Understanding and Selecting a Data Loss Prevention Solution

Version 2.0
Released: October 21, 2010

Author's Note

The content in this report was developed independently of any sponsors. It is based on material originally posted on the Securosis blog but has been enhanced and professionally edited.

Special thanks to Chris Pepper for editing and content support.

Licensed by Websense

Websense, Inc. (NASDAQ: WBSN), a global leader in unified Web, data and email content security solutions, delivers the best security for modern threats at the lowest total cost of ownership to tens of thousands of enterprise, mid-market and small organizations around the world. Distributed through a global network of channel partners and delivered as software, appliance and software-as-a-service (SaaS), Websense content security solutions help organizations leverage new communication, collaboration and Web 2.0 business tools while protecting from advanced persistent threats, preventing the loss of confidential information and enforcing Internet use and security policies. Websense is headquartered in San Diego, Calif. with offices around the world. For more information, visit http://www.websense.com.

Contributors

The following individuals contributed significantly to this report through comments on the Securosis blog and/or follow-on review and conversations:

Wim Remes

Vince Hoang

Copyright

Table of Contents

Introduction to DLP

A (Still) Confusing Market

Data Loss Prevention has matured considerably since the release of the first version of this report three years ago. Back then, the market was dominated by startups with only a couple major acquisitions by established security companies. The entire market was probably smaller than the leading one or two providers today. Even the term 'DLP' was still under debate, with a menagerie of terms like Extrusion Prevention, Anti-Data Leakage, and Information Loss Protection still in use (leading us to wonder who, exactly, wants to *protect* information loss?).

While we've experienced maturation of the products, significant acquisitions by established security firms, and standardization on the term DLP, in many ways today's market is even more confusing than a few years ago. As customer interest in DLP increased competitive and market pressures diluted the term — with everyone from encryption tools to firewalls starting to claim they prevented "data leakage". In some cases, aspects of 'real' DLP have been added to other products as value-add features. And all along the core DLP tools continued to evolve and combine, expanding their features and capabilities.

Even today, it can still be difficult to understand the value of the tools and which products best suit which environments. We have more features, more options, and more deployment models across a wider range of products (and even services). You can go with a full-suite solution that covers your network, storage infrastructure, and endpoints; or focus on a single 'channel'. You might already have DLP embedded into your firewall, web gateway, antivirus, or a host of other tools.

So the question is no longer only "Do I need DLP and which product should I buy?" but "What kind of DLP will work best for my needs, and how do I figure that out?" This report provides the necessary background in DLP to help you understand the technology, know what to look for in a product (or service), and find the best match for your organization.

Defining DLP

There is no real consensus on what actually compromises a DLP solution. Some people consider encryption or USB port control DLP, while others limit the term to complete product suites. Securosis defines DLP as:

> *Products that, based on central policies, identify, monitor, and protect data at rest, in motion, and in use, through deep content analysis.*

Thus the defining characteristics are:

- Deep content analysis
- Central policy management

- Broad content coverage across multiple platforms and locations

DLP solutions both protect sensitive data and provide insight into the use of content within the enterprise. Few enterprises classify data beyond public vs. everything else. DLP helps organizations better understand their data, and improves their ability to classify and manage content.

Full-suite solutions provide complete coverage across your network, storage repositories, and endpoints, even if you aren't using the full capabilities. There are three other possible approaches:

- *Partial-suite* DLP solutions are dedicated DLP tools that cover two potential channels (*e.g.,* network and storage) and contain full workflow (such as incident management) and content analysis capabilities. There are very few partial suites available these days.
- *Single-channel* DLP solutions cover only one channel, but still include full DLP workflow and content analysis capabilities. While we tend to see more single channel offerings than partial suites, there are still only a few products on the market — almost all either network or endpoint.
- *DLP features* are now included in a variety of products, offer a subset of coverage and content analysis capabilities, and typically lack dedicated DLP workflow. For example, we have seen network firewalls with basic pattern-matching capabilities, vulnerability assessment scanners that look for particular data types (such as credit card numbers), and limited content analysis in an email security gateway.

More on DLP Features vs. DLP Solutions

When evaluating options it's sometimes difficult to characterize the real differences between DLP features and dedicated DLP solutions, and the value of each. The key differences are:

- *A DLP product or solution* includes centralized management, policy creation, and enforcement workflow, dedicated to the monitoring and protection of content and data. The user interface and functionality are dedicated to solving the business and technical problems of protecting content through content awareness.
- *DLP features* include some of the detection and enforcement capabilities of DLP products, but are not dedicated to protecting content and data.

This latter approach is sometimes called "DLP Light" to reflect its less-robust nature, and it's becoming extremely common in a variety of other security tools. (Gartner calls this "channel DLP").

This distinction is important because DLP products solve a specific business problem that may or may not be managed by the same business unit or administrator responsible for other security functions. We often see non-technical users such as legal or compliance officers responsible for the protection of content. Even human resources is often involved in disposition of DLP alerts. Some organizations find that the DLP policies themselves are highly sensitive or need to be managed by business unit leaders outside of security, which also may argue for a dedicated solution. Because DLP is dedicated to a clear business problem (protect my content) which is differentiated from other security problems (protect my PC or protect my network), if your primary goal is data protection you should focus on DLP solutions.

This doesn't mean that DLP Light won't be the right solution for your requirements, especially in smaller organizations. It tends to cost less and is often as easy to deploy as turning a switch on a management console. If you only need basic credit card protection, and are okay with coverage, analysis, and workflow limitations, the DLP features of another product might meet your needs. Or you might start with a DLP feature included in a product you already paid for to dip your toes in the water and get a better sense of how big a problem you might have before migrating to a full, dedicated

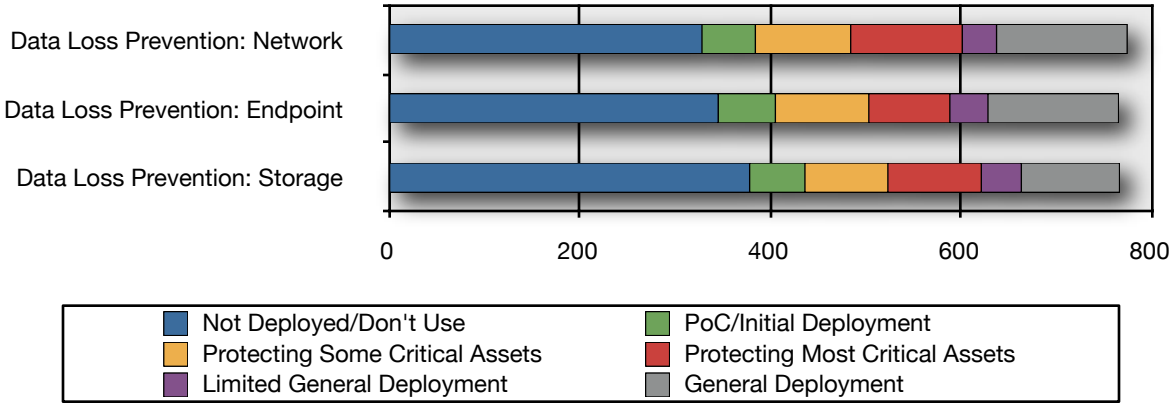
solution. We will provide more direct guidance on how to choose between the two in the *DLP Selection Process* section, and describe common implementations in the *Technical Architecture* section.

DLP Usage and Effectiveness Metrics

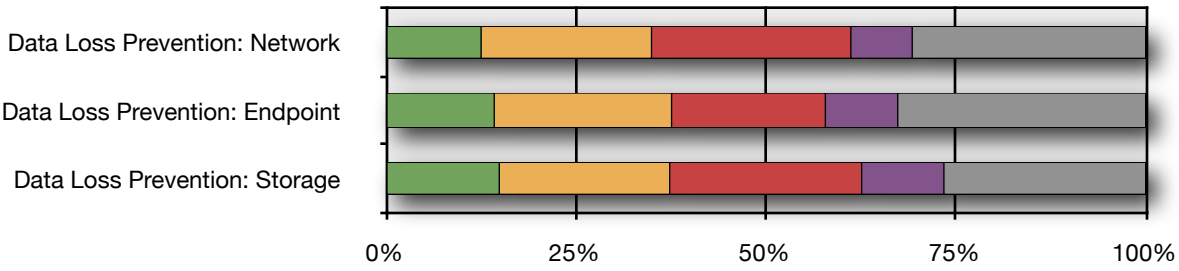
In 2010 Securosis ran an extensive data security survey of over 1,000 information technology professionals across all verticals and organization sizes. Data Loss Prevention rated among the top five security controls among the 19 surveyed across three different effectiveness categories. We also asked respondents about the scale and scope of their deployments. For a full copy of that report, showing all the controls and responses, and an anonymized version of the raw data, please visit Securosis.com.

Usage

We first asked respondents about to the scope of their deployments (note that here we removed the non-DLP answers):

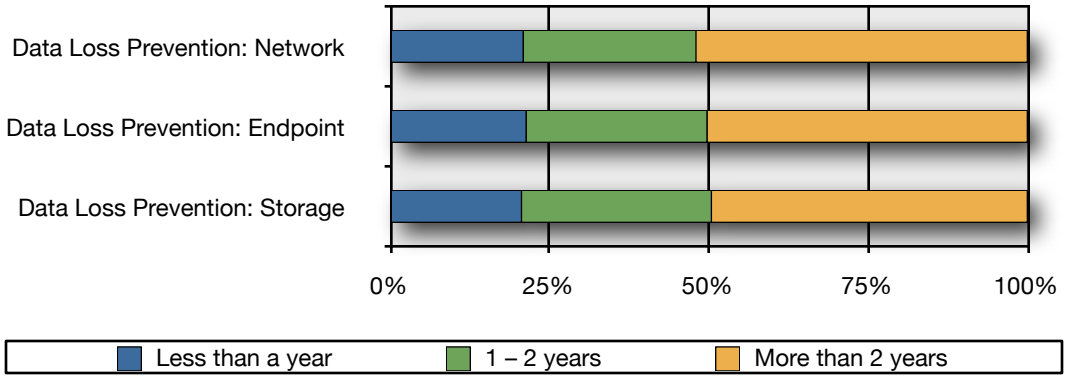


To focus only on DLP deployments, we can remove the "Not Deployed" responses and convert to a percentage scale:



Over 25% of organizations using DLP have it in general deployment.

We also asked *how long* they have been using these technologies:

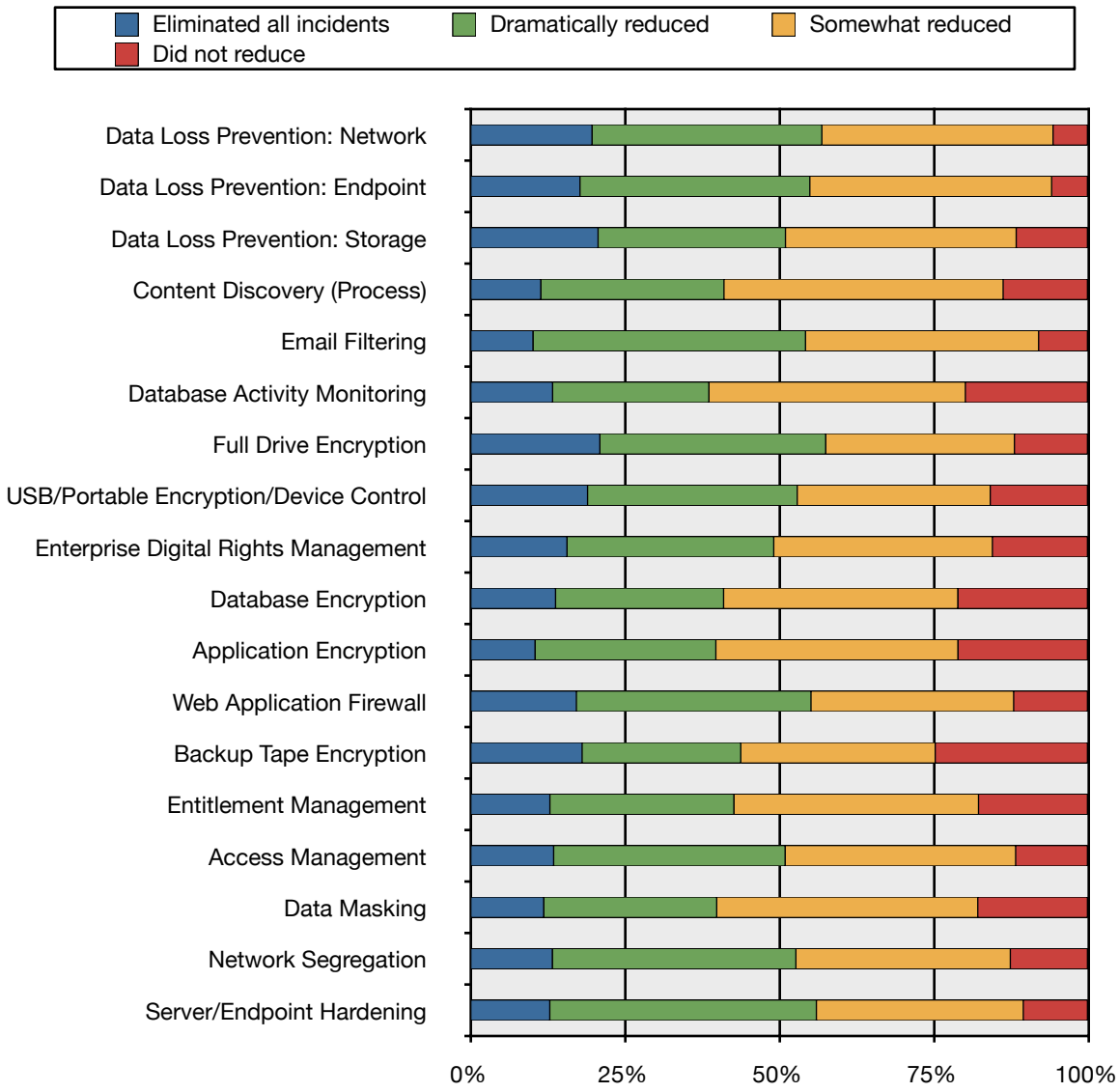


As you can see, half the DLP deployments have been in place for more than two years. Over 40% of respondents have deployed some version of DLP, which shows the technology has firmly entered the early mainstream of the market. Note that we do believe there is some degree of response bias in these results, and thus overall DLP penetration is likely lower than our results indicate, but even accounting for the bias there is still strong penetration.

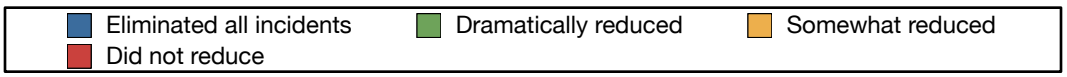
When we asked what controls organizations are considering deploying in the next 12 months, network DLP rated second and endpoint DLP third of 19 options.

Perceived Effectiveness

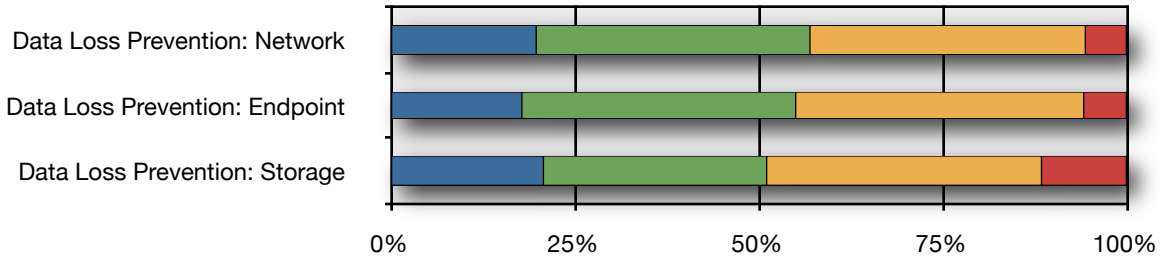
The focus of this survey was to assess how people rate the effectiveness of their various data security controls. Across all three effectiveness categories (reducing incidents, reducing incident severity, and reducing compliance costs), Data Loss Prevention consistently rated highly — even compared to traditional security tools like network segregation and server/endpoint hardening. Here is a raw response view of effectiveness at reducing incidents, with "Do not use" removed, converted to a percentage scale to make it easier to compare the controls (since we covered usage in the section above):



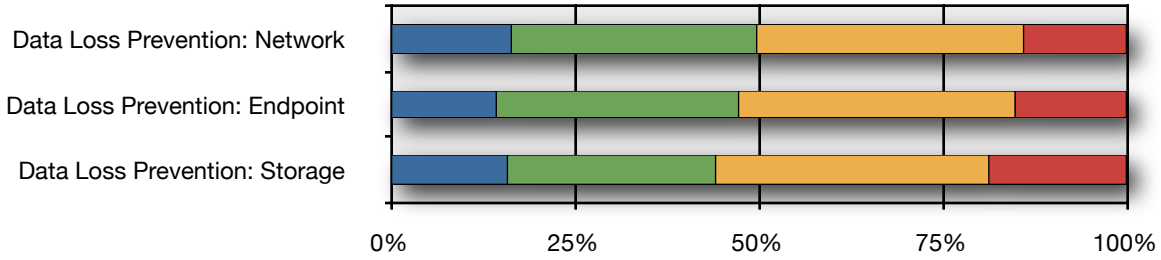
If we focus only on actual DLP deployments, we see:



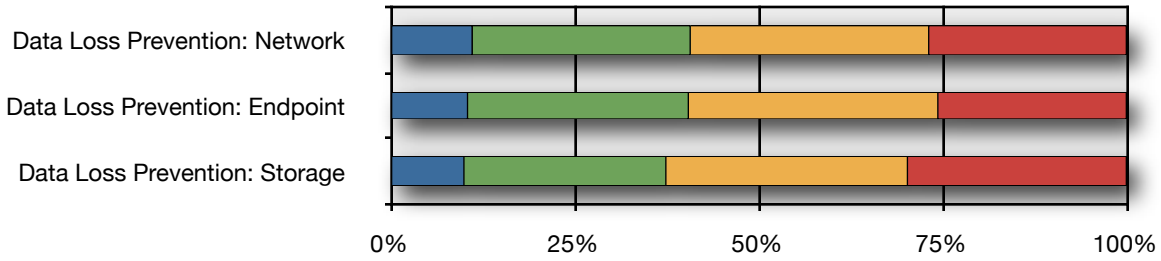
Incident Reduction Effectiveness (Controls in Use, Precentage Scale)



Incident Severity Reduction Effectiveness (Controls in Use, Percentage Scale)



Compliance Cost Reduction Effectiveness (Controls in Use, Percentage Scale)



As these results show, people using DLP rate it as extremely competitive compared to other data security controls — especially for reducing incidents. Although the compliance cost reduction effectiveness rates lower, it's possible this is affected by non-compliance driven projects.

In the following sections we'll dig into the technologies behind Data Loss Prevention, including the different deployment models, and finish the report with a detailed selection process and guidelines.

Content Awareness

Content vs. Context

Before delving into the particulars of the different content analysis techniques, we need to distinguish content from context. One of the defining characteristics of DLP solutions is their *content awareness*. This is the ability to analyze deep content using a variety of techniques, and is very different from analyzing context. It's easiest to think of content as a letter, and context as the envelope and environment around it. Context would include source, destination, size, recipients, sender, header information, metadata, time, format, and anything else short of the content of the letter itself. Context is highly useful and any DLP solution should include contextual analysis as part of an overall solution.

A more advanced version of contextual analysis is *business context analysis,* which involves deeper analysis of the content, its environment at the time of analysis, and the use of the content at that time. For example, while the envelope might tell you the sender and the destination, the business context will tell you which business unit the current holder of the envelope belongs to, their virtual location, what application they are reading it with, and so on.

Content awareness involves peering inside containers and analyzing the content itself. Its advantage is that while we use context, we're not restricted by it. If I want to protect a piece of sensitive data I want to protect it everywhere — not just in obviously sensitive containers. I'm protecting the data, not the envelope, so it makes a lot more sense to open the letter, read it, and decide how to handle it. This is more difficult and time consuming than basic contextual analysis, and is the defining characteristic of DLP solutions.

Contextual Analysis

Early contextual analysis used to be pretty simple — often little more than email headers or the metadata for a given file. Since then contextual analysis has evolved considerably to evaluate factors such as:

- File ownership and permissions.
- Use of encrypted file formats or network protocols.
- User role and business unit (through directory integration).
- Specific web services — such as known webmail providers and social networking sites.
- Web addresses (not just the session content).
- USB device information, such as manufacturer or model number.
- The desktop application in use (*e.g.,* something was copied from an Office document and then pasted into an encryption tool).

It's the contextual analysis that often provides the *business context* for the subsequent content analysis policies. This is one of the major benefits of DLP — rather than looking at packets or files in a vacuum, you can build policies that account for everything from the employee's job role to the application in use.

Content Analysis

The first step in content analysis is capturing the envelope and opening it. The DLP engine then needs to parse the context (we'll need that for the analysis) and dig into it. For a plain text email this is easy, but when you want to look inside binary files it gets a little more complicated. All DLP solutions solve this using *file cracking*. File cracking is the technology used to read and understand the file, even if the content is buried multiple levels down. For example, it's not unusual for the cracker to read an Excel spreadsheet embedded in a zipped Word file. The product needs to unzip the file, read the Word doc, analyze it, find the Excel data, read that, and analyze it. Other situations get far more complex, like a .pdf embedded in a CAD file. Many of the products on the market today support around 300 file types, embedded content, multiple languages, double byte character sets for Asian languages, and pulling plain text from unidentified file types. Quite a few use the Autonomy content engines to help with file cracking, but all the serious tools have quite a bit of proprietary capability in addition to the embedded content engine. Some tools support analysis of encrypted data if enterprise encryption is used with recovery keys, and most can identify standard encryption and use that as a contextual rule to block/quarantine content. Note that this is often limited to email and certain file transfers or types, and you need to ask carefully if you require more expansive encryption detection (*e.g.,* for encrypted `.rar` files on unusual ports or protocols).

Content Analysis Techniques

Once the content is accessed, seven major analysis techniques are used to find policy violations, each with its own strengths and weaknesses.

1. Rules-Based/Regular Expressions: This is the most common analysis technique available in both DLP products and other tools with DLP features. It analyzes the content for specific rules — such as 16 digit numbers that meet credit card checksum requirements, medical billing codes, and other textual analyses. Most DLP solutions enhance basic regular expressions with their own additional analyses (*e.g.,* a name in proximity to an address near a credit card number).

What it's best for: As a first-pass filter, or for detecting easily identified pieces of structured data like credit card numbers, Social Security numbers, and healthcare codes/records.

Strengths: Rules process quickly and can be easily configured. Most products ship with initial rule sets. The technology is well understood and easy to incorporate into a variety of products.

Weaknesses: Prone to higher false positive rates. Offers very little protection for unstructured content such as sensitive intellectual property.

2. Database Fingerprinting: Sometimes called Exact Data Matching, this technique takes either a database dump or live data (via ODBC connection) from a database and only looks for exact matches. For example, you could generate a policy to look only for credit card numbers in your customer base, thus ignoring your own employees buying online. More advanced tools look for combinations of information, such as the magic combination of first name or initial, with last name, with credit card or Social Security number, that triggers most US state-level breach disclosure laws. Make sure you understand the performance and security implications of nightly extracts vs. live database connections.

What it's best for: Structured data from databases.

Strengths: Very few false positives (close to 0). Allows you to protect customer & sensitive data while ignoring other, similar, data used by employees, such as their personal credit cards for online orders.

Weaknesses: Nightly dumps don't include transaction data since the last extract. Live connections can affect database performance. Large databases affect product performance.

3. Exact File Matching: With this technique you take a hash of a file and monitors for any files that match that exact fingerprint. Some consider this to be a contextual analysis technique because the file contents themselves are not analyzed.

What it's best for: Media files and other binaries where textual analysis isn't necessarily possible, such as photos, audio, movies, and certain proprietary design and engineering files.

Strengths: Works on any file type, low false positives (effectively none) with large enough hashes.

Weaknesses: Trivial to evade. Worthless for content that's edited, such as standard office documents and edited media files.

4. Partial Document Matching: This technique looks for a complete or partial match to protected content. You could build a policy to protect a sensitive document, and the DLP solution will look for either the complete text of the document, or even excerpts as small as a few sentences. For example, you could load up a business plan for a new product and the DLP solution would alert if an employee pasted a single paragraph into an instant message. Most solutions are based on a technique known as cyclical (or overlapping) hashing, where you take a hash of a portion of the content, offset a predetermined number of characters, then take another hash, and keep adding hashes of document segments until done. Outbound content is run through the same hash technique, and the hash values compared for matches. Some products use cyclical hashing as a base, then add more advanced linguistic analysis.

What it's best for: Protecting sensitive documents, or similar content with text, and source code. Unstructured content that's known to be sensitive.

Strengths: Ability to protect unstructured data. Generally low false positives (some vendors will say zero false positives, but any common sentence/text in a protected document can trigger alerts). Doesn't rely on complete matching of large documents; can find policy violations on even a partial match.

Weaknesses: Performance limitations on the total volume of content that can be protected. Common phrases/verbiage in a protected document may trigger false positives. Must know exactly which documents you want to protect. Trivial to avoid (ROT 13 encryption is sufficient for evasion).

5. Statistical Analysis: Use of machine learning, Bayesian analysis, and other statistical techniques to analyze a corpus of content and find policy violations in content that resembles the protected content. This category includes a wide range of statistical techniques which vary greatly in implementation and effectiveness. Some techniques are very similar to those used to block spam. Of all the techniques listed, this is the least commonly supported by different products.

What it's best for: Unstructured content where a deterministic technique, such as partial document matching would be ineffective. For example, a repository of engineering plans that's impractical to load for partial document matching due to high volatility or extreme volume.

Strengths: Can work with more nebulous content where you may not be able to isolate exact documents for matching. Can enforce policies such as "Alert on anything outbound that resembles the documents in this directory."

Weaknesses: Prone to false positives and false negatives. Requires a large corpus of source content — the bigger the better.

6. Conceptual/Lexicon: This technique uses a combination of dictionaries, rules, and other analyses to protect nebulous content that *resembles* an 'idea'. It's easier to give an example — a policy that alerts on traffic that resembles insider trading, which uses key phrases, word counts, and positions to find violations. Other examples are sexual harassment, running a private business from a work account, and job hunting.

What it's best for: Completely unstructured ideas that defy simple categorization but are similar to known documents, databases, or other registered sources.

Strengths: Not all corporate policies or content can be described using specific examples — conceptual analysis can find loosely defined policy violations other techniques can't even try to monitoring for.

Weaknesses: In most cases these are not user-definable, and the rule sets must be built by the DLP vendor with significant effort (costing more). Because of the loose nature of the rules, this technique is very prone to both false positives and false negatives.

7. Categories: Pre-built categories with rules and dictionaries for common types of sensitive data, such as credit card numbers/PCI protection, HIPAA, etc.

What it's best for: Anything that neatly fits a provided category. Typically easy to describe content related to privacy, regulations, or industry-specific guidelines.

Strengths: Extremely simple to configure. Saves significant policy generation time. Category policies can form the basis for more advanced enterprise-specific policies. For many organizations, categories can meet a large percentage of their data protection needs.

Weaknesses: One size fits all might not work. Only good for easily categorized rules and content.

These 7 techniques serve as the basis for most of the DLP products on the market. Not all products include all techniques, and there can be significant differences between implementations. Most products also support chaining techniques — building complex policies from combinations of different content and contextual analysis techniques.

Technical Architecture

Protecting Data in Motion, at Rest, and in Use

The goal of DLP is to protect content throughout its lifecycle — on the network, in storage, and on endpoints. In terms of DLP, this includes three major aspects:

- **Data in Motion** protection is monitoring (and potentially filtering) of traffic on the network (passively or inline via proxy) to identify content being sent across specific communications channels. This would include monitoring email, instant messages, and web traffic for snippets of sensitive source code. In motion tools can often block based on central policies, depending on the type of traffic.
- **Data at Rest** is protected by scanning of storage and other content repositories to identify where sensitive content is located. We often call this content discovery. For example, you can use a DLP product to scan your servers and identify documents with credit card numbers. If the server isn't authorized for that kind of data, the file can be encrypted or removed, or a warning sent to the file owner.
- **Data in Use** is addressed by endpoint solutions that monitor data as the user interacts with it. For example, they can identify when you attempt to transfer a sensitive document to a USB drive and block it (as opposed to blocking use of the USB drive entirely). Data in use tools can also detect things like copy and paste, or use of sensitive data in an unapproved application (such as someone attempting to encrypt data to sneak it past the sensors).

As we hinted in the description, these translate to three major architectural components: *network monitoring/filtering, storage scanning, and endpoint agents.* Even DLP features of other security products or DLP services fit within one or more of these major architectural pieces.

In the rest of this section we will walk through these components and describe how they work and the different deployment options. We will focus on dedicated DLP solutions, but many of these details also apply to DLP features. We will also talk more about the common DLP Light based architectures we see at the end of this section.

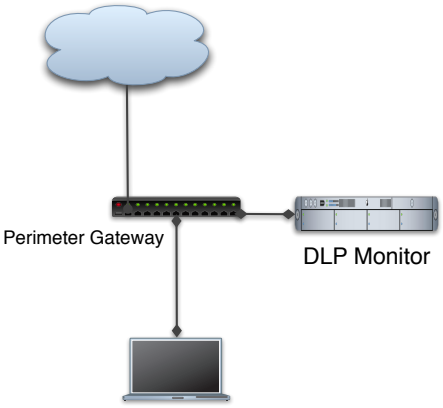
Networks

Many organizations first enter the world of DLP with network based products that provide broad protection for managed and unmanaged systems. It's typically easier to start a deployment with network products to gain broad coverage quickly. Early products limited themselves to basic monitoring and alerting but all current products include advanced capabilities to integrate with existing network infrastructure to provide protective, not just detective, controls.

Network Monitor

At the heart of most DLP solutions lies a passive network monitor. The network monitoring component is typically deployed at or near a gateway on a SPAN or mirror port (or a similar tap). It performs full packet capture, session reconstruction, and content analysis in real time. Performance is more complex and subtle than vendors normally

discuss. First, on the client expectation side, most clients claim they need performance to match their full peak bandwidth, but that level of performance is unnecessary except in very unusual circumstances because few organizations are really running that high a level of communications traffic. DLP is a tool to monitor employee communications, not all network traffic. Realistically we find that small enterprises normally run under 50mbyte/sec of relevant traffic, medium enterprises run in the 50-200mbyte/s range, and large enterprises around 300mbyte/s (as high as 500 in a few cases). Because of the content analysis overhead, not every product runs full packet capture. You might have to choose between pre-filtering (and thus missing non-standard traffic) or buying more boxes and load balancing. Also, some products lock monitoring into pre-defined port and protocol combinations, rather than using service/channel identification based on packet content. Even if full application channel identification is included, you will need to make sure it's enabled. Otherwise you might miss non-standard communications such as connecting over an unusual port. Most of the network monitors are dedicated general-purpose server hardware running DLP software. A few vendors deploy true specialized appliances.



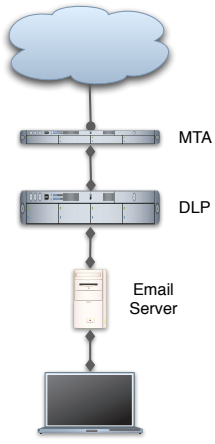
Passive Monitoring Architecture

Also keep in mind, especially when testing, that performance is often tied to the number and scope of DLP policies you deploy. If you perform large amounts of partial document matching or database fingerprinting (both of which rely on repositories of hash values) you might find performance slowing and need to move toward load balancing or separating traffic streams. For example, you can offload email to a dedicated monitor because, as we'll discuss in a moment, email monitoring using a different architectural model than web and most other traffic.

Finally, there are some organizations which exceed the average monitoring requirements in terms of both bandwidth and port/protocol coverage due to either their size or the nature of threat they face (advanced attackers more commonly exfiltrate data over unusual ports & protocols). If you fall into this category, make sure you include it in your DLP requirements and test it in the selection process. You should also coordinate your DLP program with any other egress filtering projects, which may be able to reduce your DLP load.

Email Integration

The next major component is email integration. The store and forward architecture of Internet email enables several additional capabilities including quarantine, encryption integration, and filtering — without the tight throughput constraints required to avoid blocking synchronous traffic. Most products embed an MTA (Mail Transport Agent) in the product, allowing you to simply add it as another hop in the email chain. Quite a few also integrate with some of the major existing MTAs and email security solutions directly for better performance. One weakness of this approach is it doesn't give you access to internal email. If you're on an Exchange server, internal messages that never make it through the external MTA because there's no need to transmit that traffic externally. To monitor internal mail you'll need direct Exchange/Lotus integration, which is surprisingly rare. Full integration is different than just scanning logs/libraries after the fact, which is what some companies call internal mail support. Good email integration is absolutely critical if you ever want to do any filtering, as opposed to just monitoring.



Email Architecture

Filtering/Blocking and Proxy Integration

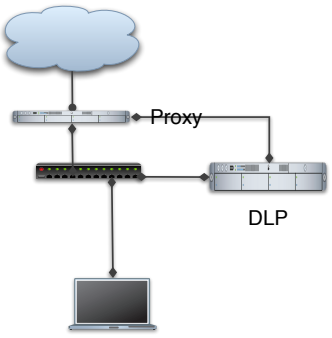
Nearly anyone deploying a DLP solution will eventually want to start blocking traffic. You can watch all your juicy sensitive data flowing out to the nether regions of the Internet for so long before you start taking action. But blocking isn't the easiest thing in the world, especially because we need to allow all good traffic, block only bad traffic, and make the decision using real-time content analysis. Email, as we just mentioned, is fairly straightforward to filter. It's not quite real-time and is proxied by its very nature. Adding one more analysis hop is a manageable problem in even the most complex environments. Outside of email, however, most communications traffic is synchronous — everything runs in real time. So if we want to filter it we either need to bridge the traffic, proxy it, or poison it from the outside.

Bridge

A bridge is simply a system with two network cards, which performs content analysis in the middle. If it sees something bad, the bridge breaks the connection for that session. Bridging isn't the best approach for DLP because it might not stop all the bad traffic before it leaks out. It's like sitting in a doorway watching everything go past with a magnifying glass; by the time you have enough traffic to make an intelligent decision, you may have missed the really good stuff. Very few products take this approach, although it does have the advantage of being protocol agnostic and able to block any form of traffic flowing through the bridge.

Proxy

In simplified terms, a proxy is a protocol/application specific application which queues up traffic before passing it on, allowing for deeper analysis. We see gateway proxies mostly for HTTP, FTP, and IM protocols. Few DLP solutions include their own proxies — they tend to integrate with existing gateway/proxy vendors because most customers prefer integration with tools they have already deployed. Integration for web gateways is typically through the iCAP protocol — this enables the proxy to grab the traffic, send it to the DLP product for analysis, and cut communications if there's a violation. This means (assuming you already use an iCAP compatible gateway) you don't have to add another piece of hardware in front of your network traffic and the DLP vendors can avoid building dedicated network hardware for inline analysis. If the gateway includes a reverse SSL proxy you can also sniff SSL connections. You will need to make changes on your endpoints to deal with all the certificate alerts, but can now peer into encrypted traffic. For instant messaging you'll need an IM proxy and a DLP product that specifically supports your IM protocol.



Proxy/Gateway

We *highly* recommend that if you monitor web traffic, you deploy and integrate your DLP with a reverse SSL proxy so you can view encrypted traffic. Otherwise you lose the ability to monitor any SSL connections, such as to Gmail, Hotmail, Facebook, etc.

TCP Poisoning

The last method of filtering is TCP poisoning. You can monitor the traffic and when you see something bad inject a TCP reset packet to kill the connection. This works on every TCP protocol but isn't very efficient. For one thing, some protocols will keep trying to get the traffic through. If you TCP poison a single email message, the server will keep trying to send it for several days, as often as every 15 minutes. The other problem is the same as bridging: since you don't queue the traffic at all, by the time you notice something bad it might be too late. It's a good stop-gap for broad protocol coverage, but you should rely on proxies for filtering as much as possible.

Internal Networks

Although technically capable of monitoring internal networks, DLP is rarely used on internal traffic other than email. Perimeter gateways provide convenient choke points — without them internal monitoring is a daunting prospect due to

concerns about cost, performance, and policy management/false positives. A few DLP vendors have partnerships for internal monitoring but this is a lower priority feature for most organizations.

Distributed and Hierarchical Deployments

All medium to large enterprises, and many smaller organizations, have multiple locations and web gateways. A DLP solution should support multiple monitoring points, including a mix of passive network monitoring, proxy points, email servers, and remote locations. While processing/analysis can be offloaded to remote enforcement points, they should send all events back to a central management server for workflow, reporting, investigations, and archiving. Remote offices are usually easy to support because you can just push policies down and reporting back, but not every product offers this capability.

The more advanced products support hierarchical deployments for organizations that want to manage DLP differently between multiple geographic locations, or by business unit. International companies often need this to meet legal monitoring requirements which vary by country. Hierarchical management supports coordinated local policies and enforcement in different regions, running on their own management servers, communicating back to a central management server. Early products only supported one management server but now we have options to deal with these distributed situations, with a mix of corporate/regional/business unit policies, reporting, and workflow. With so much sensitive information moving around, it's important each tier in the hierarchy is well secured and all communications between DLP nodes is encrypted.

Storage

While catching leaks on the network is fairly powerful, it's only one small part of the problem. Many customers are finding that it's just as valuable, if not more valuable, to figure out where all that data is stored in the first place. We call this *content discovery.* Enterprise search or electronic discovery tools might be able to help with this, but they really aren't tuned well for this specific problem. Enterprise data classification tools can also help, but based on discussions with a number of clients they don't seem to work well for finding specific policy violations. Thus we see many clients opting to use the content discovery features of their DLP products.

Content discovery consists of three components, based on where information is stored:

1. Storage: scanning mass storage including file servers, SAN, and NAS.
2. Applications/Servers: application-specific scanning of stored data on email servers, document management systems, and databases.
3. Endpoints: scanning workstations and laptops for content.

We will mostly focus on storage and application/server scanning in this section because we will cover endpoints in more detail later.

Content Discovery Techniques

There are four basic techniques for content discovery:

1. **Remote Scanning:** Either the central policy server or a dedicated scanning server accesses storage repositories via network shares or other administrative access. Files are scanned for content violations. Connections are often made using administrative credentials, and any content transferred should be encrypted, but this may require reconfiguration of the storage repository and isn't always feasible. Most tools allow bandwidth throttling to limit network impact, and scanning servers are often placed close to the storage to increase speed and limit network impact. This technology supports scanning nearly any storage repository, but even with optimization performance is limited by the network.

2. **Agent-Based Scanning:** An agent is installed on the system (server) to be scanned and scanning is performed locally. Agents are platform specific and use local CPU cycles, but can potentially perform significantly faster than remote scanning, especially for large repositories.
3. **Memory-Resident Agent Scanning:** Rather than deploying a full-time agent, a memory-resident agent is installed which performs a scan and then exits without leaving anything running or stored on the local system. This offers the performance of agent-based scanning in situations where you don't want an agent running all the time.
4. **Application Integration:** Direct integration (often using an agent) with document management, content management, or other storage repositories. This integration not only supports visibility into management content, but allows the discovery tool to understand local context/metadata and possibly enforce actions within the system.

Any of these technologies can work for any of the modes, and enterprises typically deploy a mix depending on policy and infrastructure requirements. We currently see deployments guided by technology limitations of each approach:

- Remote scanning can significantly increase network traffic and has performance limitations based on network bandwidth and target and scanner network performance. Some solutions can scan gigabytes per day (sometimes hundreds, but not terabytes per day) per server due to these practical limitations — which may be inadequate for very large storage. The advantage is that you only need access to a file share for scanning.
- Agents, ephemeral or permanent, are limited by processing power and memory on the target system, which may translate into restrictions on the number of policies that can be enforced, and the types of content analysis that can be used.
- Agents don't support all platforms.

Storage Enforcement

Once a policy violation is discovered, a DLP tool can take a variety of actions:

- **Alert/Report:** Create an incident in the central management server just like a network violation.
- **Warn:** Notify the user via email that they may be in violation of policy.
- **Quarantine/Notify:** Move the file to the central management server and leave a text file with instructions on how to request recovery of the file.
- **Quarantine/Encrypt:** Encrypt the file in place, usually leaving a plain text file describing how to request decryption.
- **Quarantine/Access Control:** Change access controls to restrict access to the file.
- **Remove/Delete:** Either transfer the file to the central server without notification, or simply delete it.

The combination of different deployment architectures, discovery techniques, and enforcement options creates a powerful combination for protecting data at rest and supporting compliance initiatives. For example, we often see deployments of DLP to support PCI compliance — more for the ability to ensure (and report) that no cardholder data is stored in violation of PCI than to protect email or web traffic.

Integration and Additional Features

The most common integration we see is with document management systems, including Microsoft SharePoint. When integrated with a DMS more information is available than typically provided by a file share, sometimes even including file usage. DMS metadata and privileges are usually more robust than those on generic file shares, providing greater context for building policies.

On the database side some products now include the ability to look for sensitive information over an ODBC connection. Rather than looking at *all* data in the database, the tool scans the first *n* rows of a table and the column headers to see if

any sensitive data might be present. This is especially useful for evaluating the many *ad hoc* databases commonly found within business units.

Finally, we are seeing growing interest among users in usage monitoring (who is accessing files) and better privilege and entitlement management, which are typically handled by a different set of tools. We see these features as the next step in DLP for storage, and expect to see them becoming integrated into most DLP solutions over time.

Endpoints

DLP usually starts on the network because that's the most cost-effective way to get the broadest coverage. Network monitoring is unintrusive (unless you have to crack SSL) and offers visibility to any system on the network, managed or unmanaged, server or workstation. Filtering is more difficult, but again still relatively straightforward (especially for email) and covers all systems connected to the network. But it's clear this isn't a complete solution: it doesn't protect data when someone walks out the door with a laptop, and can't even prevent people from copying data to portable storage like USB drives. To move from a "leak prevention" solution to a "content protection" solution, products need to expand not only to stored data, but to the endpoints where data is used.

> *Note: Although there have been large advancements in endpoint DLP, endpoint-only solutions are not recommended for most users. As we'll discuss, they normally require compromise on the numbers and types of policies that can be enforced, offer limited email integration, and offer no protection for unmanaged systems. Long-term, you'll need both network and endpoint capabilities, and nearly every network DLP solution offers at least some endpoint protection.*

Adding an endpoint agent to a DLP solution not only gives you the ability to discover locally stored content, but also to potentially protect systems no longer on the network or even protect data as it's being actively used. While extremely powerful, this has been problematic to implement. Agents need to perform within the resource constraints of a standard laptop while maintaining content awareness. This can be difficult if you have large policies such as, "Protect all 10 million credit card numbers from our database." as opposed to something simpler, like "Protect any credit card number." — which will generate false positives every time an employee visits Amazon.com.

Key Capabilities

Agents include four generic layers/features:

1. ***Content Discovery:*** Scanning of stored content for policy violations.
2. ***File System Protection:*** Monitoring of and enforcement on file operations as they occur (as opposed to discovery, which is scanning of content already written to media). This is most often used to prevent content from being written to portable media (USB). It's also where tools hook in for automatic encryption or application of DRM rights.
3. ***Network Protection:*** Monitoring and enforcement of network operations. An endpoint agent can provide protection similar to gateway DLP when an endpoint is off the corporate network. As most endpoints handle printing and faxing as a form of network traffic, this is where most print/fax protection can be enforced (the rest comes from special print/fax hooks).
4. ***GUI/Kernel Protection:*** A more generic category to cover data in use scenarios, such as Copy & Paste, application restrictions, and Print Screen.

Between these four categories we cover most of the day to day operations a user might perform that places content at risk. It hits our primary drivers from the last section: protecting data from portable storage, protecting systems off the corporate network, and supporting discovery on the endpoint. Most tools on the market start with file and then networking features before moving on to some of the more complex GUI/kernel functions.

Use Cases

Endpoint DLP is evolving to support a few critical use cases:

- Ongoing scanning of local storage for sensitive data that shouldn't ever appear on an endpoint, such as credit cards or customer Personally Identifiable Information.
- Enforcing network rules off the managed network, or modified rules on more hostile networks.
- Restricting sensitive content from portable storage, including USB drives, CD/DVD drives, home storage, and devices such as smartphones and PDAs.
- Restricting Copy and Paste of sensitive content.
- Restricting applications allowed to use sensitive content — *e.g.,* only allowing encryption with an approved enterprise solution, but not tools downloaded online that don't support enterprise data recovery.
- Integration with Enterprise Digital Rights Management to automatically apply access control to documents based on the included content.
- Auditing use of sensitive content for compliance reporting.

Agent Content Awareness

Even if you have an endpoint with a quad-core processor and 8gb of RAM, it would be wasteful to devote all that horsepower to enforcing DLP — especially if it interferes with the primary purpose of the system.

Content analysis may be resource intensive, depending on the types of policies you are trying to enforce. Additionally, different agents have different enforcement capabilities, which do not always match up to their gateway counterparts. At minimum most endpoint tools support rules/regular expressions, some degree of partial document matching, and a whole lot of contextual analysis. Others support their entire repertoire of content analysis techniques, but you will likely have to tune policies to run on more constrained endpoints.

Some tools rely on the central management server for aspects of content analysis, to offload agent overhead. Rather than performing all analysis locally, they ship content back to the server and act on any results. This obviously isn't ideal, because those policies can't be enforced when the endpoint is off the enterprise network, and it sucks up a bit of bandwidth. But this does enable enforcement of policies that are otherwise totally unrealistic on an endpoint, such as fingerprinting of a large enterprise database.

One emerging option is policies that adapt based on endpoint location. For example, when you're on the enterprise network most policies are enforced at the gateway. Once you access the Internet outside the corporate walls, a different set of policies is enforced. You might use database fingerprinting of the customer database at the gateway when the laptop is in the office or on a (non-split-tunneled) VPN, but drop to a rule/regex for Social Security Numbers or account numbers for mobile workers. Sure, you'll get more false positives, but you're still able to protect your sensitive information within performance constraints.

Agent Management

Agent management consists of two main functions: deployment and maintenance. On the deployment side, most tools today are designed to work with whatever workstation management tools your organization already uses. As with other software tools, you create a deployment package and then distribute it along with any other software updates. If you don't already have a software deployment tool, you'll want to look for an endpoint DLP tool that includes basic deployment capabilities. Since all endpoint DLP tools include central policy management, deployment is fairly straightforward. There's little need to customize packages based on user, group, or other variables beyond the location of the central management server.

The rest of the agent's lifecycle, aside from major updates, is controlled through the central management server. Agents should communicate regularly with the central server to receive policy updates and report incidents & activity. When the central management server is accessible, this should happen in near-real-time. When the endpoint is off the enterprise network (without VPN/remote access), the DLP tool will store violations locally in a secure repository that's encrypted and inaccessible to the user. The tool will then connect to the management server next time it's accessible, receiving policy updates and reporting activity. The management server should produce aging reports to help you identify endpoints which are out of date and need to be refreshed. Under some circumstances, the endpoint may be able to communicate remote violations through encrypted email or another secure mechanism from outside the corporate firewall.

Aside from content policy updates and activity reporting, there are a few other features that require central management. For content discovery, you'll need to control scanning schedule/frequency, as well as bandwidth and performance (*e.g.*, capping CPU usage). For real time monitoring and enforcement you'll also want performance controls, including limits on how much space is used to store policies and the local cache of incident information.

Once you set your base configuration, you shouldn't need to do much endpoint management directly. Things like enforcement actions are handled implicitly as part of policy, so integrated into the main DLP policy interface.

Enforcement Options

Because any given endpoint agent might be monitoring data in motion, at rest, and in use, there's a wide range of potential alerting and enforcement options. Rather than listing every possible option, many of which are also available in network and storage DLP, here are some endpoint-specific examples:

- Blocking transfer of a file to portable storage.
- Allowing transfer to portable storage, but requiring the user to submit a "business justification" in a pop-up window which is sent back with the incident information to the central DLP server.
- Application of Digital Rights Management to a discovered file.
- Creating hidden 'shadow' copies of any files transferred to portable storage, which are sent to the DLP server the next time the user is on the organization's network and later reviewed to determine whether incident investigation is warranted.
- Allowing Copy & Paste only between approved applications (application control). Attempting to Copy & paste protected content into an unapproved application is blocked (*e.g.,* to prevent pasting into an encryption application).
- Only allowing printing to approved print servers.

DLP Features and Integration with Other Security Products

Up until now we have mostly focused on describing aspects of dedicated DLP solutions, but we also see increasing interest in DLP Light tools for four main use cases:

- Organizations which turn on the DLP feature of an existing security product, like an endpoint suite or IPS, to generally assess their data security issues. Users typically turn on a few general rules and use the results more to scope out their issues than to actively enforce policies.
- Organizations which only need basic protection on one or a few channels for limited data types, and want to bundle the DLP with existing tools if possible — often to save on costs. The most common examples are email filtering, endpoint storage monitoring, or content-based USB alerting/blocking for credit card numbers or customer PII.
- Organizations which want to dip their toes into DLP with plans for later expansion. They will usually turn on the DLP features of an existing security tool that is also integrated with a larger DLP solution. These are often provided by larger vendors which have acquired a DLP solution and integrated certain features into their existing product line.
- To address a very specific, and very narrow, compliance deficiency that a DLP Light feature can resolve.

There are other examples, but these are the four cases we encounter most often. DLP Light tends to work best when protection scope and content analysis requirements are limited, and when cost is a major concern. There is enough market diversity now that full DLP solutions are available even for cost-conscious smaller organizations, so we suggest that if more-complete data protection is your goal, you take a look at the DLP solutions for small and mid-size organizations rather than assuming DLP Light is your only option.

Although there are a myriad of options out there, we do see some consistencies between the various DLP Light offerings, as well as full-DLP integration with other existing tools. The next few paragraphs highlight the most common options in terms of features and architectures, including the places where full DLP solutions can integrate with existing infrastructure:

Content Analysis and Workflow

Most DLP Light tools start with some form of rules/pattern matching — usually regular expressions, often with some additional contextual analysis. This base feature covers everything from keywords to credit card numbers. Because most customers don't want to build their own custom rules, the tools come with pre-built policies. The most common is to find credit card data for PCI compliance, since that drives a large portion of the market. We next tend to see PII detection, followed by healthcare/HIPAA data discovery; all of which are designed to meet clear compliance needs.

The longer the tool/feature has been on the market, the more categories it tends to support, but few DLP light tools or features support the more advanced content analysis techniques we've described in this paper. This usually results in more false positives than a dedicated solution, but for some of these data types, like credit card numbers, even a false positive is something you usually want to take a look at.

DLP Light tools or features also tend to be more limited in terms of workflow. They rarely provide dedicated workflow for DLP, and policy alerts are integrated into whatever existing console and workflow the tool uses for its primary function. This might not be an issue, but it's definitely important to consider before making a final decision, as these constraints might impact your existing workflow and procedures for the given tool.

Network Features and Integration

DLP features are increasingly integrated into existing network security tools, especially email security gateways. The most common examples are:

- **Email Security Gateways:** These were the first non-DLP tools to include content analysis, and tend to offer the most policy/category coverage. Many of you already deploy some level of content-based email filtering. Email gateways are also one of the top integration points with full DLP solutions: all the policies and workflow are managed on the DLP side, but analysis and enforcement are integrated with the gateway directly rather than requiring a separate mail hop.
- **Web Security Gateways:** Some web gateways now directly enforce DLP policies on the content they proxy, such as preventing files with credit card numbers from being uploaded to webmail and social networking services. Web proxies are the second most common integration point for DLP solutions because, as we described in the *Technical Architecture* section, they proxy web and FTP traffic and make a perfect filtering and enforcement point. These are also the tools you will use to reverse proxy SSL connections to monitor those encrypted communications, since that's a critical capability these tools require to block inbound malicious content. Web gateways also provide valuable context, with some able to categorize URLs and web services to support policies that account for the web destination, not just the content and port/protocol.
- **Unified Threat Management:** UTMs provide broad network security coverage, including at least firewall and IPS capabilities, but usually also web filtering, an email security gateway, remote access, and web content filtering

(antivirus). These are a natural location to add network DLP coverage. We don't see many yet integrated with full DLP solutions, and they tend to build their own analysis capabilities (primarily for integration and performance reasons).

- **Intrusion Detection and Prevention Systems:** IDS/IPS tools already perform content inspection, and thus make a natural fit for additional DLP analysis. This is usually basic analysis integrated into existing policy sets, rather than a new, full content analysis engine. They are rarely integrated with a full DLP solution, although we do expect to see this over time, because they are already effective at killing active sessions.

Endpoint Features and Integration

DLP features have appeared in various endpoint tools aside from dedicated DLP products since practically before there was a DLP market. This continues to expand, especially as interest grows in controlling USB usage without onerous business impact.

- **USB/Portable Device Control:** A frequent inhibitor to deployment of portable storage management tools is their impact on standard business processes. There is always a subset of users who legitimately needs some access to portable storage for file exchange (*e.g.,* sales presentations), but the organization still wants to audit or even block inappropriate transfers. Even basic content awareness can clearly help provide protection while reducing business impact. Some tools include basic DLP capabilities, and we are seeing others evolve to offer somewhat extensive endpoint DLP coverage — with multiple detection techniques, multivariate policies, and even dedicated workflow. This is also a common integration/partner point for full DLP solutions, although due to various acquisitions we don't see those partnerships quite as often as we used to. When evaluating this option, keep in mind that some tools position themselves as offering DLP capabilities but lack *any* content analysis; instead relying on metadata or other context. Finally, despite its incredible usefulness, we see creation of shadow copies of files in many portable device control products, but almost never in DLP solutions.
- **Endpoint Protection Platforms:** For those of you who don't know, EPP is the term for comprehensive endpoint suites that include antivirus, host intrusion prevention, and everything from remote access and Network Admission Control to application whitelisting. Many EPP vendors have acquired full or endpoint-only DLP products and are in various stages of integration. Other EPP vendors have added basic DLP features — most often for monitoring local files or storage transfers of sensitive information. So there are options for either basic endpoint DLP (usually some preset categories), all the way up to a DLP client integrated with a dedicated DLP suite.
- **'Non-Antivirus' EPP:** There are also endpoint security platforms that are dedicated to more than just portable device control, but not focused around antivirus like other EPP tools. This category covers a range of tools, but the features offered are generally comparable to the other offerings.

Overall, most people deploying DLP features on an endpoint (without a dedicated DLP solution) are focused on scanning the local hard drive and/or monitoring/filtering file transfers to portable storage. But as we described earlier you might also see anything from network filtering to application control integrated into endpoint tools.

Storage Features and Integration

We don't see nearly as much DLP Light in storage as in networking and endpoints — in large part because there aren't as many clear security integration points. Fewer organizations have any sort of storage security monitoring, whereas nearly every organization performs network and endpoint monitoring of some sort. But while we see less DLP Light, as we have already discussed, we see extensive integration on the DLP side for different types of storage repositories.

- **Database Activity Monitoring and Vulnerability Assessment:** DAM products, many of which now include or integrate with Database Vulnerability Assessment tools, now sometimes include content analysis capabilities. These are designed to either find sensitive data in large databases, detect sensitive data in unexpected database responses, or help automate database monitoring and alerting policies. Due to the high potential speeds and transaction volumes

involved in real time database monitoring, these policies are usually limited to rules/patterns/categories. Vulnerability assessment policies may include more options because the performance demands are different.

- **Vulnerability Assessment:** Some vulnerability assessment tools can scan for basic DLP policy violations if they include the ability to passively monitor network traffic or scan storage.
- **Document Management Systems:** This is a common integration point for DLP solutions, but we don't see DLP included as a DMS feature.
- **Content Classification, Forensics, and Electronic Discovery:** These tools aren't dedicated to DLP, but we sometimes see them positioned as offering DLP features. They do offer content analysis, but usually not advanced techniques like partial document matching and database fingerprinting/matching.

Other Features and Integrations

The lists above include most of the DLP Light, feature, and integration options we've seen; but there are few categories that don't fit quite as neatly into our network/endpoint/storage divisions:

- **SIEM and Log Management:** All major SIEM tools can accept alerts from DLP solutions and possibly correlate them with other collected activity. Some SIEM tools also offer DLP features, depending on what kinds of activity they can collect to perform content analysis on. Log management tools tend to be more passive, but increasingly include some similar basic DLP-like features when analyzing data. Most DLP users tend to stick with their DLP solutions for incident workflow, but we do know cases where alerts are sent to the SIEM for correlation or incident response, as well as when the organization prefers to manage all security incidents in the SIEM.
- **Enterprise Digital Rights Management:** Multiple DLP solutions now integrate with Enterprise DRM tools to automatically apply DRM rights to files that match policies. This makes EDRM far more usable for most organizations, since one major inhibitor is the complexity of asking users to apply DRM rights. This integration may be offered both in storage and on endpoints, and we expect to see these partnerships continue to expand.
- **Email Encryption:** Automatic encryption of emails based on content was one of the very first third party integrations to appear on the market, and a variety of options are available. This is most frequently seen in financial and healthcare organizations (including insurance) with strict customer communications security requirements.

DLP Software as a Service (SaaS)

Although there aren't currently any completely SaaS-based DLP services available due to the extreme internal integration requirements for network, endpoint, and storage coverage, some early SaaS offerings are available for limited DLP deployments. Due to the ongoing interest in the cloud and SaaS in general, we expect to see new options appear on a regular basis.

Current DLP SaaS offerings fall into the following categories:

- **DLP for email:** Many organizations are opting for SaaS-based email security rather than installing internal gateways (or a combination of the two). This is clearly a valuable and straightforward integration point for monitoring outbound email. Most services don't yet include full DLP analysis capabilities, but many major email security service providers have also acquired DLP solutions (sometimes before buying the email SaaS provider), so we expect integration to expand. Ideally, if you obtain your full DLP solution from the same vendor as your email security SaaS, the policies and violations will synchronize from the cloud to your local management server.
- **Content Discovery:** While still fairly new to the market, it's possible to install an endpoint (or server, usually limited to Windows) agent that scans locally and reports to a cloud-based DLP service. This targets smaller to mid-size organizations which don't want the overhead of a full DLP solution, and don't have such deep requirements.

- **DLP for web filtering:** As with email, we are seeing organizations adopt cloud-based web content filtering to block web based attacks before they hit the local network, and better support remote users and locations. All the content is already being scanned, so this is a nice fit for DLP SaaS. With the same acquisition trends we saw in email services, we also hope to see integrated policy management and workflow for organizations obtaining their DLP web filtering from the same SaaS provider which supplies their on-premise DLP solution.

There are definitely other opportunities for DLP SaaS, and we expect to see more options develop over the next few years. But before jumping in with a SaaS provider keep in mind that they won't be merely assessing and stopping external threats, but scanning for extremely sensitive content and policy violations. This may limit most DLP SaaS to focusing on common low hanging fruit, like those ubiquitous credit card numbers and customer PII, as opposed to sensitive engineering plans or large customer databases.

A Note on Installation Architecture Options

Due to a combination of product design and acquisition activity, there are huge differences in how the different DLP solutions are packaged and installed.

Some are based on all-in-one appliances or bare-metal installs that condense all functions onto a single system. Other tools may require more complex installations that even involve different operating systems and a separate, dedicated database installation requiring a database license and DBA.

Even when the entire user interface is fully unified, that doesn't mean the infrastructure behind it isn't heterogenous. This shouldn't inhibit your DLP usage, but it's important to keep in mind during the selection and planning.

Central Administration, Policy Management, and Workflow

As we've discussed throughout this report, all current DLP solutions include a central management server for administering enforcement and detection points, creating and administering policies, incident workflow, and reporting. These features are frequently the most influential in the selection process. There are many differences between the various products on the market, so rather than try to cover every possible feature we'll focus on the baseline of functions which are most important.

User Interface

Unlike other security tools, DLP tools are often used by non-technical staff ranging from HR to executive management to corporate legal and business unit heads. As such, the user interface must account for this mix of technical and non-technical staff and be easily customizable to meet the needs of any particular user group. Due to the complexity and volume of information a DLP solution may deal with, the user interface can make or break a DLP product. For example, simply highlighting the portions of an email in violation of a policy when displaying the incident can shave minutes off handling time and avoid misanalyses. A DLP user interface should include the following elements:

- **Dashboard:** A good dashboard will have user-selectable elements and defaults for technical and non-technical users. Individual elements may be available only to authorized users or groups which are typically kept in enterprise directories. The dashboard should focus on the information valuable to that user, and not be just a generic system-wide view. Obvious elements include number and distribution of violations based on severity and channel and other top-level information, to summarize the overall risk to the enterprise.
- **Incident Management Queue:** The incident management queue is the single most important component of the user interface. This is the screen incident handlers use to monitor and manage policy violations. The queue should be concise, customizable, and easy to read at a glance. Due to the importance of this feature we detail recommended functionality later in this report.
- **Single Incident Display:** When a handler digs into an incident, the display should cleanly and concisely summarize the reason for the violation, the user involved, the criticality, the severity (criticality is based on which policy is violated, severity on how much data is involved), related incidents, and all other information needed to make an intelligent decision on incident disposition.
- **System Administration:** Standard system status and administration interface, including user and group administration.
- **Hierarchical Administration:** Status and administration for remote components of the DLP solution, such as enforcement points, remote offices, and endpoints, including comparisons of which rules are active where.
- **Reporting:** A mix of customizable pre-built reports and tools to facilitate *ad hoc* reporting.

- **Policy Creation and Management:** After the incident queue, this is the most important element of the central management server. Policy creation and management is important enough that we'll cover it in more detail later.

A DLP interface should be clean and easy to navigate. That may sound obvious, but we're all far too familiar with poorly designed security tools that rely on the technical skills of the administrator to get around. DLP is used outside of security — possibly even outside of IT — so the user interface needs to work for a wide range of users.

Hierarchical Management, Directory Integration, and Role-Based Administration

Hierarchical Management

DLP policies and enforcement often need to be tailored to the requirements of individual business units or geographic locations. Hierarchical management allows you to establish multiple policy servers throughout the organization, with a hierarchy of administration and policies. For example, a geographic region might have its own policy server slaved to a central policy server. That region can create its own specific policies, ignore central policies with permission, and handle local incidents. Violations would be aggregated on the central server while certain policies are always enforced centrally. The DLP tool would support the creation of global and local policies, assign policies for local or global enforcement, and manage workflow and reporting across locations.

Directory Integration

DLP solutions also integrate with enterprise directories (typically Microsoft Active Directory) so violations can be tied to users, not just IP addresses. This is complicated because they must deal with a mix of managed and unmanaged (guest/temporary) employees without assigned addresses. The integration should tie DHCP leases to users based on their network login, and update automatically to avoid accidentally tying a policy violation to an innocent user. For example, an early version of a current product associated a user to an IP address until the address was reassigned to another user. One reference almost fired an employee because a contractor (not in Active Directory) was the next person to use that IP and committed a policy violation. The tool linked the violation to the innocent employee. Directory integration also streamlines incident management by eliminating the need to reference external data sources for user identities and organization structure.

Role-Based Administration

The system should allow internal role-based administration for both internal administrative tasks and monitoring & enforcement. Internally, users can be assigned to administrative and policy groups for separation of duties. For example, someone might be given the role of enforcing any policy assigned to the accounting group, without access to administer the system, create policies, see violations for any other group, or alter policies. Since your Active Directory might not reflect the user categories needed for monitoring and enforcement, the DLP system should provide flexible support for monitoring and enforcement based on DLP-product specific groups and roles.

Policy Creation and Management

Policy creation and management is a critical function at the heart of DLP — it's also (potentially) the most difficult part of managing DLP. The policy creation interface should be accessible to both technical and non-technical users, although creation of heavily customized policies will nearly always require technical skills.

For policy creation, the system should let you identify the kind of data to protect, a source for the data if appropriate, destinations, which channels to monitor and protect, what actions to take for violations, which users to apply the policy

to, and which handlers and administrators can access the policy and violations. Not all policies are created equal, so each should also be assigned a sensitivity, and severity thresholds based upon volume of violations. Policies should be usable as templates for new policies, and if the system includes categories (which you really want) the policies associated with a given category should also be editable and available as templates for custom policies. Policy wizards are also useful, especially for policies such as protecting a single document.

Most users tend to prefer interfaces that use clear, graphical layouts for policies — preferably with an easy-to-read grid of channels monitored and disposition for violations on that channel. The more complex a policy, the easier it is to create internal discrepancies or accidentally assign the wrong disposition to the wrong channel or violation.

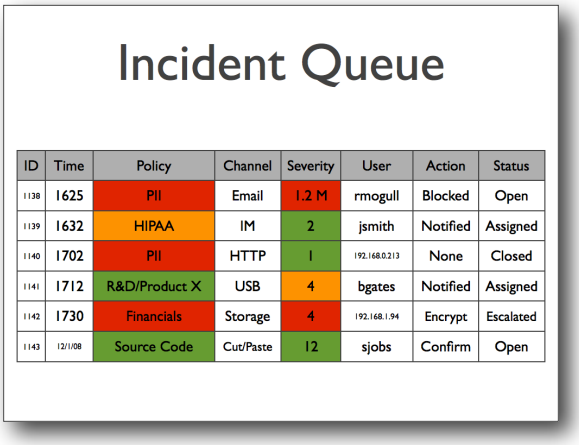
Essentially every policy will need some level of tuning, and a good tool will allow you to create a policy in test mode that shows how it would react in production, without filling incident handlers' queues or taking any enforcement action. Some tools can test draft polices against previously recorded traffic.

Policies include extremely sensitive information, so they should be hashed, encrypted, or otherwise protected within the system. Some business units may have extremely sensitive policies which will need to be protected against system administrators without explicit permission to see a particular policy. All policy violation records should also be protected.

Incident Workflow and Case Management

Incident workflow will be the most heavily used part of the DLP system. This is where violations are reported, incidents managed, and investigations performed.

The first stop is the incident handling queue, which is a summary of all incidents either assigned to that handler, or unassigned but within the enforcement domain of the handler. Incident status should be clearly indicated with color-coded sensitivity (based on the policy violated) and severity (based on the volume of the transgression, or some other factor defined in the policy). Each incident should appear on a single line, and be sortable or filterable on any field. Channel, policy violated, user, incident status (open, closed, assigned, unassigned, investigation) and handler should also be indicated and easily changed for instant disposition. By default, closed incidents shouldn't clutter the interface — basically treating it like an email Inbox. Each user should be able to customize anything to better suit his or her work style. Incidents with either multiple policy violations, or multiple violations of a single policy, should only appear once in the incident queue. An email with 10 attachments shouldn't show up as 10 different incidents, unless each attachment violated a different policy.



When a single incident is opened, it should list all the incident details, including (unless otherwise restricted) highlighting what data in the document or traffic violated which policy. A valuable feature is a summary of other recent violations by that user, and other violations with that data (which could indicate a larger event). The tool should allow the handler to make comments, assign additional handlers, notify management, and upload any supporting documentation.

More advanced tools include case management for detailed tracking of incidents and any supporting documentation, including time-stamps and data hashes. This is valuable in cases where legal action is taken, and evidence in the case management system should be managed to increase its suitability for admission in court.

System Administration, Reporting, and Other Features

As with any security tool, a DLP solution should include all the basic system administration features, including:

- Backup and restore: both full system and system configuration only for platform migrations.
- Import/Export: for policies and violations. There should be some provision for extracting closed violations to free up space.
- Load balancing/clustering
- Performance monitoring and tuning
- Database management

Tools tend to mix these functions between the tool itself and the underlying platform. Some organizations prefer to completely manage the tool internally without requiring the administrator to learn or manage the platform. As much as possible, you should look for a DLP tools that lets you manage everything through the included interface.

Reporting varies widely across solutions; some use internal reporting interfaces while others rely on third party tools such as Crystal Reports. All tools ship with some default reports, but not all allow you to create your own reports. You should look for a mix of technical and non-technical reports, and if compliance is an issue consider tools that bundle compliance reports, which are ubiquitous among dedicated DLP solutions.

When you use storage or endpoint features, you'll need a management interface that allows you to manage policies over servers, storage, and endpoints. The tool should support device grouping, performance and bandwidth management, rolling signature updates, and other features needed to manage large numbers of devices.

Beyond these basic features, products start to differentiate themselves with other advances to help meet particular enterprise needs, including:

- Third party integration, from web gateways to forensics tools.
- Language support, including double-byte character sets for Asia.
- Anonymization of policy violations to support international workplace privacy requirements.
- Full capture for recording all traffic, not just policy violations.

The DLP Selection Process

Define Needs and Prepare Your Organization

Before you start looking at any tools you need to understand why you might need DLP, how you plan on using it, and the business processes around creating policies and managing incidents.

Define the Selection Team

Identify business units which need to be involved and create a selection committee. We tend to include two kinds of business units in the DLP selection process: content owners with sensitive data to protect, and content protectors with responsibility for enforcing controls on the data. Content owners include business units which hold and use the data. Content protectors tend to include departments such as Human Resources, IT Security, Corporate Legal, Compliance, and Risk Management. Once you identify the major stakeholders you'll want to bring them together for the next few steps.

Business Unit	Representative
IT Security	
CIO/IT Operations	
Legal	
Human Resources	
Risk Management	
Compliance	
Networking	
Email	
Storage	
Workstation/Endpoint	
Business Unit/Content Owner	
Business Unit/Content Owner	
Business Unit/Content Owner	

This list covers a superset of the people who tend to be involved with selection. Depending on the size of your organization you may need more or less, and in most cases the primary selection work will be done by 2-3 IT and IT security staff, but we suggest you include this larger group in the initial requirements generation process. The members

of this team will also help obtain sample data/content for content analysis testing, and provide feedback on user interfaces and workflow if they will eventually be users of the product.

Stack rank your data protection priorities and define data types

The first step is to list out which major categories of data/content/information you want to protect. While it's important to be specific enough for planning purposes, it's okay to stay fairly high-level. Definitions such as "PCI data", "engineering plans", and "customer lists" are good. Overly general categories like "corporate sensitive data" and "classified material" are insufficient — too generic, and they cannot be mapped to specific data types. This list must be prioritized; one good way of developing the ranking is to pull the business unit representatives together and force them to sort and agree to the priorities, rather than having someone who isn't directly responsible (such as IT or security) determine the ranking.

For each category of content listed in the first step, define the data type, so you can map it to your content analysis requirements:

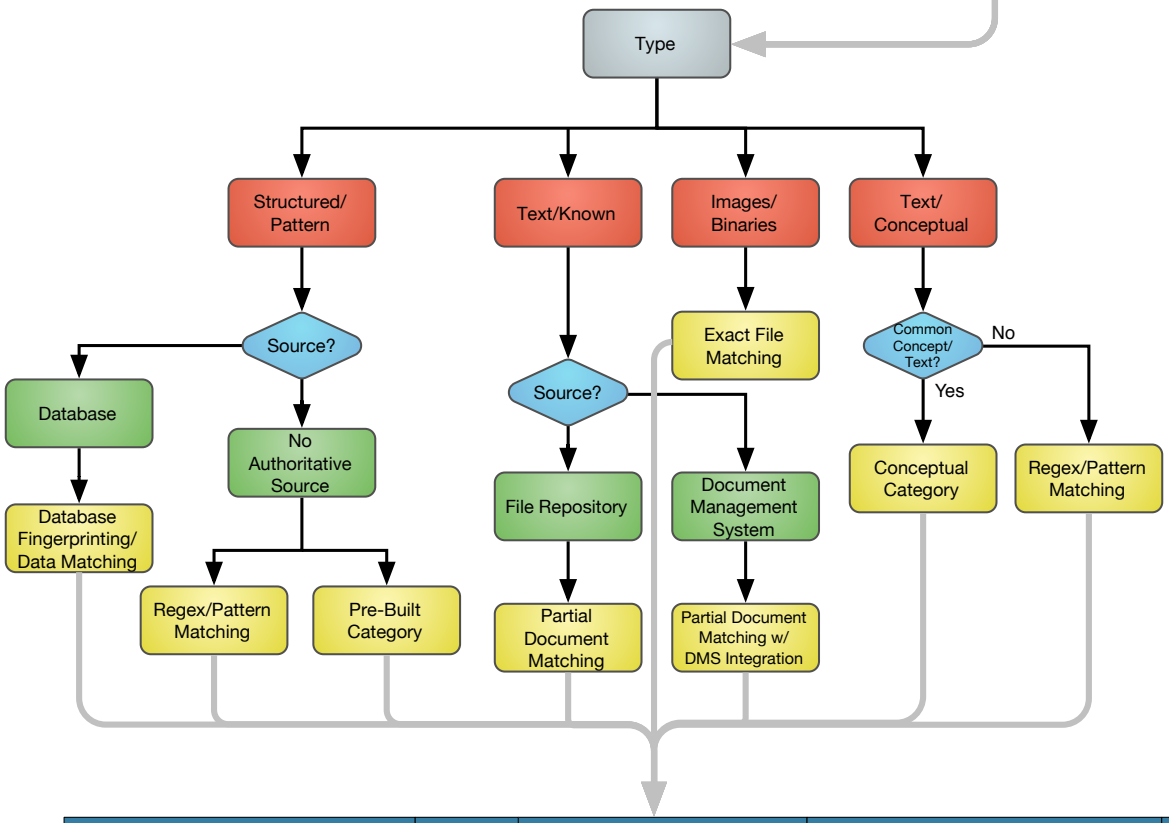
- Structured or patterned data includes credit card numbers, Social Security Numbers, and account numbers — it follows a defined pattern we can test for.
- Known text is unstructured content, typically found in documents, where we know the source and want to protect that specific information. Examples are engineering plans, source code, corporate financials, and customer lists.
- Images and binaries are non-text files such as music, video, photos, and compiled application code.
- Conceptual text is information that doesn't come from an authoritative source like a document repository but may contain certain keywords, phrases, or language patterns. This is pretty broad but examples include insider trading, job seeking, and sexual harassment.

Rank	Data/Content	Type
1		
2		
3		
4		
5		

Match data types to required content analysis techniques

Using the flowchart below, determine required content analysis techniques based on data types and other environmental factors, such as the existence of authoritative sources. This chart doesn't account for every possibility but is a good starting point and should define the high-level requirements for a majority of situations.

Rank	Data/Content	Type
1		
2		
3		
4		
5		



Content Analysis Techinque	Required	Additional Requirements	Requirements/Notes	Phase
Regex/Pattern Matching				
Pre-Built Categories		Categories? (*e.g.*, PCI/HIPAA)		
Database Fingerprinting/Data Matching		DB platform support		
		DB connection type support (ODBC vs. file extract)		
Partial Document Matching		Unusual document types?		
Partial Document Matching (DMS)		Document Management System platform support		
Conceptual Categories		Concept (*e.g.*, insider trading, job seeking)		
Exact File Matching				

Determine additional requirements

Some content analysis technique include additional requirements, such as support for specific database platforms and document management systems. If you are considering database fingerprinting, also determine whether you can work against live data in a production system or will rely on data extracts (periodic database dumps reduce performance overhead on the production system).

Define rollout phases

While we haven't yet defined formal project phases, you should have an idea early on whether each data protection requirement is immediate or something you can roll out later in the project. One reason is that many DLP projects are initiated to address some sort of breach or compliance deficiency relating to only a single data type. This could lead to selecting a product based only on that requirement, which might entail problematic limitations down the road as you expand your deployment to protect other kinds of content.

Determine Monitoring/Alerting Requirements

Start by figuring out where you want to monitor your information: which network channels, storage platforms, and endpoint functions. Your high-level options are:

- Network
  - Email
  - Webmail
  - HTTP/FTP
  - HTTPS
  - IM/Messaging
  - Generic TCP/IP
- Storage
  - File Shares
  - Document Management Systems
  - Databases
- Endpoint
  - Local Storage
  - Portable Storage
  - Network Communications
  - Copy & Paste
  - Print/Fax
  - Screenshots
  - Application Control

You might have some additional requirements, but these are the most common ones we encounter.

Determine Enforcement Requirements

As we've discussed already, most DLP tools offer various enforcement actions, which tend to vary by channel and platform. The most basic enforcement option is 'Block': the activity is stopped when a policy violation is detected. For example, an email will be filtered, a file not transferred to a USB drive, or an HTTP URL will fail. But most products also include other options, such as:

- **Encrypt:** Encrypt the file or email before allowing it to be sent/stored.
- **Quarantine:** Move the email or file into a quarantine queue for approval.

- **Shadow:** Allow a file to be moved to USB storage, but send a protected copy to the DLP server for later analysis.
- **Justify:** Warn the user that this action may violate policy, and require them to enter a business justification to store with the incident alert on the DLP server.
- **Change rights:** Add or modify Digital Rights Management on the file.
- **Change permissions:** Modify file permissions.

Location/Channel	Alert	Enforce
Network		
Email		
HTTP/FTP		
HTTPS		
IM/Messaging		
Generic TCP/IP		
Webmail		
Other		
Other		
Storage		
File Shares		
Document Management Systems		
Database		
Other		
Endpoint		
Local Storage		
Portable Storage		
Network Communications		
Copy & Paste		
Print/Fax		
Screenshots		
Application Control		
Other		

Enforcement Actions	
Encrypt	
Block	
Quarantine	
Justify	

Enforcement Actions	
Encrypt	
Quarantine	
Change Rights	
Change Permissions	

Enforcement Actions	
Encrypt	
Block	
Shadow	

Map Content Analysis Techniques to Monitoring/Protection Requirements

As we have discussed, DLP products vary in which policies they can enforce on which locations, channels, and platforms. Most often we see limitations on the types or size of policies that can be enforced on an endpoint, which change based as the endpoint moves off and back onto the corporate network, because some require communication with the central DLP server.

For the final step in this part of the process, list your content analysis requirements for each monitoring/protection requirement you just defined. These directly translate to the RFP requirements at the core of most DLP projects: what you want to protect, where you need to protect it, and how.

Location/Channel	Content Analysis Techniques

Determine Infrastructure Integration Requirements

To work properly, all DLP tools need some degree of integration with your existing infrastructure. The most common integration points are:

- **Directory servers** to determine users and build user, role, and business unit policies. At minimum, you need to know who to investigate when you receive an alert.
- **DHCP servers** so you can correlate IP addresses to users. You don't need this if all you are looking at is email or endpoints, but for any other network monitoring it's critical.
- **SMTP gateway** — this can be as simple as adding your DLP tool as another hop in the MTA chain, but might be more involved.
- **Perimeter router/firewalls** for passive network monitoring. You need someplace to position the DLP sensor — typically a SPAN or mirror port, as discussed earlier.
- Your **web gateway** will probably integrate with your DLP system if you want to filter web traffic with DLP policies. If you want to monitor SSL traffic (you do!), you'll need to integrate with something capable of serving as a reverse proxy (man in the middle).
- **Storage platforms** to install client software to integrate with your storage repositories, rather than relying purely on remote network/file share scanning.
- **Endpoint platforms** must be compatible to accept the endpoint DLP agent. You may also want to use an existing software distribution tool to deploy it.

For each of these list out not only the platform you plan to integrate with, but any additional requirements — such as bandwidth or other performance requirements.

Infrastructure Component	Platform/Requirement
Network	
Directory Servers	
DHCP Servers	
Perimeter Router/Firewalls	
SMTP Gateway	
Email Server	
Email Encryption System	
Web Gateway	
SSL/TLS Reverse Proxy	
IM/Messaging Gateway	
Other	
Other	
Storage	
File Servers	
Document Management Systems	
SharePoint	
Database Management Systems	
Digital Rights Management	
Other	
Other	
Endpoint	
Operating Systems	
Software Distribution/Update Tool	
Email Client	
Device Control Tool	
Remote Access Client	
DRM Client	
Other	
Other	
General	
SIEM	
Workflow Management	
Other	
Other	

We don't mean to make this overly complex — many DLP deployments only integrate with a few of these infrastructure components, or the functionality is included within the DLP product. Integration might be as simple as plugging a DLP server into a SPAN port, pointing it at your directory server, and adding it into the email MTA chain. But for developing requirements, it's better to over-plan than to miss a crucial piece that blocks expansion later.

Finally, if you plan to deploy any database or document based policies, fill out the storage section of the table. Even if you don't plan to scan your storage repositories, you'll be using them to build partial document matching and database fingerprinting policies.

Determine Management, Workflow, and Reporting Requirements

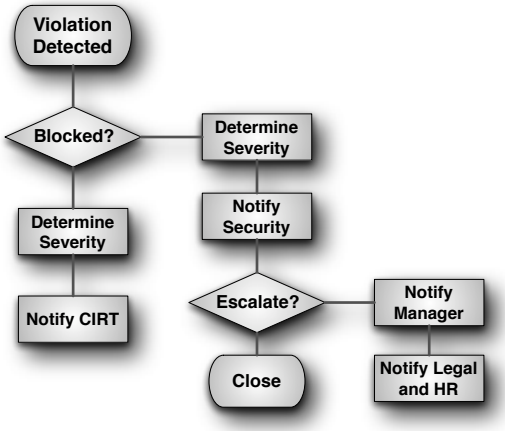
The last major set of requirements includes central management features, workflow, and reporting support. The following table includes common requirements, but is far from exhaustive:

Feature	Requirement
Management	
Consolidated UI for all DLP features	
Incident backup/restore	
Configuration backup/restore	
Policy backup/restore	
Hierarchical management	
Role-based management	
Per-role policy restrictions	
Policy creation wizards	
Regional policy support	
Endpoint agent management and performance tuning	
Storage repository management and performance tuning	
Business unit policy creation	
Automatic incident archiving	
Other	
Other	
Workflow	
Unified incident queue (network, endpoint, storage)	
Role-based incident handling (with regional and business unit support)	
Incident display restrictions (based on policy and sensitivity)	
Incident escalation and transfer	
Incident investigation (other similar violations or incidents by user)	
Internal case management	
Per-handler incident queue display customization	
Email notification	
Non-technical UI support	

Feature	Requirement
Organization dashboard	
Web-based interface	
Other	
Other	
Reporting	
Pre-defined compliance reports	
Additional pre-defined reports	
Internal reporting support	
Third-party reporting support	
Business/executive reports	
Scheduled email report generation	
Other	
Other	
Other	

Outline Process Workflow

One of the biggest stumbling blocks for DLP deployments is failure to prepare the enterprise. In this stage you define your expected workflows for creating new protection policies and handling incidents involving insiders and external attackers. Which business units are allowed to request protection of data? Who is responsible for building the policies? When a policy is violated, what's the workflow to remediate it? When is HR notified? Legal? Who handles day-to-day policy violations? Is it a technical security role, or non-technical, such as a compliance officer? The answers to these kinds of questions will guide you toward different solutions to meet your workflow needs.



By the end of this phase you should have defined key stakeholders; convened a selection team; prioritized the data you want to protect; determined where you want to protect it; listed specific protection, integration, and management requirements; and roughed out workflow requirements for building policies and remediating incidents.

Formalize Requirements

This phase can be performed by a smaller team working under the mandate of the selection committee. Here, the generic needs determined earlier are translated into specific technical features, while any additional requirements are considered. This is the time to come up with any detailed criteria for directory integration, gateway integration, data

storage, hierarchical deployment, endpoint integration, and so on that aren't already specified. You can always refine these requirements after you proceed to the selection process and get a better feel for how the products work.

At the conclusion of this stage you develop a formal RFI (Request For Information) to release to vendors, and a rough RFP (Request For Proposals) that you'll clean up and formally issue in the evaluation phase.

Evaluate Products

As with any products, it's sometimes difficult to cut through the marketing to figure out whether a product really meets your needs. The following steps should minimize your risk and help you feel confident in your final decision:

1. **Issue the RFI:** Larger organizations should issue an RFI though established channels and contact a few leading DLP vendors directly. If you're a smaller organization, start by sending your RFI to a trusted VAR and email a few of the DLP vendors which seem appropriate for your organization.
2. **Perform a paper evaluation:** Before bringing anyone in, match any materials from the vendor or other sources to your RFI and draft RFP. Your goal is to build a short list of 3 products which match your needs. You should also use outside research sources and product comparisons.
3. **Bring in 3 vendors for an on-site presentation and risk assessment:** Nearly every DLP vendor will be happy to come in and install their product on your network in monitoring mode for a few days (if you are big enough), with a suite of basic rules. You'll want to overlap the products as much as possible to directly compare results based on the same traffic over the same time period. This is also your first chance to meet directly with the vendors (or your VAR) and get more specific answers to any questions. Some vendors may (legitimately) desire a formal RFP before dedicating resources to any on-site demonstrations.
4. **Finalize your RFP and issue it to your short list of vendors:** At this point you should completely understand your specific requirements and issue a formal RFP.
5. **Assess RFP responses and begin product testing:** Review the RFP results and drop anyone who doesn't meet any of your minimal requirements (such as directory integration), as opposed to "nice to have" features. Then bring in any remaining products for in-house testing. To properly test products, place them on your network in passive monitor mode and load up some sample rule-sets that represent the kinds of rules you'd deploy in production. This lets you compare products side by side, running equivalent rules, on the same traffic. If testing endpoint or storage support, use a consistent set of endpoints or storage repositories. You'll also want to test any other specific features that are high on your priority list.
6. **Select, negotiate, and buy:** Finish testing, take the results to the full selection committee, and begin negotiating with your top choice.

Internal Testing

In-house testing is the last chance to find problems during your selection process. Make sure you test the products as thoroughly as possible. A few key aspects to test, if you can, are:

- Policy creation and content analysis. Violate policies and try to evade or overwhelm the tool to learn where its limits are.
- Email integration.
- Incident workflow: Review the working interface with those employees who will be responsible for enforcement.
- Directory integration.
- Storage integration on major platforms to test performance and compatibility for data at rest protection.
- Endpoint functionality on your standard image.

- Network performance: Not just bandwidth, but any requirements to integrate the product with your network and tune it. Do you need to pre-filter traffic? Do you need to specify port and protocol combinations?
- Network gateway integration.
- Enforcement actions.

Conclusion

Navigating the Maze

Data Loss Prevention is a confusing market, but by understanding the capabilities of DLP tools and using a structured selection process you can choose an appropriate tool for your requirements.

I've worked with a hundred or more organizations evaluating DLP since the inception of the market. Not all of them bought a product, and not all of them implemented one, but those which did generally found the implementation easier than many other security products. From a technical standpoint, that is — the biggest obstacles to a successful DLP deployment tend to be inappropriate expectations and failing to prepare for the business process and workflow of DLP.

Many of you probably hear that DLP is too complex to deploy, or generates too many false positives. This isn't what I hear from most of the people who have actually deployed or managed DLP, but by the same token I've yet to meet a security tool that's as easy to use and effective as the vendor's sales presentations. When I dig into struggling or failed DLP deployments, I most often find a mix of poorly structured implementations (like turning on a string of checklist policies and being overwhelmed with incidents) or improper expectations (such as high false positives from a DLP Light feature using a very crude regular expression-based rule). I've even talked with organizations who have bought and deployed DLP without bothering to figure out what content they wanted to protect, or who would be responsible for handling incidents.

The key to a successful DLP deployment is knowing your needs, understanding the capabilities of your tool, and properly setting expectations. Know what you want to protect, how you want to protect it, and where you need to integrate with your existing infrastructure before you let the first vendor in the door.

Have a clear understanding of which business units will be involved and how you plan to deal with violations before you begin the selection process. *After* deployment is a bad time to realize that the wrong people see policy violations, or your new purchase isn't capable of protecting the sensitive data of a business unit not included in the selection process.

DLP products provide very high value for organizations which plan properly and understand how to take full advantage of them. Focus on the features which are most important to you as an organization, paying particular attention to policy creation and workflow, and work with key business units early in the process.

Most organizations that deploy DLP properly see significant risk reduction with few false positives and business interruptions.

Selection Worksheet

Process Checklist

Step	Completion Date	Completed By
Define Selection Team		
Stack rank content protection priorities		
Define data types		
Match data types to content analysis techniques		
Determine additional content analysis requirements		
Define rollout phases		
Determine monitoring/alerting requirements		
Determine enforcement requirements		
Map analysis to monitoring/enforcement requirements		
Determine infrastructure integration requirements		
Determine management/workflow/reporting requirements		
Outline process workflow		
Formalize requirements		
Issue RFI		
Perform paper evaluation		
Shortlist vendors and complete vendor presentations		
Issue RFP		
Assess responses		
Select and negotiate		
Issue purchase order		

Define the Selection Team

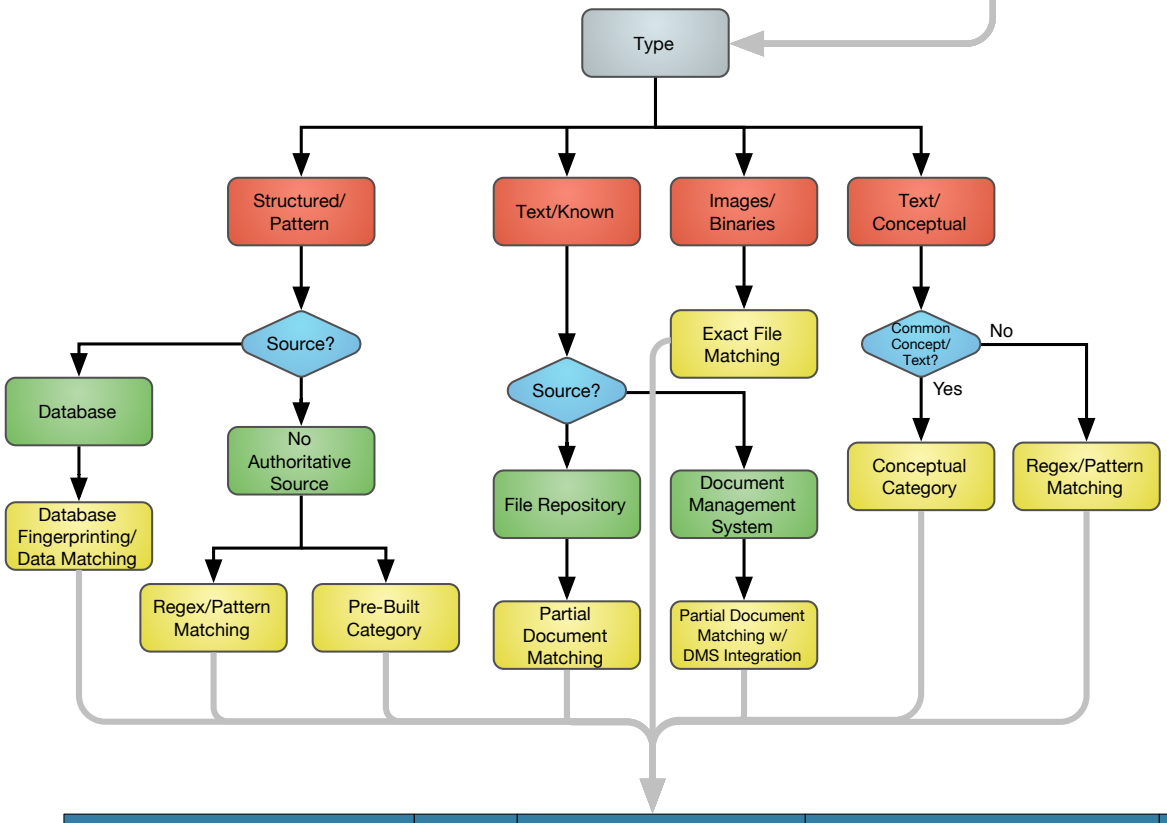
Business Unit	Representative
IT Security	
CIO/IT Operations	
Legal	
Human Resources	
Risk Management	
Compliance	
Networking	
Email	
Storage	
Workstation/Endpoint	
Business Unit/Content Owner	
Business Unit/Content Owner	
Business Unit/Content Owner	

Stack rank your data protection priorities and define data types

Rank	Data/Content	Type
1		
2		
3		
4		
5		

Match data types to required content analysis techniques

Rank	Data/Content	Type
1		
2		
3		
4		
5		



Content Analysis Techinque	Required	Additional Requirements	Requirements/Notes	Phase
Regex/Pattern Matching				
Pre-Built Categories		Categories? (*e.g.*,PCI/HIPAA)		
Database Fingerprinting/Data Matching		DB platform support		
		DB connection type support (ODBC vs. file extract)		
Partial Document Matching		Unusual document types?		
Partial Document Matching (DMS)		Document Management System platform support		
Conceptual Categories		Concept (*e.g.*, insider trading, job seeking)		
Exact File Matching				

Define rollout phases

Phase	Target Requirements/Scope	Target Date
1		
2		
3		
4		
5		

Determine Monitoring/Alerting and Enforcement Requirements

Location/Channel	Alert	Enforce
Network		
Email		
HTTP/FTP		
HTTPS		
IM/Messaging		
Generic TCP/IP		
Webmail		
Other		
Other		
Storage		
File Shares		
Document Management Systems		
Database		
Other		
Endpoint		
Local Storage		
Portable Storage		
Network Communications		
Copy & Paste		
Print/Fax		
Screenshots		
Application Control		
Other		

Enforcement Actions	
Encrypt	
Block	
Quarantine	
Justify	

Enforcement Actions	
Encrypt	
Quarantine	
Change Rights	
Change Permissions	

Enforcement Actions	
Encrypt	
Block	
Shadow	

Map Content Analysis Techniques to Monitoring/Protection Requirements

Location/Channel	Content Analysis Techniques

Determine Infrastructure Integration Requirements

Infrastructure Component	Platform/Requirement
Network	
Directory Servers	
DHCP Servers	
Perimeter Router/Firewalls	
SMTP Gateway	
Email Server	
Email Encryption System	
Web Gateway	
SSL/TLS Reverse Proxy	
IM/Messaging Gateway	
Other	
Other	
Storage	
File Servers	
Document Management Systems	
SharePoint	
Database Management Systems	
Digital Rights Management	
Other	
Other	
Endpoint	
Operating Systems	
Software Distribution/Update Tool	
Email Client	
Device Control Tool	
Remote Access Client	
DRM Client	
Other	
Other	
General	
SIEM	
Workflow Management	
Other	
Other	

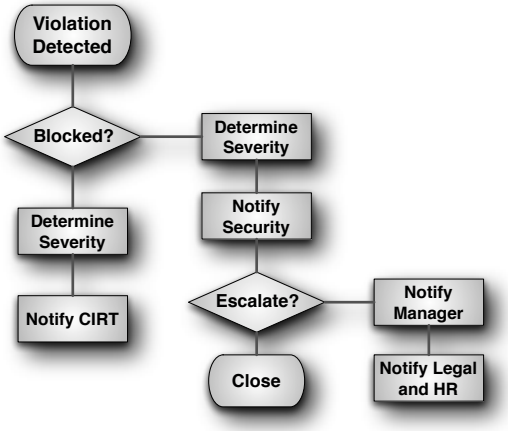
Determine Management, Workflow, and Reporting Requirements

Feature	Requirement
Management	
Consolidated UI for all DLP features	
Incident backup/restore	
Configuration backup/restore	
Policy backup/restore	
Hierarchical management	
Role-based management	
Per-role policy restrictions	
Policy creation wizards	
Regional policy support	
Endpoint agent management and performance tuning	
Storage repository management and performance tuning	
Business unit policy creation	
Automatic incident archiving	
Other	
Other	
Workflow	
Unified incident queue (network, endpoint, storage)	
Role-based incident handling (with regional and business unit support)	
Incident display restrictions (based on policy and sensitivity)	
Incident escalation and transfer	
Incident investigation (other similar violations or incidents by user)	
Internal case management	
Per-handler incident queue display customization	
Email notification	
Non-technical UI support	
Organization dashboard	
Web-based interface	
Other	
Other	
Reporting	

Feature	Requirement
Pre-defined compliance reports	
Additional pre-defined reports	
Internal reporting support	
Third-party reporting support	
Business/executive reports	
Scheduled email report generation	
Other	
Other	
Other	

Outline Process Workflow

This is only a sample:



Attach any additional documentation (RFI/RFP/Vendor Evaluations)

Who We Are

About the Author

Rich Mogull, Analyst and CEO

Rich has twenty years of experience in information security, physical security, and risk management. He specializes in data security, application security, emerging security technologies, and security management. Prior to founding Securosis, Rich was a Research Vice President at Gartner on the security team where he also served as research co-chair for the Gartner Security Summit. Prior to his seven years at Gartner, Rich worked as an independent consultant, web application developer, software development manager at the University of Colorado, and systems and network administrator. Rich is the Security Editor of TidBITS, a monthly columnist for Dark Reading, and a frequent contributor to publications ranging from Information Security Magazine to Macworld. He is a frequent industry speaker at events including the RSA Security Conference and DefCon, and has spoken on every continent except Antarctica (where he's happy to speak for free — assuming travel is covered).

About Securosis

Securosis, L.L.C. is an independent research and analysis firm dedicated to thought leadership, objectivity, and transparency. Our analysts have all held executive level positions and are dedicated to providing high-value, pragmatic advisory services.

We provide services in four main areas:

- Publishing and speaking: Including independent objective white papers, webcasts, and in-person presentations.
- Strategic consulting for end users: Including product selection assistance, technology and architecture strategy, education, security management evaluations, and risk assessments.
- Strategic consulting for vendors: Including market and product analysis and strategy, technology guidance, product evaluations, and merger and acquisition assessments.
- Investor consulting: Technical due diligence including product and market evaluations, available in conjunction with deep product assessments with our research partners.

Our clients range from stealth startups to some of the best known technology vendors and end users. Clients include large financial institutions, institutional investors, mid-sized enterprises, and major security vendors.

Securosis has partnered with security testing labs to provide unique product evaluations that combine in-depth technical analysis with high-level product, architecture, and market analysis.