DATA LOSS PREVENTION: A HOLISTIC APPROACH

Introduction

Data breach has been one of the biggest fears that organizations face today. While DLP is not a panacea to such attacks, it should certainly be in the arsenal of tools to defend against such risks. The term DLP, which stands for Data Loss Prevention, first hit the market in 2006 and gained some popularity in early part of 2007. DLP is not a plug-and-play solution. The successful implementation of this technology requires significant preparation and diligent ongoing maintenance. While a great deal of attention has been given to protecting companies' electronic assets from outside threats – from intrusion prevention systems to firewalls to vulnerability management – organizations must now turn their attention to an equally dangerous situation: the problem of data loss from the inside. There is gaping hole in many Organizations which is the ubiquitous way businesses and individuals communicate with each other–over the Internet.

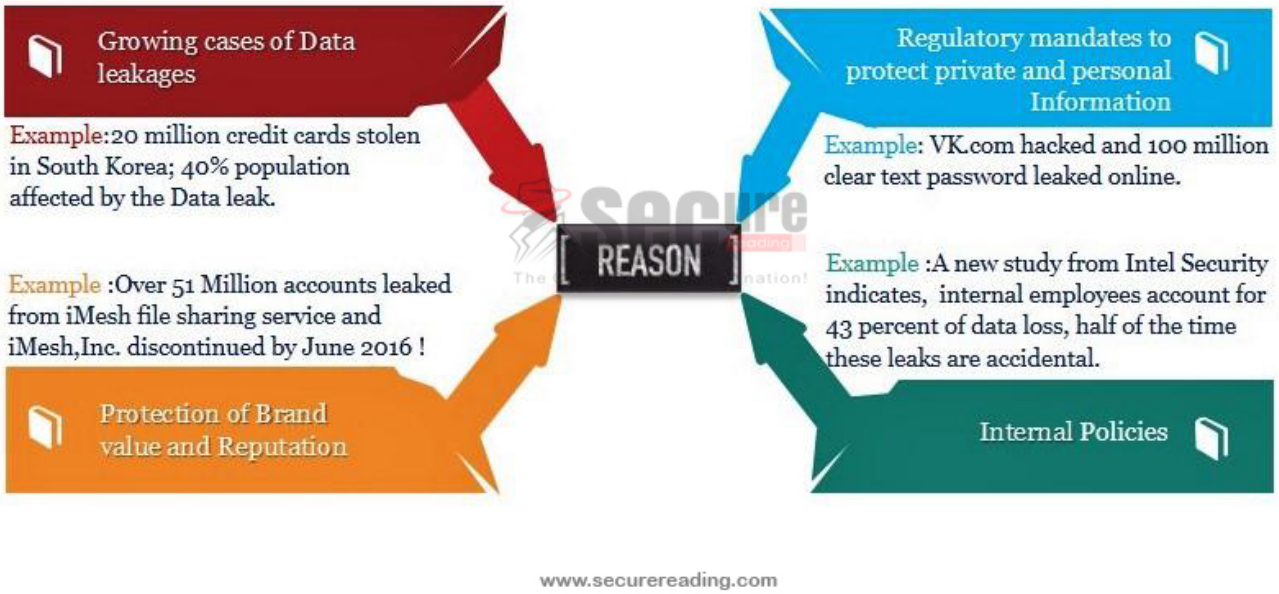
Given today's strict regulatory and ultra-competitive environment, data loss prevention (DLP) is one of the most critical issues facing CIOs, CSOs and CISOs. For those creating and implementing a DLP strategy, the task can seem daunting. Fortunately, effective technical solutions are available. This paper presents best practices that organizations can leverage as they seek solutions for preventing leaks, enforcing compliance, and protecting the company's brand value and reputation.

What is DLP?

Data loss prevention (DLP) is a solution for identifying, monitoring and protecting sensitive data or information in an organization according to policies. Organizations can have varied policies, but typically they tend to focus on preventing sensitive data from leaking out of the organization and identifying people or places that should not have access to certain data or information.

Importance of DLP

Following are the major reasons that make an organization think about deploying DLP solutions:

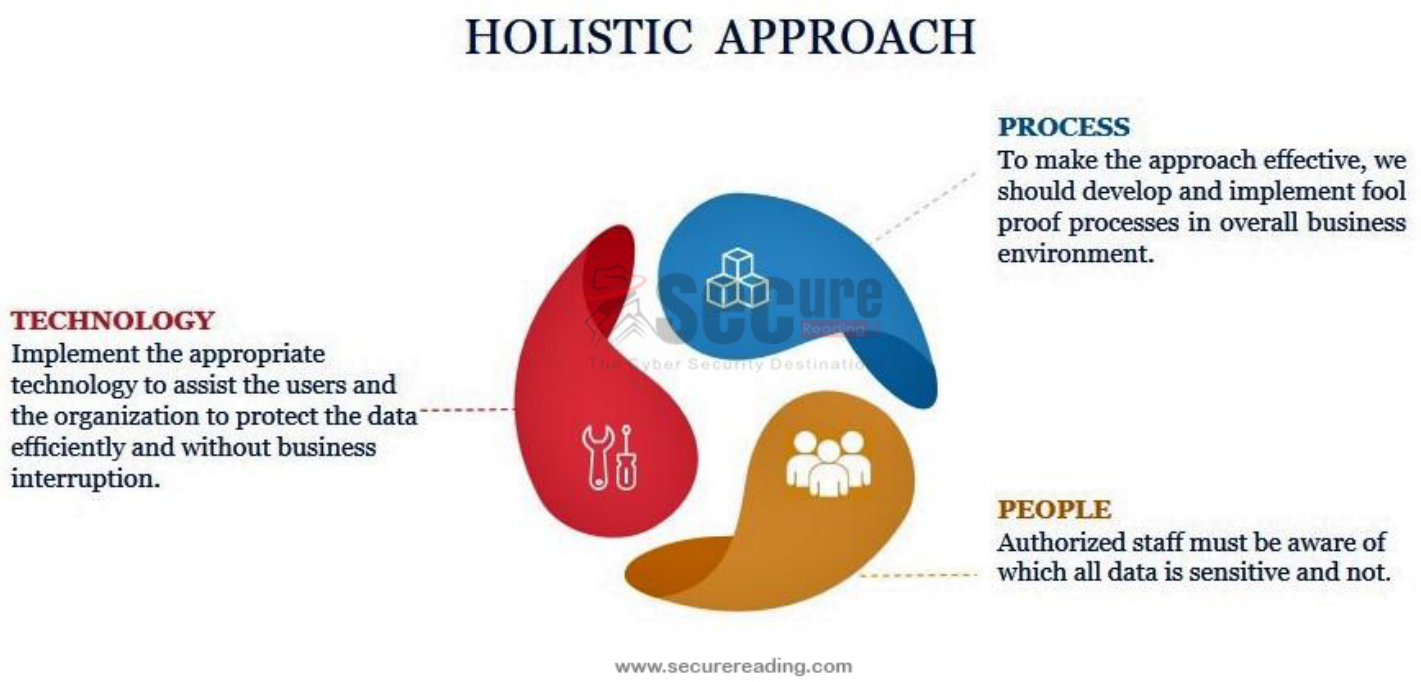


Until a few years ago, organizations thought of data/information security only in terms of protecting their network from hackers. But with growing amount of data, rapid growth in the sizes of organizations , rise in number of data points and easier modes of communication accidental or even deliberate leakage of data from within the organization has become a painful reality. This has lead to the growing awareness about information security in general and about outbound content management in particular.

How is DLP different from any other security technology?

While tools such as firewalls and IDS/IPS look for anything that can pose a threat to an organization, DLP is interested in identifying sensitive data. It looks for content that is critical to an organization. While DLP can prevent data breaches from Intruders, more often than not this solution is used as a mechanism for discovering broken processes in the normal course of business. We know for a fact that majority of all malware outbreaks companies suffer are due to unwitting user actions. This trend has not changed much even with the ongoing user awareness training. There have been cases of data loss, where employees were part of such act at will. For example, an American Multinational Corporation Morgan Stanley has a new kind of data breach: an old employee named Galen Marsh, who was recently promoted as financial advisor, stole account information from up to 10% of its total wealth management clients, including account names and numbers. How could this incident have been prevented? Proper implementation of DLP would have marked this data as sensitive and rated it a high criticality.

The most effective approach for data leakage prevention is by addressing it through people, process and technology.



Two Technical Approaches to DLP:

DLP technology is based upon content-level inspection which is fundamental to the DLP overlay and network-based approaches presented here.

The DLP Overlay Approach

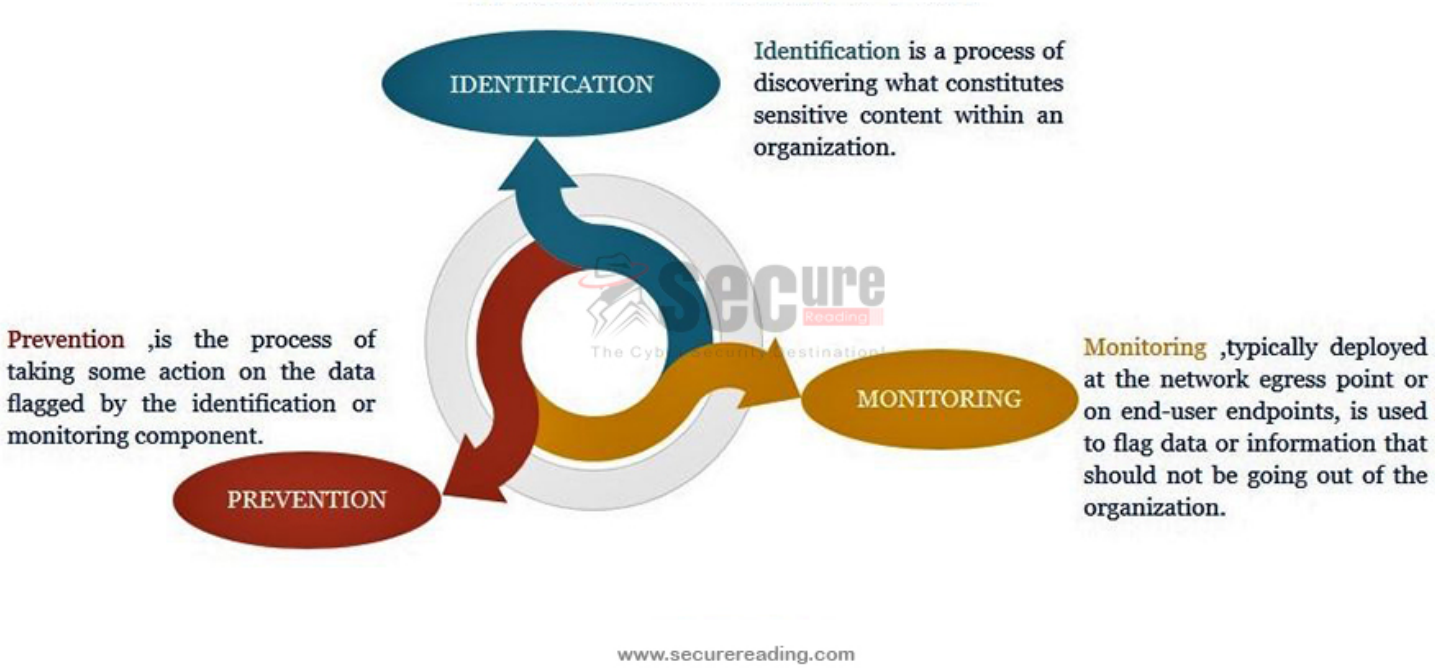
The DLP overlay is based upon IT identifying content it needs to monitor and the DLP overlay does so at every point in the IT infrastructure to prevent data loss. DLP overlay solutions provide large amounts of information concerning how data is used and is thus effective at protecting against accidental data loss. But DLP overlays have to be used in conjunction with other data security technology to protect against all types of data loss such as accidental, negligent, data theft, identity theft, etc.

The Network-Based DLP Approach

McAfee, Symantec and others believe that DLP is a separate security system while others such as Cisco believe that data loss is best mitigated by understanding what data needs to be protected, and then leveraging the network to prevent data loss as the network touches every IT asset. The network-based DLP approach is an efficient and reasonable way to achieve data loss prevention. The network approach to DLP allows IT leaders to measure risk by identifying its most valuable data and then creating the right strategy to prevent data loss. In addition data security policy is augmented while providing content monitoring and inspection over high-risk channels in the network. This affords a broad approach to DLP as every corporation has unique data loss vulnerabilities it needs to mitigate.

DLP COMPONENTS

The core DLP process can be broadly classified into three components: Identification, Monitoring and Prevention.



From a data loss perspective, the industry has adopted three standard terms related to the states in the data lifecycle:

• **Data at rest** is data that is stored within the IT infrastructure and on media. Common components containing data at rest are servers, databases, file shares, intranet sites, workstations, laptops, mobile devices, portable storage, backup tapes, and removable media. Data at rest can also be stored externally with third parties or through external extensions of the IT infrastructure, such as cloud storage.

• **Data in motion** (Network) is data that is in transit, flowing across internal networks and to the outside world (i.e., data on the wire and in the air).

• **Data in use** (end-point) is data that is being accessed or used by a system at a point in time. Examples include data in temporary memory on a local machine, an open report or running query on a workstation, an email that has been drafted but not sent, a file being copied to a USB drive, and data being copied and pasted from one local document to another.

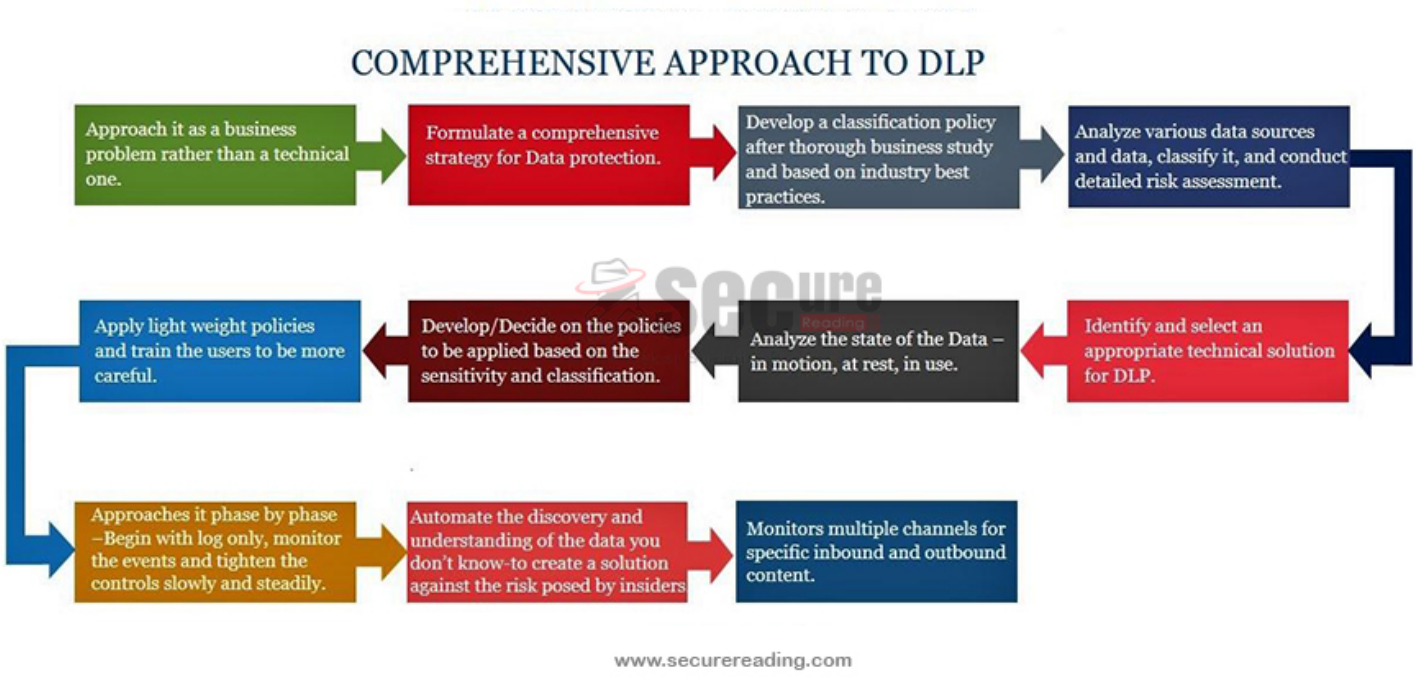
How to identify sensitive data?

An effective DLP program requires an understanding of following questions:

DLP is shipped with hundreds of predefined policies. In addition, vendors are even willing to create a custom policy based on customer requirements. This is based on the business model of a particular customer. By closely working with the vendor, default policies can be fine-tuned to suit your needs. One of the key challenges to securing your critical data is the fact that there are so many ways for it to leave. In developing your DLP strategy, a holistic view should be taken to ensure that the combination of controls employed is geared to protect the most sensitive data that the organization holds.

How can we protect those sensitive data?

Here gives a comprehensive approach to prevent data leakage.



DLP Technology

Generally speaking, there are two levels of DLP technologies: Full Suite and Channel Data Loss Prevention. Full Suite DLP technologies are focused exclusively on the task of preventing sensitive data loss, while Channel DLP solutions make DLP a single feature among a long list of non-DLP functions.

- **Full Suite DLP**

Coverage: Most Full Suite DLP solutions were developed with the idea of data loss prevention in mind and include comprehensive coverage for the greatest effectiveness. These solutions provide coverage across the complete spectrum of leakage vectors, namely, data moving through the network gateway or data in motion, stored data on servers and workstations or data at rest, and data at the workstation/endpoint level or data in use. Equally as important, Full Suite DLP solutions address the full range of network protocols, including email, HTTP, HTTPS, FTP and other non-specific TCP traffic.

Detection Methodologies: Another critical distinction of most Full Suite DLP solutions is in the depth and breadth of sensitive data detection methodologies. The earliest DLP technologies relied exclusively on pattern matching on text strings, looking for patterns that matched account numbers or a dictionary of words. These early detection methodologies can detect very specific patterns, but often result in a high number of false positives as well. Over time, a number of new detection methodologies have been introduced that have drastically improved the effectiveness of DLP solutions.

One critical detection methodology, data fingerprinting, is now common across leading full suite DLP vendors. The fingerprinting process can be used on databases (structured data) and files or documents (unstructured data) by initially creating and storing a one-way hash on the DLP system. The DLP solution then analyzes content, compares it with the stored hashes and returns an incident if there is a match. This methodology can be used to accurately identify sensitive database content, such as a last name and account number as well as exact or partial matches of documents.

Central Management Console: Another unique feature of Full Suite DLP solutions DLP solutions is a central management console for configuring coverage across data in motion, at rest and data in use, creating and managing policies, reporting and incident workflow. This sidesteps the need for different management interfaces for each component of DLP, significantly reducing the management overhead of a comprehensive DLP initiative.

- **Channel DLP**

Most Channel DLP solutions were designed for some other function besides DLP and were modified in order to take advantage of the DLP visibility by providing some limited DLP functionality. Some common Channel DLP solutions include email security solutions, device control software and secure web gateways. In each case, Channel DLP solutions are limited both in their coverage and detection methodologies. For example, a number of email security vendors – both on-premise and cloud-based – have the capability to scan email content for sensitive data. In most cases, detection methodologies are limited to pattern matching across email. Among other widely-used protocols, such as HTTP, HTTPS and FTP, content is not inspected in any way.

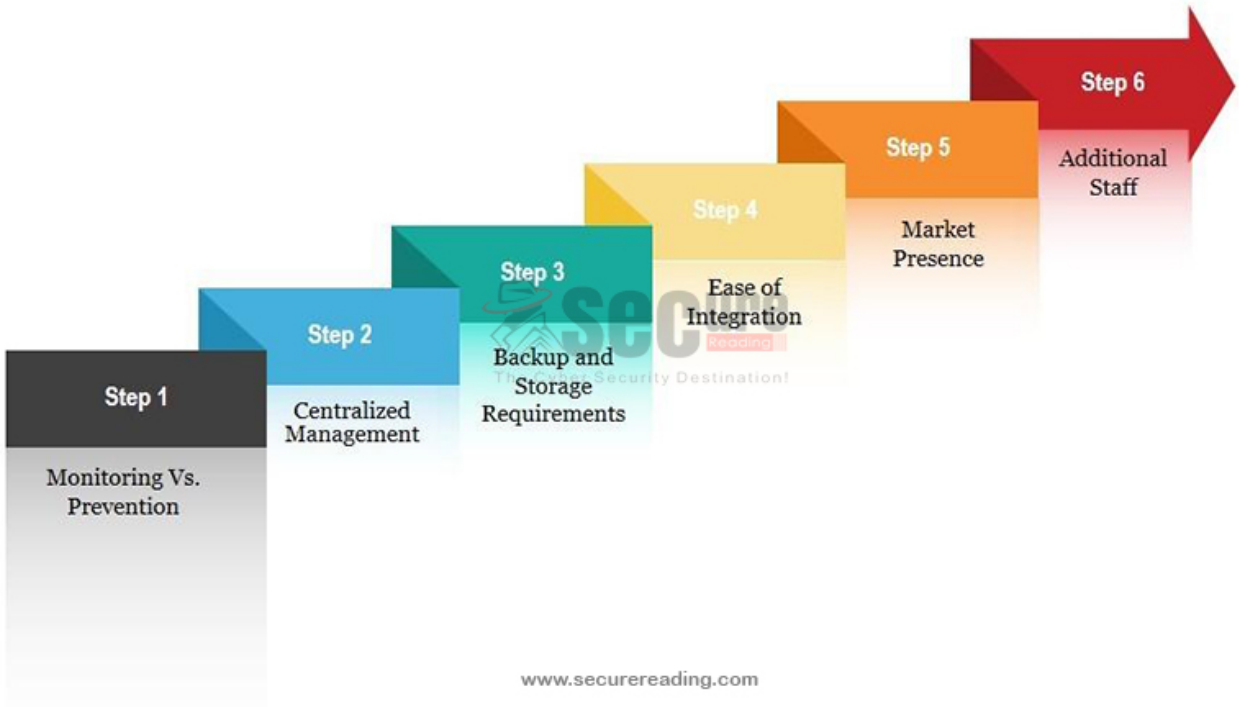
How to choose a vendor?

There are several write-ups factoring in a variety of elements in choosing a vendor. After all we are after finding a DLP solution that will meet the business needs as best as possible. We will point out the key steps that customers should look into.

- **Monitoring Vs. Prevention**

These two features may be represented by Vendors with complex and fancy names. Though taking a deeper look into the solution might reveal a few unique features to each vendor, at a higher level they simply refer to DLP functioning in monitoring mode and prevention mode. A good analogy to the discussion of

whether or not content protection technology should run in monitoring or preventive mode is the comparison between intrusion detection systems (IDS) and intrusion prevention systems (IPS). When IPS was first introduced, there was a misconception of this technology that it will be able to block most of the attacks and the false positives will simply disappear. Customers did not know at the time that only a handful of signatures could go in block mode and a thorough study of the environment was critical to extend the blocking. Yet, practically there was no significant increase in blocking. Same rule applies to DLP as well. The accuracy of a signature is very critical before deciding to quarantine or block a certain activity. Moreover, DLP requires additional hardware and software in order to enforce prevention. To quote an example, if we choose to block an email containing sensitive data, some vendors require the integration with an enterprise class MTA, such as Ironport, Sendmail, Proofpoint, etc. Settling for prevention mode can be very costly, especially if you do intend to block multiple channels. In addition to the cost, the ease of integration should be factored in as well. It is important that all future goals be included in the scope. That way if prevention mode is in scope, an organization is better informed of the additional software and hardware requirements that will be needed to enforce blocking effectively. Keep in mind that some of the technologies that we need for blocking might already be in place in your environment. This might take off some of the financial burden.



- **Centralized Management**

Maintenance overhead is every organization's nightmare. Centralized management can reduce a lot of overhead. Some of the key features to include are policy creation and enforcement, reporting, and data filters.

- **Backup and Storage Requirements**

Each organization has a set of requirements for data storage. While most DLP vendors are software based, there are some that are appliance based. The product arrives in a hardware appliance and has the capability to retain data for significant length of time. If the data retention policy states that data must be kept for six months, some appliance based products are built to handle terabytes of data. This can be a good solution for organizations on a tighter budget. Reconnex is an example of a hardware based solution.

- **Ease of Integration**

Few elements can play a significant role in ease of integration. Vendors do not always have the solution in hand to meet a customer's requirements. Several complex issues will come into light only while the implementation takes place. One of the issues I have run into is an agent less approach for data discovery feature. All operating platforms that will be part of the scanning should be taken into consideration. In some cases, the scanning feature was agent less for windows based systems, however required an agent to be installed on AIX OS. If the company policy states that such agents are not allowed to be running on critical servers, deployment will come to a standstill. Often times, this exception will call for a meeting with technology steering board (TSB) and can delay the project significantly.

If preventive mode is in scope, ease of integration is a key element to consider in addition to software and hardware required. In some cases, organizations come to the realization of the difficulty in implementing DLP in preventive mode only after significant amount of work has been done. If this gets overlooked, the overall deployment can get very cumbersome.

- **Market Presence**

This is a key factor to consider in choosing a vendor. A vendor with good market presence has already experienced and dealt with problems in implementation. Secondly, this can help with policy creation, which is the core of this technology and has a direct impact on the workflow. For those that are required to meet government regulations, there are predefined policies that organizations can utilize. If a particular vendor has already served healthcare organizations, if not all, most requirements are very similar on a regulatory standpoint and this particular vendor can be a good fit for other healthcare organizations. I highly recommend requesting for reference from customers in similar industry.

- **Additional Staff**

When IDS made its first entry into the security industry, very few organizations realized the need for dedicated staff to weed out false positives from actual threat. In present day, almost all organizations that have deployed IDS devices, employ enough staffs to cover a 24/7 operation. DLP is in its early stages to conclude how much additional work this can create and the need for dedicated staff. We have seen enough false positives in the IDS world to realize that DLP is no exception. So, in order for DLP signatures to be more accurate than IDS signatures, is there a better matching mechanism used? Of course, not. While the content being sought is different, the mechanism is the same. With the exposure we have gotten in the IDS world, it should be obvious that there will be need for additional staff. Vendors often use confusing terms to get customers to buy into their solution. Once false positives were apparent, the advent of SIEM tool and its ability to correlate was supposed to do the magic. The end result has not been any different as far as the need for additional staff goes. Vendors are fully aware of the budget constraints of their prospective buyer. In order for them land their technology they will present it as though there is no need for staff. Hence the total cost of ownership will seem to fit within the budget. Besides supporting the technology, there is need for resources for escalation/follow up/remediation for all violations detected.

Conclusion

Data Loss Prevention is an ongoing process. To achieve high level of network and information security, the participants considered that security should be a concern all along the development lifecycle of products and services. . DLP solutions offer a multifaceted capability to significantly increase an enterprise's ability to manage risks to its key information assets. Sharing best practices, which should be distinguished from common practice, was also mentioned as an efficient means to increase the security level.