help source

Do you have something to say about this article?
Visit the *Journal* pages of the ISACA® website *(www.isaca.org/journal)*, find the article and click on the Comments link to share your thoughts.

http://bit.ly/2ySgksA

Q We are in the process of selecting a data loss prevention (DLP) tool. After discussing it with vendors, we realized that successful implementation of DLP depends on classifying data to identify key words that enable the DLP to recognize data to be protected. The challenge is we have a huge amount of data that are scattered all over. Therefore, we are still struggling with the right approach that will help us classify the data. How should we approach this problem?

A Data or information is a primary enabler for any organization, as established in COBIT® 5. Organizations today generate, process, use and store volumes of data/information. Many organizations face similar problems when classifying data/information. Although there is no panacea to this problem, it can be addressed based on the approaches used by various organizations.

ISACA's *Data Leak Prevention*[1] white paper identifies three key objectives for a DLP solution:

- Locate and catalog sensitive information stored throughout the enterprise. (Data classification)
- Monitor and control the movement of sensitive information across enterprise networks. (Network-level controls)
- Monitor and control the movement of sensitive information on end-user systems. (End-user controls)

The white paper provides guidelines for implementing DLP. These guidelines are:

- Data classification should be the first step of the program.
- Define and implement data classification and protection policies.
- Implement and configure DLP solutions per policy.
- Identify and monitor the risk associated with limitations of DLP solutions in protecting the organization's data.

The major objective of DLP is to prevent secret and confidential information from reaching unintended recipients. Organizations expect that DLP should be able to detect whenever such secret or confidential information is transmitted beyond the boundaries of the organization. An effective DLP implementation requires careful planning and cultural change, which are not possible without identifying and classifying the organization's data and information.

One more point also needs to be considered: implementing only DLP solutions may not provide the required level of assurance on the protection of data. It may have to be supplemented with implementing and integrating digital rights management (DRM) and access management solutions.

Other aspects to consider while implementing a DLP solution include:

- Generally, regulatory requirements mean data leaks can be catastrophic for organizations, and the possibility of liability and litigation are main drivers for organizations to consider DLP technologies.
- Many times, DLP is deployed by organizations with a focus on protecting intellectual property rights and trade secrets only.
- DLP and digital rights managements (DRM) implementation should be considered as an organizational program rather than as an IT initiative.
- Such programs may have multiple projects/phases and may require one to three years to fully implement depending on the size of the organization.
- DLP can protect data/information within the organization's perimeter, but cannot be extended beyond boundaries such as DRM.
- Data classification forms the foundation for DLP to be successful.
- DLP is not an adequate protection in cases where the organization uses cloud technologies.

Sunil Bakshi, CISA, CRISC, CISM, CGEIT, ABCI, AMIIB, BS 25999 LI, CEH, CISSP, ISO 27001 LA, MCA, PMP
Has worked in IT, IT governance, IS audit, information security and IT risk management. He has 40 years of experience in various positions in different industries. Currently, he is a freelance consultant and visiting faculty member at the National Institute of Bank Management, India.

At this point, the focus is on the first step: the classification of data.

Data classification best practices suggest the following steps:

1. Define a classification scheme in which the data/information within the organization is identified and classified in predefined buckets (e.g., top secret, confidential, sensitive, internal, public). Organizations may adopt a different scheme depending upon the nature of their data/information.
2. Identify the organization's data that are in soft form (digital or electronic) and hard form (physical documents). Also, note that there is a great deal of data/information in the heads of employees who deal with data while carrying out their responsibilities.
3. Define a data classification and protection policy that will apply across the organization. The policy should address the privacy policy and related compliance.
4. Determine the method to classify the data. The best approach for this is to use a risk management framework to help in determining the nature of the data.
5. Classify and label the data.
6. Implement controls for protection.

Organizations face major challenges while executing the fifth step, primarily due to:

- The volume of data generated, processed and stored
- Multiple data owners and coordination among them
- Cross-functional dependency and accesses required by such teams
- Classifying and labeling historical data

The following suggestions may be considered while executing the data classification process:

- Educate business process owners about data classification and the protection policy, including the privacy policy.
- Ask business process owners to identify data elements and the source of the data. This will help in identifying data owners/custodians. For example, employee data generated and owned by the human resources (HR) function, but used by other departments, must be classified by HR, and others must use that classification.
- Form small data sets that make meaningful information from data elements and classify them. For example, employee number, name, date of birth, address and date of hire can form a data set that is generally used by other functions such as payroll and physical security. Many independent data elements cannot be classified, with a few exceptions (e.g., credit card numbers).
- Identify the data sets (partial or complete) used in the report/document when classifying such information and determine the classification level of the report/information based on the classification of the data sets. Most information or reports generally contain multiple data sets. Generally, the highest level shall prevail. For example, employee personal information is confidential; therefore, the payslip of the employee is automatically classified as confidential.
- Determine and document exceptions.
- Maintain a function-focused and centralized data set inventory that validates the data's classification.
- Implement a process for periodic review.
- Implement an ongoing classification process.

Once the classification process is underway, further steps to optimize security may be considered. Labels used to classify data can be used as key words while implementing DLP solutions.

A last point to be noted is that though DLP solutions significantly improve an organization's ability to manage risk associated with data leaks, implementation of these solutions is complex and prone to errors and mistakes that may hamper achieving objectives. Careful planning and preparation, communication and training are essential for successful DLP programs.

Endnotes

1 ISACA®, *Data Leak Prevention*, USA, 2010, *www.isaca.org/Knowledge-Center/Research/ResearchDeliverables/Pages/Data-Leak-Prevention.aspx*