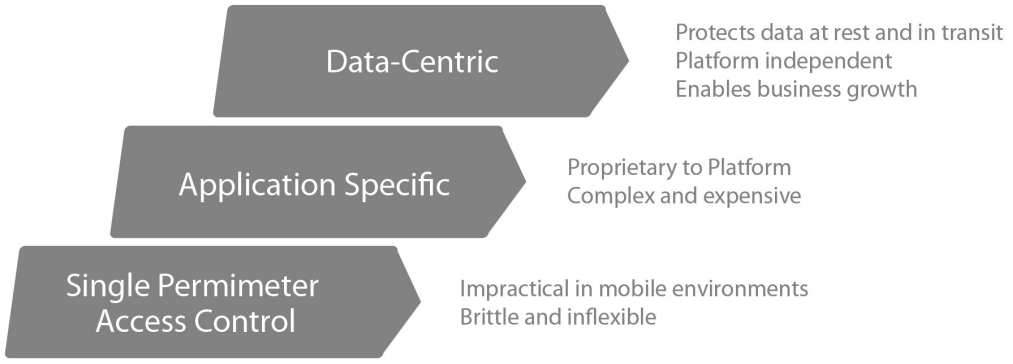
SecurityCurve

Stay ahead of the curve

The Data-Centric Security Challenge

Before massively distributed networking became commonplace in the late 1980s, most enterprise data was physically restricted to a mainframe in a single data center. Terminals provided access, but data was rarely stored on individuals' personal desktops. As desktop processing power increased and the cost of storage decreased, the paradigm shifted permanently.

At first, organizations dealt with the new world of data distribution by cordoning off internal "trusted" networks from the outside. In this model, networks could be interconnected over the Internet or leased lines, with a gatekeeper, such as a firewall, monitoring the perimeter to prevent loss and intrusion. The myth of the "trusted insider" assumed that internal users could always be trusted. While this model was a bit dubious even in the '90s, it is clearly no longer viable in the current reality of de-perimeterization. Placing security into applications emerged as an interim model when the perimeters began breaking down, but it does not adequately address the total data security problem.



High profile cases prove that even centrally stored data is vulnerable to attack.

De-perimeterization refers to the fact that single gateway perimeters do not adequately protect data that is shared and transmitted by an organization in today's IT architecture. There is no simple "inside" and "outside" – there are, in fact, many complex zones. Some zones are relatively fixed, such as the DMZ that houses a company's web servers. Other zones are in constant motion. For example, laptops and mobile smart devices travel from inside a corporate campus to wherever the user goes, such as from a local coffee shop to a conference room half-way around the world.

Some data architectures lend themselves to distributed protection fairly easily. When data is stored in a central place, such as on a protected server (and is not allowed to be duplicated on a remote system), access control can also be managed centrally. Logs and monitors track access to the data and record which users and applications have viewed or altered it. Even in this scenario, however, there are still data-centric protection concerns. Access must be limited to only authorized users and chinks in the protection armor could lead to external attackers viewing or exporting protected data.

High profile cases, such as the computer intrusion at TJX[i], prove that even centrally stored data is vulnerable to attack. Data centers with authentic copies of sensitive information alleviate data leakage when replication copies are prohibited. This approach is not without reasons for caution, however. For some business cases, a single copy of data that is only accessed live limits use because it requires an active connection. Data centers are also potential single points of failure if the center's physical and logical access is not properly controlled. The controls must be able to support data-centric security and business access

needs simultaneously. A secure data center solution that violates Service Level Agreements (SLAs) and is unable to accommodate large numbers of users is not meeting the needs of the business. Enterprises need solutions that can meet both SLAs and security requirements.

The data-centric protection problem is exacerbated by the fact that much of an enterprise's most critical data is not kept in a single, non-replicated place. Users require access to data from multiple devices that are not always on a network. Copies of high-value data, from intellectual property, to customer lists, to medical images, are duplicated for essential business purposes. Data is copied and sent to a partner or colleague via email or backed up by an individual for mobile transport on a pen drive or smart device.

Organizations must apply a multi-phased approach that encompasses all types of data.

Secure pipes, such as SSL VPN tunnels and encrypted email transport, protect this data from prying eyes in transit, but when the data arrives at its destination it becomes the property of the target device. If that target or the data itself is not properly protected, the replicated data is at risk.

Consequences of failed data protection range from the unfortunate to the catastrophic. Loss of protected customer data can wreak havoc from a compliance perspective resulting in disclosure fees and penalties. For many organizations, it also translates to a loss of customer confidence and, potentially, lost revenues. If the breach is significant enough, as it was in the case of CardSystems[ii], it can result in the company's rapid demise.

A Multi-Phased Problem

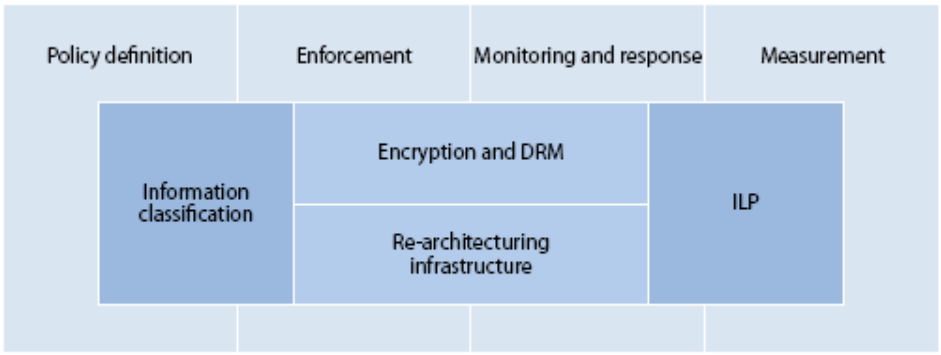
Data-centric security applies to many types of data, residing on multiple devices and platforms, accessed by varied classes of users. There's no single type of data and there's no single solution to data-centric protection. The solution is a phased, prioritized approach that matches protection to the specific data-types and use cases.

Underlying this point in Forrester Research's report, "Making Data-Centric Security Real," analyst Paul Stamp determines that effective data-centric security requires multiple initiatives. Stamp breaks down the facets of a data-centric security program into four initiatives:

1. Information Classification
2. Encryption and Digital Rights Management
3. Re-architecting the Infrastructure
4. Information Leak Prevention (ILP) a.k.a. Data Leak Prevention (DLP)

These are bounded and governed by policy definition, enforcement, monitoring and measurement. The multiple disciplines are illustrated in the following figure:

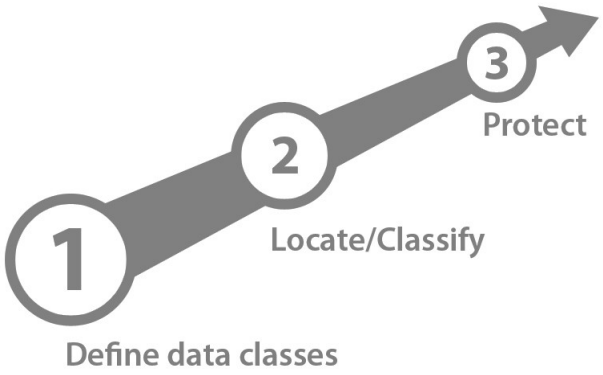
The Multiple Disciplines of Data-Centric Security



44543 Source: Forrester Research, Inc.

In order to implement the disciplines, organizations must apply a multi-phased approach that encompasses all types of data and applies granular policies and solution types. In this schematic, we can refine Stamp's disciplines and initiatives into a strategic three-step framework that forms the basic platform for a data-centric security approach.

The Three Step Approach to Data-Centric Security



Policies are worthless if they're not being applied based on information classification standards.

Define Data Classes

From a security standpoint, it's essential to understand two basic things about data: who and/or what can view it and who/what can alter it. Policies are worthless if they're not being applied based on information classification standards. While it's extremely tempting to apply complex layers of classification rankings to information, most organizations find that starting simply, with three (or maximum five) levels of classification works best as a starting point for security and protection purposes.

Data and information protection requirements vary depending on the classification of the data itself. Some data is meant to be public and accessible by everyone - world viewable, but changed or altered by very few. For example, the name of a company is rarely meant to be secret; company names are not simply public, they're often advertised. On the other hand, ability to change this information is most likely protected and requires a high level of assurance that unauthorized changes are disallowed. Therefore the data may be classified as:

View = *Public* and Alteration = *Highly Protected*

Contrast this with a collaborative product plan being developed by an organization's research and development (R&D) team. The plans themselves will form the basis for the company's strategy going forward and contain trade secrets. Viewing of this information is highly sensitive and restricted. However, for the members of the team, the ability to change and alter information in the plan as brainstorming ideas occur is part of the creative process. So, while alteration is highly protected at one level, restricted only to approved team members, it must also be highly flexible to accommodate the collaborative process. So this data may be classified as:

View = *Sensitive* and Alteration = *Highly Protected/Open*

As noted earlier, organizations should be careful not to create Byzantine sensitivity classification levels that become too complex to implement effectively. Focus instead on three, or

at most five, categories of sensitivity for the classification work. Concentrate on putting only the most critical and sensitive data into the top protection category. Keep in mind that protecting highly sensitive data is often costly and requires expenditures for enforcement controls and management. Once data has been classified to a protection level, policies can be defined that indicate control requirements.

There are pros and cons to each type of technology based on use cases.

Locate/Classify

With a baseline of protection needs based on information classification categories, the next step is for the organization to locate the classified data. This is a non-trivial endeavor in today's highly distributed and duplicated world. Data location and management is the underlying tenet of Data Lifecycle Management (DLM). DLM follows data from original inception to (if appropriate) final deletion. During the lifecycle, the data has to be tracked and monitored for access, views, alteration and duplication.

Mapping access and duplication is a difficult process for new data but it can also be quite complicated for existing data. Again, take for example a customer data list. While the canonical list may reside on a central server, duplicates of this list probably exist in hundreds or thousands of places in the organization: e.g. on each sales person's laptop, on mobile smart device's contact lists, on thumb drives and in spreadsheets that are shared with partners. If, after the classification phase the information on this list is deemed "sensitive," an organization may find it difficult to apply proper policy enforcement to all iterations of the data.

Part of the process may entail deleting some of the instances of the data if the resident device is deemed outside of the scope of the governing enforcement policies.

Protect

With knowledge of the data's classification level and all of the instantiations of that data, a company is ready to move to the protection phase of the process. It is in this phase that data protection tools, such as rights management, encryption and leak protection can be applied.

There are pros and cons to each type of technology, based on use cases. While filtering and leak prevention technologies are effective during some phases of transit, they are not the best choice for protecting data at rest. Looking at data in transit, motion and rest, the technologies can be best applied as shown in the following table.

Phase	Protection Technology
Transit	Encryption Data and Information Leak Prevention/Filtering
Use	Rights Management
Rest	Encryption Access Control

All of the technologies listed can be important parts of an organization's comprehensive data-centric security strategy. For the purposes of this paper, we will focus the rest of the discussion on encryption and its benefits for protecting unstructured data.

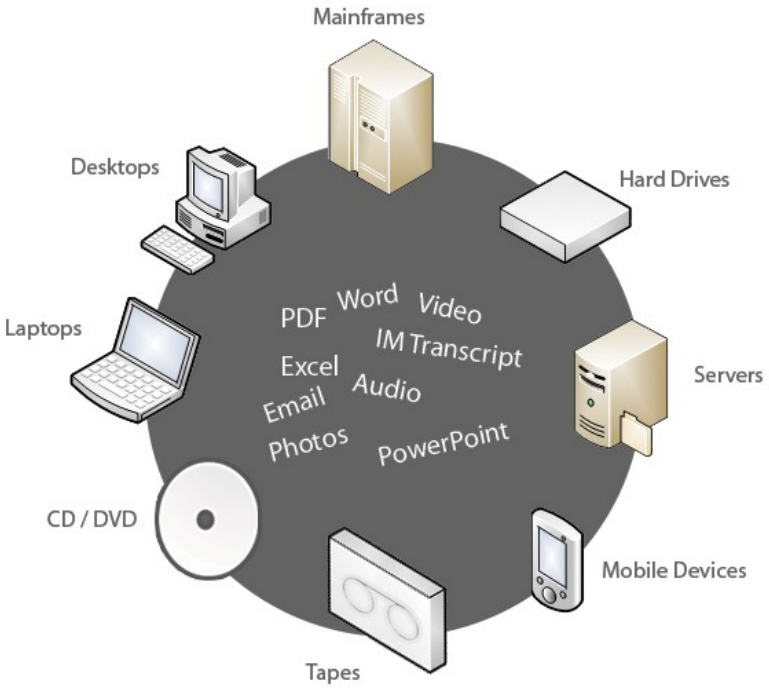
Controlling Unstructured Data

Well-planned data storage and management has structure. The data is categorized, classified and tagged with meta-data that makes it easy to locate and search efficiently. In a relational database, the data is stored in tables and columns that are defined by a schema governing the relationships, constraints and policies. In the "*classify, know, protect*" approach outlined above, structured data fits fairly easily into a data-lifecycle management (DLM) framework for management and control because it has already been categorized and tagged.

In most organizations, 70-90% of business data is in an unstructured or semi-structured state and recent research indicates that only 23% of organizations feel this data is properly protected. Unstructured data includes Word and Excel documents, images BLOBs (Binary Large Objects), not to mention the billions of emails and instant messages generated every day. Much of this is sensitive data, such as personally identifiable information (PII) and intellectual property (IP) that must be protected with appropriate measures. Some of the data is considered semi-structured because it has some meta-data information attached, such as a "Word Document" or an XML web-page, but this structure is fairly coarse. Protection must be applied using technologies that work with the meta-data available and able to support voluminous amounts of data. Some large organizations can generate terabytes of email in a few days or weeks.

In most organizations, 70-90% of business data is unstructured.

Another challenge of unstructured data-centric security is that the data must support multiple distribution needs. Data is shared cross-platform, from mainframes to UNIX-based web servers to portable handhelds, as well as cross-unit between entities and partners of the enterprise. Sometimes the success of the business depends on the speed with which the information can be quickly disseminated; excessive gating and controls can be seen to stand in the way of revenue generation.



Organizations know that the unstructured data has to be protected but that protection cannot get in the way of the business. It must be portable (cross-platform, user, business unit) and persistent. Protection of unstructured data in transit and by filter provides control from prying eyes as the data moves cross-boundary; however, the data is left vulnerable once it reaches its destination.

One limitation to encrypted tunnels is that the protection does not extend past the tunnel termination point.

Protecting Unstructured Data with Encryption

One approach to protecting unstructured data is to use encryption technologies. When the data is encrypted it is not viewable by unauthorized users. Encryption can be used to protect the data while it is being transported as well as when it is at rest.

Transit Only

Almost everyone has used some form of encryption in transit – SSL is in common use for secure transactions on the web and VPNs have been in use by businesses for over a decade. Use of an encrypted tunnel for transit protection is an effective way to shield sensitive and unstructured data when it travels over untrusted networks such as wireless LANs and the Internet.

Encrypted tunnels work on a variety of networks and devices and can be used along with strong authentication and access controls for added security. One limitation to encrypted tunnels for data protection is that the protection does not extend past the tunnel termination point. Once the data has reached its destination, if it is not otherwise protected, it is exposed to the same vulnerabilities that the target device is exposed to.

Rest Only

Some encryption can be used exclusively while the data is at rest; this includes native database encryption that encrypts data before it is stored and full-disk encryption that encrypts the entire contents of a hard drive. These forms of encryption at rest work well with unstructured data especially on a hard drive where everything is encrypted regardless of form or classification.

Companies that wish to use this type of encryption must ensure that their databases support native encryption. For full-disk encryption, there are a number of options for a variety of devices including laptops, servers and handhelds. Drawbacks to this approach are performance degradation if systems, especially databases, are not sized properly for the additional overhead requirements. Also, rest only encryption does not protect the unstructured data in transit or in use, so other methods of protection will be required for those use cases.

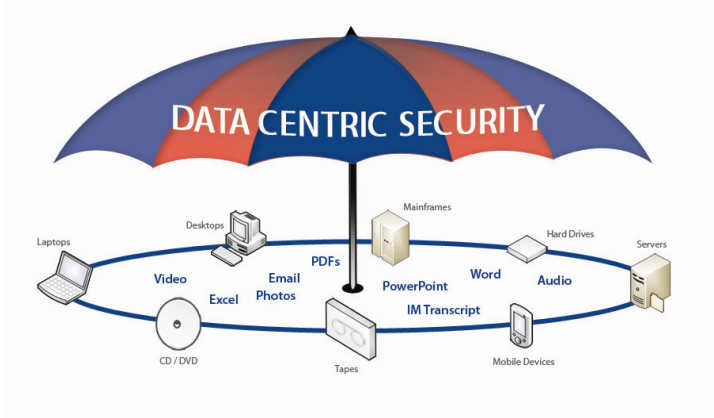
Transit and Rest

File and data encryption can provide protection both in transit and at rest. The unstructured data that requires protection is encrypted before it is transferred or stored. As with full-disk encryption, device support is required, but there are file encryption programs that work on a wide variety of platforms and devices.

Performance is still a concern, but is less invasive because only sensitive files are encrypted rather than the entire hard drive. In cases where decryption is not supported by a target, the protection is persistent because the data itself is still encrypted and unreadable. One drawback to this approach is that data is not protected while it is in use by an authorized user.

Selection Criteria for File/Data Encryption

For organizations seeking effective file/data encryption solutions to protect unstructured data at rest and in transit, there are a number of important criteria to review before making a purchase. Specific requirements will vary by a company's unique needs and priorities. However, this guide distills the most common requirements that companies assess when selecting a file/data encryption solution. Requirements are broken down into two main categories. Business requirements address needs of the organization independent of the underlying technology. Technical requirements list the architectural and technical specifics that a solution must meet. Taken together, the requirements create a protection picture for an enterprise and serve to focus the selection process on solutions that will provide best-fit protection for the organization's unstructured data.



The requirements create a protection picture for the enterprise and focus the selection process on best-fit solutions.

Business Requirements

Though every business is unique, there are some recurring patterns of need in the assessment process. Most companies find that most, or all, of the following must be considered when selecting a file/data encryption solution.

Compliance

Many of the disclosure laws, both in the United States and internationally, require notification if data has been lost or *may* have been lost. This includes data exported from a database or resident on a stolen laptop. Many of the laws, however, do exempt the notification requirement if the data was encrypted at the time it was lost.

This exemption is a core driver behind file encryption purchases. Review the disclosure laws to determine if a specific type of encryption algorithm is required to ensure that the proposed solution will meet the criterion for exemption.

Other compliance mandates also call for data encryption. The Payment Card Industry Data Security Standard (PCI DSS) calls for encryption of credit card data in transit, at rest and when sent via email attachments. Going forward, as data-centric security becomes an increased focus of compliance work, new and emerging mandates may call for encryption of key data in transit and at rest. Policy managers that can implement encryption requirements from a central policy store will be prepared for these emerging requirements.

Time to Deploy

Comprehensive data-centric strategies can take months, and even years, to deploy. But companies often need to act quickly to deploy tactical, critical protection of their most sensitive data. Assess the time involved with a solution to determine how quickly it can be deployed.

What is the acceptable time for prototyping and roll-out? Is it two years, two months or two weeks? Does the vendor have proof points validating that comparable organizations can deploy in the target time-frame? Two weeks may sound like an unrealistic time-frame for some, but with the right parameters and scope, it is possible. Low-touch solutions that do not require extensive overhauls to the infrastructure or business operating procedures can be implemented rapidly if the project is managed well and the scope has been clearly defined in advance. Another aspect of deployment is the amount of user education required. If a solution is familiar to users or use is transparent, it will be up and running more quickly. File encryption tools that are intuitive to use will be adopted more easily and require less training time, both for end-users and administrators, than complicated ones. Review the interface and management tools before deployment to determine how quickly the solution can reasonably be rolled-out.

Companies often need to act quickly to deploy tactical protection of their most sensitive data.

In some cases, file and data encryption can be rolled-out to very large organizations in as little as two weeks. Vendors under consideration should be able to provide verifiable reference customers that can validate the success of a fast roll-out.

Business Process Alignment

How well a solution fits into the business can be a key indicator of success. Solutions that require business process reorganization can get bogged down in lengthy implementation cycles and face adoption hurdles.

Assess how the business uses files and data within the company as well as with business partners. Generate common use cases for how the data will be shared and accessed. Then look for solutions that support these business processes. For example, if users send sensitive information to a wide variety of partner companies that require a long approval process by multiple stakeholders, the need for key distribution may interrupt the normal business process. In this case, the best fit may be a solution that can be rapidly extended to new partners as needed.

User Experience

Not everyone is a technology guru. Most users concentrate on getting their work done, not on the underlying technology powering that work. And when security solutions are deemed too difficult to use, many users will circumvent the solution as well as the security. So take into account the comfort level of users when selecting a solution.

Keep in mind that there may be different user classes, such as internal users, customers, partners and consultants, all with different levels of technical expertise. If a solution is to be used by a very broad group, such as the corporation's entire customer base, look for an interface that is familiar and user-friendly.

Interviewing peers at other companies can provide insight into usability in production. Does the vendor have references that are willing to discuss their experiences? How many customers does the vendor have currently using the solution in a production environment? Real-world implementation references and size of user base are strong indicators of usability.

Another way to make the user experience more palatable is to look for solutions that can provide transparent security. For example, look for a solution that works inside of an application or with an email server to encrypt the file/data automatically with no user intervention required.

Cost

Though some organizations place few limits on security spending, most have to take into account budgetary constraints. When selecting a solution, be sure to include not just the initial cost for the solution but also the ongoing maintenance and any additional headcount for management.

One hidden cost is that of help-desk personnel. Security solutions that lock users out of their files or data can result in a dramatic spike of costly support calls, not to mention angry users who can't get their jobs done. This cost is linked with the user experience requirement – if a solution is easy to use it is less likely help-desk calls will increase.

Technical Requirements

Even more than business requirements, technical requirements vary by organizational needs and constraints. Use this list as a starting point for generating your own specific list of assessment points.

Availability

Files live on a number of devices and, for most companies, an effective file encryption solution must work on all of them. After the locate/know phase of a data-centric process, companies should have a list of the devices, platforms and operating systems that have to be supported for storage of sensitive data. Use that list as a baseline for assessment of file encryption products. If a key platform isn't supported, create a compensating control or simply move on to the next solution provider to find one who supports everything your company needs.

At a minimum, most companies will need a solution that works consistently across multiple operating systems.

At a minimum, most enterprises will need a file encryption solution that works consistently across multiple operating systems including mainframes, midranges, UNIX/Linux/Windows servers and Windows desktops.

Architecture

Organizations have complex IT architectures and long-term plans for growth via the corporate reference architecture. Security solutions must work within the existing architecture and be sufficiently robust to grow as the organization's reference architecture does. Security solutions that call for massive restructuring of the architecture are often delayed or rejected as being too intrusive.

For this reason, many companies look for low-touch file encryption solutions that work with the current and planned architecture with no infrastructure changes required. Look for end-point solutions that can be installed on an as-needed basis on the key devices and servers and work with existing security stores such as Microsoft Active Directory and Sun Directory Server. Also look for gateways that can be placed in the current environment without requiring configuration changes to security devices such as firewalls and access control servers.

Extensibility/Interoperability

Before purchasing a file/data encryption solution, determine the flexibility of the product architecture and how extensible it will be going forward. Specifically, understand the technical requirements for partners and others that integrate with your own architecture. Also, assess which applications and new initiatives, such as SOA, will need to be supported.

Transparent integration is a requirement for many companies. Available APIs into a proposed solution will enable application developers to write transparent, policy-based encryption

capabilities directly into new and existing applications.

For users, inventory which applications are in highest use. Most companies use Microsoft Office with Outlook for the bulk of their desktop work. A solution that integrates with these products transparently will be easier to deploy quickly with minimum disruption.

Conversely, a product that only works with a limited number of file formats may not be an effective solution for many organizations. Determine the file formats that need to be part of the unstructured data protection solution. If a solution is limited to certain extension types, (for example, .docs can be protected but .tiffs can't) the long term extensibility and usefulness of the product will be negatively impacted.

Another key integration point to consider is PKI interoperability. What certificate stores are in use for key storage? For ease of integration, look for solutions that work with existing directory stores such as Active Directory, eDirectory and Directory Server.

File/data encryption provides protection that stays with unstructured data as it is replicated across multiple devices, platforms and zones.

Certifications

Encryption products use encryption modules to store and manage algorithms. How well the product manages the security services depends on robustness of implementation. To help organizations assess the reliability and functions of cryptographic modules and algorithms, the U.S. National Institute of Standards and Technology (NIST), in cooperation with the Communications Security Establishment Canada (CSEC), runs validation programs that certify solutions to the U.S. Government's Federal Information Processing Standards (FIPS) 140-2 program.

Does your organization require FIPS 140-2 certification for encryption solutions? If so, look for products that have been certified or can be implemented using certified modules. For more information on NIST's FIPS 140-2 cryptographic module validation program, please visit http://csrc.nist.gov/groups/STM/cmvp/index.html.

Scalability

Successful companies face rapidly increasing user populations as partnerships expand, acquisitions are made and customer bases grow. The advent of highly virtualized data centers has resulted in a mushrooming effect on server installations and new applications are being brought online to support Web 2.0 and SOA initiatives.

File/data encryption solutions that are installed today will have to scale to this growth in order to be effective. Assess solutions based on how well they will scale to new users and new servers and services. How many files or documents per day is the product handling in production environments? Is it 500 or 5,000,000? And what is the size of those files? Also determine how easily the solution can handle the sheer growth of files and data – is the solution dependent on a central processing point that could suffer from slowdowns and bottlenecks? Or is the solution based on distributed processing power and as scalable as the rest of the infrastructure? Before installation, make sure the solution selected is designed to grow along with the company's data/file encryption needs.

Conclusion

A comprehensive data-centric security strategy is a long-term, complex undertaking. There are no short-cuts and no single product will address every data-centric security need. File/data encryption provides protection that stays with unstructured data as it is replicated across multiple devices, platforms and zones. Look for products that meet business and technical requirements and can be implemented rapidly and grow along with the company's long-term data-centric solution strategy.

SecurityCurve

Stay ahead of the curve

diana@securitycurve.com
www.securitycurve.com

Security Curve gives companies the market and technology insight they need to make agile business moves so they can stay ahead of the security curve. Our clients benefit from targeted intelligence, comprehensive research, and focused solutions to stay ahead of the competition in a rapidly changing market.

Diana Kelley, Partner

Diana Kelley has extensive experience delivering strategic, competitive knowledge to large corporations and security software vendors. She was Vice President and Service Director for the Security and Risk Management Strategies (SRMS) service at Burton Group, the Executive Security Advisor for CA's eTrust Business Unit, and a Manager in KPMG's Financial Services Consulting organization.

References

[i] "Frequently Asked Questions," http://www.tjx.com/tjx_faq.html

[ii] Pay By Touch Outbids CyberSource to Acquire CardSystems," http://www.greensheet.com/PriorIssues-/051101-/3.htm

[iii] "Unstructured data at risk in most firms, survey finds: Ponemon study shows dearth of corporate data ownership rules, user monitoring policies," Brian Fonseca, ComputerWorld, July 1, 2008