Securing Unstructured Data

In classic terms, *structured data* is data that conforms to some sort of strict data model and is confined by that model. For most IT and security professionals, structured data is the information that lives in the database and is organized based on the database schema and associated database rules. The data itself is structured in a manner that typically allows for easy classification of the data. There is tight control over the access of structured data. Security controls are relatively easy to define and apply to structured data using either the built-in features of the structure or third-party tools designed for the specific structure.

By contrast, *unstructured data* is much more difficult to manage and secure. Unstructured data can live anywhere, in any format, and on any device, and can move across any network. Unstructured data has no strict format.

Securing information when stored as structured data is relatively straightforward. But it can be very difficult for unstructured data. Many analysts say 80 percent or more of digital information in an organization is unstructured, and that the amount of unstructured data is growing at a rate 10 to 20 times the rate of structured data.

Different States of Unstructured Data

Unstructured data changes are constantly occurring. Such data can be in one of three states at any given time. They are

- At rest
- In transit
- In use

If the data is at rest, it is sitting quietly on a storage device. It can also be in transit (sometimes referred to as “in flight”), which means it is being copied from one location to another. Or, it can be in use, in which case the data is actively open in some application.

Approaches for Securing Unstructured Data

- **Databases**: The database is the center of the data world. With new developments in database technology, increasing amounts of unstructured data are now stored in the database.

  The most common approach to securing the data in a database is encryption. Encryption of data that resides in a database can be approached in various ways:

  - Encryption of the actual data itself such that it is stored in normal data files in an encrypted state. The database doesn't necessarily know whether or how the data is encrypted, so it passes the encrypted data to the application to decrypt.
  - Partial encryption of the database schema so that specific rows, columns, or records are encrypted as a function of the storage of the data. In this case the database handles the encryption of data and performs the decryption to the application.
  - Full encryption of the database data files such that any information that resides in them is encrypted.

  It is important to control who or what can connect to the database and perform queries. Database access controls play a key role in restricting access to data. The approaches that are used by different databases vary, from authentication with a simple username and password to gain access to a database schema, to a complex set of rules that define for various levels of data classification who can access what, from where, at what time, and using what application.

  Many databases provide functionality for the mass export of data into other databases. This presents security challenges. From an unstructured data perspective, this can be a significant problem, because the information that resides in database files can be easily shipped from location to location. Encryption can be applied at the export phase. This usually is a different mechanism from that used for the encryption of data in the schema or the encryption applied to system-wide database backups.

- **Applications**: Unstructured data is typically created in either of two ways: through user activity on their workstations, or as applications access and manipulate structured data and reformat it into a document, e-mail, or image.

  Securing applications is one of the most important ways to protect data, because applications are the interface between the end user and the data. As a result, a great deal of the security investment is devoted to the development of the application.

  Application security can be categorized into the following groups:
  - Application access controls that ensure an identity is authenticated and authorized to view the protected data, to which that identity is authorized, via the application
  - Network and session security to ensure the connection between the database, application, and user is secure
  - Auditing and logging of activity to provide reporting of valid and invalid application activity
  - Application code and configuration management that ensure code and changes to the application configuration are secure

- **Networks**: Data moves from the protected realm of the database into the application and on to the end user. Network security technologies have developed into complex systems that are able to analyze traffic and detect threats. Network intrusion prevention systems actively monitor the network for malicious activity and, upon detection, prevent intrusion into the network. Malware protection technologies prevent Trojans from deploying and planting back doors on your trusted network clients.

- **Computers**: Once a legitimate user has securely connected across the network to the application to access data residing in the database, the information is ultimately presented in a web page that is rendered by a web browser. The security on the computer from which the user interacts with the application and resulting unstructured content becomes critical.

- It is important to ensure that only legitimate users can access the computer.
- Once a user is authenticated and provided with a session, they are able to proceed to access the network, read data on the local hard disk, and connect USB and other storage devices and transfer data to and from the computer. Hence, Controlling the flow of information over network interfaces and other information connection points is required.
- Securing data residing at rest on the computer is immensely paramount. Henceforth, security for storage devices is essential.

- **Storage**: Storage is one of the most effective areas of unstructured data security and is often the one which receives the most attention after a data loss incident. Storage security solutions mainly deal with data at rest, but sometimes stored data exists in another state, either in transit or in use. Typically, storage security solutions focus on encryption or access control. Disk encryption and File System encryption can be helpful in regards to storage security through encryption.

- **Physical World**: A lot of time and effort is devoted to finding solutions for securing unstructured data in the digital world; however, data loss incidents due to the loss or inappropriate disposal of paper documents must also be considered—that is, finding solutions for how to secure information in the hardcopy world.

  Before being printed, all confidential documents should have any non-essential contents hidden or deleted, so unnecessary confidential information is not included in the printout. All printed confidential documents should have a front cover page and the pages should be numbered. Each copy of an entire document should be labeled. Confidential documents should be printed to a private printer, whenever possible. All paper documents containing confidential information should be locked in a secured container. When no longer needed, documents should be immediately shredded or placed in a secure container for a shredding service to destroy.