Interested in learning
more about security?

SANS Institute
InfoSec Reading Room

This paper is from the SANS Institute Reading Room site. Reposting is not permitted without express written permission.

Data Loss Prevention

Copyright SANS Institute
Author Retains Full Rights

AD
Symantec™
Endpoint Protection 12
The next generation of reputation-based security
Download the beta
Symantec.
Confidence in a connected world.

© SANS Institute 2008, Author retains full rights.

Data Loss Prevention

GIAC Gold Certification

Author: Prathaben Kanagasingham

Advisor: John C.A Bambenek

Accepted: August 15th 2008

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

Table of Contents

1. Introduction.....3

2. Deeper Look at DLP Solution.....4

3. Identification of Sensitive Data.....6

3.1 Data in Motion.....8

3.2 Data at Rest.....9

3.3 Data at End Points.....10

4. Choosing a Vendor.....11

4.1 Monitoring vs. Prevention.....11

4.2 Centralized Management.....12

4.3 Backup and Storage Requirements.....12

4.4 Ease of Integration.....13

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

4.5 Market Presence.....13

4.6 Staffing Needs.....14

5. A Feature Rich Solution.....14

6. Leak Prevention.....16

7. Proof of Concept.....18

8. Alternative to vendor solutions.....19

9. Glossary.....21

10. References……23

11. Footnotes……...25

© SANS Institute 2008, Author retains full rights.

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

1. Introduction

Data breach has been one of the biggest fears that organizations face today. Quite a few organizations have been in the news for information disclosure and a popular recent case is that of T.J.Maxx. While DLP is not a panacea to such attacks, it should certainly be in the arsenal of tools to defend against such risks.

The term DLP, which stands for Data Loss Prevention, first hit the market in 2006 and gained some popularity in early part of 2007. Just as we have witnessed the growth of firewalls, intrusion detection systems (IDS) and numerous security products, DLP has already improved considerably and is beginning to influence the security industry. While DLP has been known by several acronyms, in simple terms, it is truly a technology that provides visibility at content level into one's network. While I was researching this product, I came across a white paper by Rich Mogull, where the author had indicated the other acronyms this technology had been known to bear.[1]

Here is an excerpt on the acronyms from his whitepaper.

- Data Loss Prevention/Protection
- Data Leak Prevention/Protection

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

- Information Loss Prevention/Protection
- Information Leak Prevention/Protection
- Extrusion Prevention
- Content Monitoring and Filtering
- Content Monitoring and Protection

2. Deeper Look at DLP Solution

It is nothing new that vendors hype up the market with different levels of offering whether it be service or product. While this varies from vendor to vendor, there are commonly three different levels of this solution and they are known as data at rest, data in-motion and data at end-points.

So how is DLP different from any other security technology? While tools such as firewalls and IDS/IPS look for anything that can pose a threat to an organization, DLP is interested in identifying sensitive data. It looks for content that is critical to an organization.

It would seem as though DLP is a solution whose only purpose is to prevent data breaches from intruders. While it can prevent such data leaks, more often than not this

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

solution is used as a mechanism for discovering broken processes in the normal course of business. Such an example would be the presence of sensitive data on an associate's laptop. Organizations spend abundance of time and money on user awareness training. Hence, one would assume data leaks as a result of actions by an unwitting user should be very minimal. This is not true. We know for a fact that majority of all malware outbreaks companies suffer are due to such user actions. This trend has not changed much even with the ongoing user awareness training.

Though it would appear in writing that policies and procedures are properly followed, only when a detective control is in place, would we see the bigger picture as to how much of it is really in effect. In order to be successful in enforcing a policy, detective control alone is not sufficient. While detective control would provide visibility, preventive control is a necessity to lessen data leaks by both accidentally and intentionally. DLP is a technology that can help us enforce these policies effectively.

Vontu, one of the leading DLP vendors according to Gartner report, estimated the following on data disclosure.

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

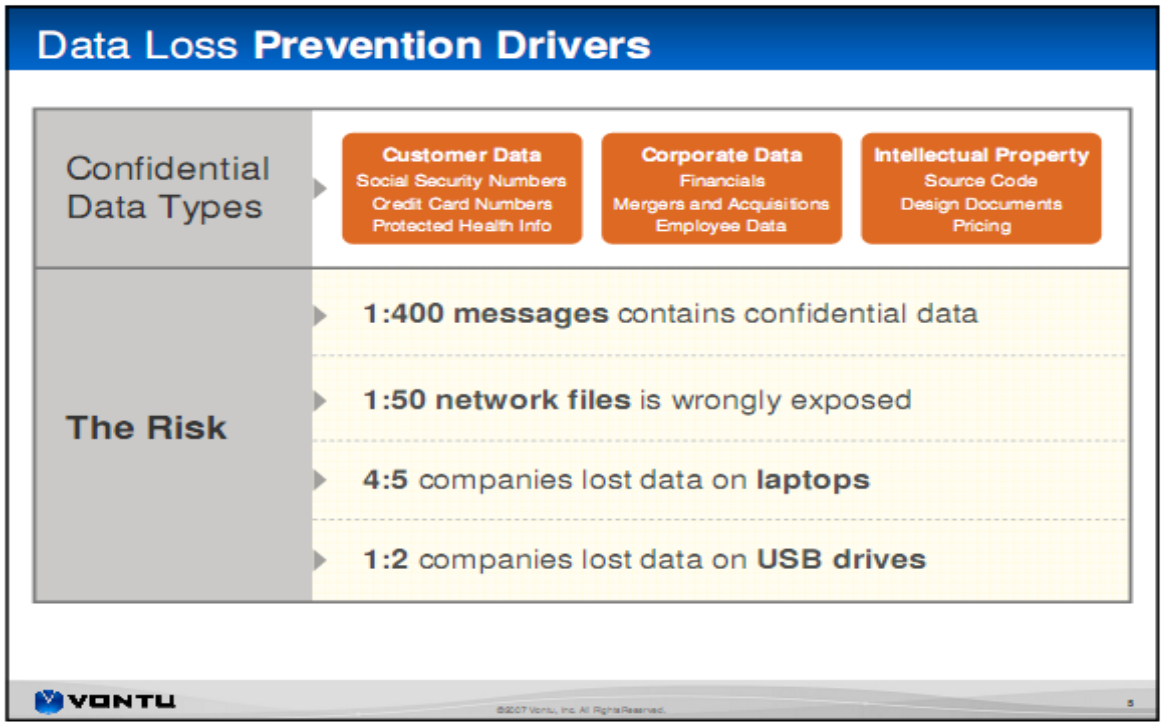


Figure 1.

It is clear from the numbers above, that data loss is largely contributed by employees. This is due to lax security policies on associate use of company assets. Organizations allow the use of Instant messaging, web mail such as Yahoo, Gmail, etc. Some even allow media streaming and P2P file sharing as well. We often find this to be the case with colleges and Universities. Why would organizations allow the use of web mail and instant messaging knowing they run into the risk of losing sensitive data? This has a lot to do with the culture of the organization. Even the ones with very strict policies find it hard to impose such rules

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

effectively. It is not easy to enforce such policies immediately either, when the employees have had the luxury of using them for many years. Governments and high profile organizations are no exception to this rule. For this reason, management is left with the choice of implementing such solution to provide visibility and to pursue the great undertaking of reducing the risk by slowly taking away this luxury from users. In some cases, P2P file sharing has resulted in company proprietary materials landing in the hands of competitors. Such a case on information leak via LimeWire P2P was published on LA times.[2] ZDNET based out of U.K has another story on data loss via USB storage devices.[3] While majority of such violations are non-intentional, they can be very detrimental to an organization. Besides losing company proprietary information, there might be a legal liability on an organization, if the material being shared is copyrighted.

There have been cases of data loss, where employees were part of such act at will. A report by Bnet shows that 45 percent of employees take data when they change jobs.[4] Such is the case with a former HP employee Atul Malhotra, who had allegedly sent copies of IBM confidential documents to his Vice Presidents at HP. Prior to joining HP, he was employed by IBM and had access to this information.[5] How could this incident have been prevented? Proper implementation of DLP would have marked this data as sensitive and rated it a high criticality. Common exit points of this type of data breach are corporate email, web mail, FTP,

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

removable drives and printing. At any of these exit points DLP would have flagged this activity.

3. Identification of Sensitive Data

DLP is shipped with hundreds of pre-defined policies. Portauthority by WebSense boasts over 140 pre-defined templates for major regulatory statutes. These policies have rules for anywhere from identification of social security numbers to US regulatory laws. The very popular ones are HIPAA, Sarbanes Oxley, GLBA, etc. In addition, vendors are even willing to create a custom policy based on customer requirements. This is based on the business model of a particular customer. By closely working with the vendor, default policies can be fine tuned to suit your needs.

Just as we have seen the great use of regular expressions in data mining, they are a useful tool for content matching in DLP as well. This gets even more accurate when data matching is applied against context. If a payroll employee is observed viewing someone else's remuneration package, this event is a normal behavior and can be ignored. However, if this event were to occur from another department, the DLP should raise a flag and hence it should be escalated. One key to point to note in writing a signature is the tradeoff between false positives and false negatives. Some vendors call them wide and narrow rules. If the

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

matching occurs at broader scope, it can result in a high number of false positives. On the other hand, we run into the risk of not catching a true positive, if we were to keep the rules too narrow. This is a business decision an organization should make based on the sensitivity level of the content vs. resources allocated for remediation. After all, customers do not want to end up in a similar situation as they are with IDS.

Almost all vendors I had spoken to use the terms structured data matching and unstructured data matching. Structured data are those that exist in defined formats, such as SSN and credit card numbers. Unstructured data are those that do not conform to a defined format, which is everything else. Some examples of unstructured data are source codes, media files, etc. As far as structured data is concerned, predefined format simplifies the construction of regular expression. As for unstructured date, we are left with no choice but to fingerprint the data due to its complex format. Fingerprints are made using one way secure hash and saved in a database. This information can then be used for identifying such sensitive content elsewhere. Depending on the outcome, a decision will be made whether or not it warrants a reason to raise a flag.

There has been misleading information of DLPs being able to identify 370 plus file formats. File type identification does not translate into content inspection. It is roughly about 180 file types that this technology can interpret and inspect the contents. In order for DLP to

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

do its job effectively, content inspection is important. Hence the identification of a file is no good. Customers tend to get sold on the sheer number of 370, when in fact DLP is equipped to tear down the file on less than half of them.

Accurate regular expressions are important to minimize false positives. This coupled with context will greatly increase true detection capability and save the pain of dealing with false positives. More time can be effectively spent on remediation efforts.

One thing I would like to see added is the granularity in data matching. For example, let's take a look at the following Snort signature.

alert tcp $EXTERNAL_NET any -> $SQL_SERVERS 1433 (msg:"SQL SA brute force"; flow:to_server,established; content:"|10|"; depth:1; content:"|00 00|"; depth:2; offset:34; content:"|00 00 00 00|"; depth:4; offset:64; pcre:"/^.{12}(\x00|\x01)\x00\x00(\x70|\x71)/smi"; byte_jump:2,48,little,from_beginning; content:"s|00|a|00|"; within:4; distance:8; nocase; threshold:type threshold, track by_src)

In addition to matching a strings using the following regular expression *pcre:"/^.{12}(\x00|\x01)\x00\x00(\x70|\x71)/smi",* we are able to define the depth in packet where the match should occur. This is accomplished using the keyword "depth". The same logic can be applied in DLP. Content matching in some directories and files can be either included or excluded as we see fit. Let's take a look at a scenario, where presence of medical

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

terms on a particular server is a concern. Medical term “MUMPS” can be confused with plus five’s MUMPS protocol. In order to avoid this confusion, DLP will need to be equipped with the ability to define inclusion and exclusion criteria for files and directories. It will be even better, if discovery scanner can factor in parent directories to be matched against context. In this case, the term MUMPS found in /etc/mumps/mumps.cfg directory can be ignored as it has no relevance to the medical term.

3.1 Data in Motion

This feature of the DLP solution applies to all data on wire. There are currently multiple protocols supported and HTTP, FTP, IM, P2P and SMTP are to name a few. Please see below for an example of placement of this device.

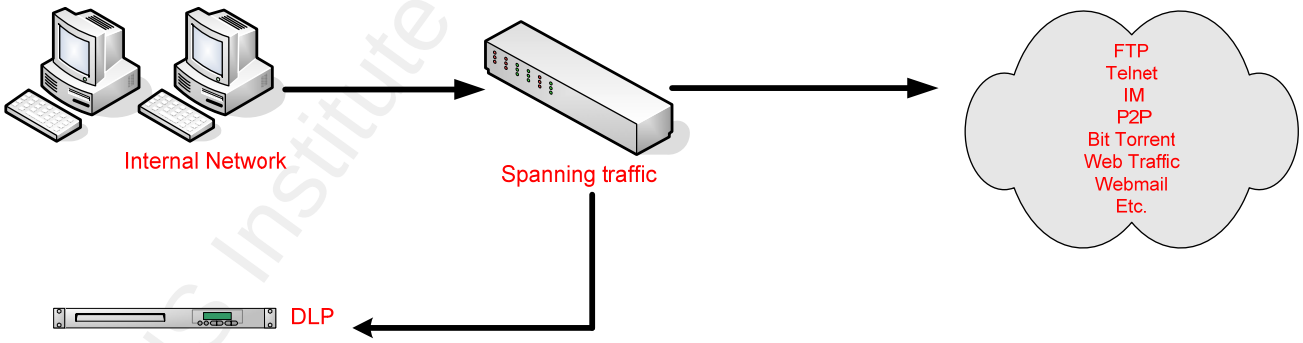


Figure 2.

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

As shown above, all traffic leaving internal network via any of the common channels indicated above will be mirrored to DLP for inspection.

This provides visibility into a large number of violations. For example, if a sensitive file was transferred using FTP, there are several things that this will bring to light. FTP, a protocol that uses clear text should be the first concern in transmitting sensitive files. Secondly, this leads to the question if this file should ever be leaving the company. Third, we will need to verify if the parties involved are authorized to view and transmit data. Most of this applies to not just FTP, but any communication channel mentioned above. Before DLP hit the market, enterprises were already inspecting network traffic for such violations. Email and web activity were very common in this regard. For example, ZIX email encryption technology examines email traffic and encrypts them on behalf of the user, if any critical data is found.

3.2 Data at Rest

Just as the name implies, this applies to anything that holds data such as file shares, databases, etc. Data discovery has two uses. Primary use of this feature is for discovering sensitive data on data depositories. This uses the existing policy to look for any sensitive data. Discovery scanning can also be used to fingerprint data to be used identifying unstructured data elsewhere.

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

Figure 3.

As depicted above, this appliance can be placed anywhere on the network with the only requirement being IP connectivity to targets that are in scope. DLPs are equipped to create multiple virtual sessions to minimize the need for several devices sitting on the network. Each virtual session can be configured to scan a set of servers in a given network. This is ideal for larger networks. While the bandwidth usage can be a concern for such high volume of traffic, there are workarounds to this problem. One of the workarounds is utilizing the incremental scanning feature. Once a server is fully scanned, an incremental scan will only look for changes since the last scan.

The most common data disclosures that are identified during discovery scanning are those where critical data residing on DMZ servers without customer's knowledge. One of the most recent such incidents is a case, where customer data resided on a DMZ server for well over a year. Please see below an excerpt from this story.[6]

"Personal information that may have included Social Security numbers and pharmacy or

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

medical data for about 128,000 WellPoint Inc. customers in several states was exposed online over the past year, the health insurer said Tuesday…exposed information for more than 128,000 customers to Internet access for about a year."

3.3 Data at End Points

Data at end-points is an agent based solution that sits on end user workstations and laptops monitoring for any data leaving via removable devices, such as floppies, CDs, USBs, etc. This also provides auditing and protection against users printing classified data.

Due to its agent based approach, it really has not been a favorable solution among customers. However it does provide a great deal of protection against data leaving via removable devices. Implementation of this solution is comparable to a host-based IDS (Intrusion Detection System).

4. Choosing a Vendor

There are several write-ups factoring in a variety of elements in choosing a vendor. I will point out the key areas that customers should look into. The evaluation criteria that we will be focusing on are the following. After all we are after finding a DLP solution that will meet the business needs as best as possible.

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

1. Monitoring vs prevention

2. Centralized management

3. Backup and storage requirements

4. Ease of integration

5. Market presence

6. Additional staff

4.1 Monitoring vs. Prevention

Vendors use complex and fancy names for these two features. Though taking a deeper look into the solution might reveal a few unique features to each vendor, at a higher level they simply refer to DLP functioning in monitoring mode and prevention mode. For example, Vontu claims one of their key features of “Vontu network protect” is its ability to relocate exposed files. A good analogy to the discussion of whether or not content protection technology should run in monitoring or preventive mode is the comparison between intrusion detection systems (IDS) and intrusion prevention systems (IPS). When IPS was first introduced, there was a misconception of this technology being able to block if not all most of

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

the attacks and the false positives will simply disappear. Little did the customers know at the time, only a handful of signatures could go in block mode out of the box and a thorough study of the environment was critical to extend the blocking capability. Yet, practically there was no significant increase in blocking.

Same rule applies to DLP as well. The accuracy of a signature is very critical before deciding to quarantine or block a certain activity. Moreover, DLP requires additional hardware and software in order to enforce prevention. To quote an example, if we choose to block an email containing sensitive data, some vendors require the integration with an enterprise class MTA, such as Ironport, Sendmail, Proofpoint, etc. Settling for prevention mode can be very costly, especially if you do intend to block multiple channels. In addition to the cost, the ease of integration should be factored in as well. It is important that all future goals be included in the scope. That way if prevention mode is in scope, an organization is better informed of the additional software and hardware requirements that will be needed to enforce blocking effectively. Keep in mind that some of the technologies that we need for blocking might already be in place in your environment. This might take off some of the financial burden.

4.2 Centralized Management

Maintenance overhead is every organization's nightmare. Centralized management

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

can reduce a lot of overhead. Some of the key features to include are policy creation and enforcement, reporting, and data filters.

4.3 Backup and Storage Requirements

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

4.4 Ease of Integration

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

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

deployment will come to a stand still. Often times, this exception will call for a meeting with technology steering board (TSB) and can delay the project significantly.

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

4.5 Market Presence

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

4.6 Staffing Needs

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

When IDS made its first entry into the security industry, very few organizations realized the need for dedicated staff to weed out false positives from actual threat. In present day, almost all organizations that have deployed IDS devices, employ enough staffs to cover a 24/7 operation. Managed security services providers (MSSP) for IDS was built around this concept. Those that could not afford to employ staff for around the clock coverage sought to MSSPs.

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

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

5. A Feature Rich Solution

While the concept of DLP is the same across all vendors, there are some features that are unique to each vendor. I will point out a few that I have found to be the most useful. Some vendors take pride in the ability of their product to be configured with automated remediation. This is certainly an interesting feature, since it takes human element out of the picture. From a financial standpoint, this can vastly reduce the cost involved in remediation. Automated remediation varies based on the type of activity. For example, we may choose to quarantine, encrypt, block and/or notify sender in the case of an email. If not all, most of the functions above can be accomplished using a secure email product.

In the case of data discovery scanning, DLP is equipped to move the data to a secure location, if it were found to be residing on a non-protected share. This is an interesting feature, in that it mitigates the risk by moving the data to a secure location.

Over a decade has gone by since the first commercially available IDS was developed. Yet, not many vendors incorporate active directory as a standard to associate alerts with a particular user/users. The RUA feature (Real-Time User Awareness) in the most recent release of Snort IDS extends this capability. DLP has certainly closed the gap in this regard. It has eliminated the manual process of user lookups by utilizing active directory/LDAP server.

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

This feature is standard among all DLP vendors.

Besides simplifying this process, DLP goes another step further. Let me explain this with an example. If a user was observed making an unauthorized access, DLP will pull up the groups this user is part of. This can benefit in a couple of ways. If the entire group is allowed to access this particular data, there is likely a broken process. If the group is not allowed to access this data, this will indicate this particular user had either special permissions to access the content or was indeed making an unauthorized access, hence it would demand immediate attention. Another key feature is the incident remediation process. Vontu uses the following steps for remediation.

- Severity Level Assignment – Assigns severity level to incidents and is highly configurable.
- Custom Attribute Lookup – This makes queries to LDAP or Active Directory server for user identity and additional attributes.
- Automated Incident Response – A number of actions can be taken using this feature. Some of the important ones are the ability to comment, block, log, etc.
- Role-based Access control – This is an interesting feature, in that it determines

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

which incidents a remediator can work on and the amount of details available. For example, if the violation originated from a staff in the DLP group, it does not do any good assigning the incident to the violator himself.

- SmartResponse – This provides detailed data to determine the remediation steps for incidents. It also allows for fast incident remediation.

Finally, I would like to touch on reporting since this is what speaks for return on investment (ROI) as far as management is concerned. There are different types of report templates shipped with DLP. Majority of the templates cover compliance reporting and executive summaries. However, in addition to basic templates, good customizable reports with the ability to drill down on data are a must for investigative purposes. Ideally, we want this level of granularity on every field in a data set that DLP captures. Some DLPs allow the use of dashboard. This will come in handy for trend analysis.

6. Leak Prevention

Let's take a deeper look at prevention mode. The name of the technology makes direct reference to preventing any data leaks. This is not all it is hyped up to be. Just as what we experienced with IPS, this is really a marketing strategy and there is a huge learning curve. The ability to block sensitive content is a process and is on-going. Vendors themselves

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

recommend that their product be run in monitoring mode for at least six months, before any blocking feature is enabled. I would look at this caution both positively and negatively. If a vendor is so confident of their product, why wait six months? On the other hand, perhaps vendor is recommending a safe approach. Overall, my take is that it requires a deeper understanding of the rules before confidently enabling blocking rules. As pointed out earlier, this is another area that could demand dedicated staff for DLP.

This discussion of preventive mode gets even more interesting with data in motion. This is another area where information can be very misleading. After the first meeting with a vendor, it would seem data in motion has the capability to block multiple channels. For now, this is an idea being entertained and still in the works. There are only a handful of channels that DLP is equipped to block and they are SMTP, HTTP, HTTPS, FTP and Telnet. The effectiveness of this feature remains to be seen.

SMTP blocking capability is enabled by integrating into MTAs. As illustrated in figure 4, all email messages are inspected by an SMTP prevent server and once the content is examined, an action item is sent to the MTA. An action item can be block, encrypt, quarantine and/or notify sender.

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

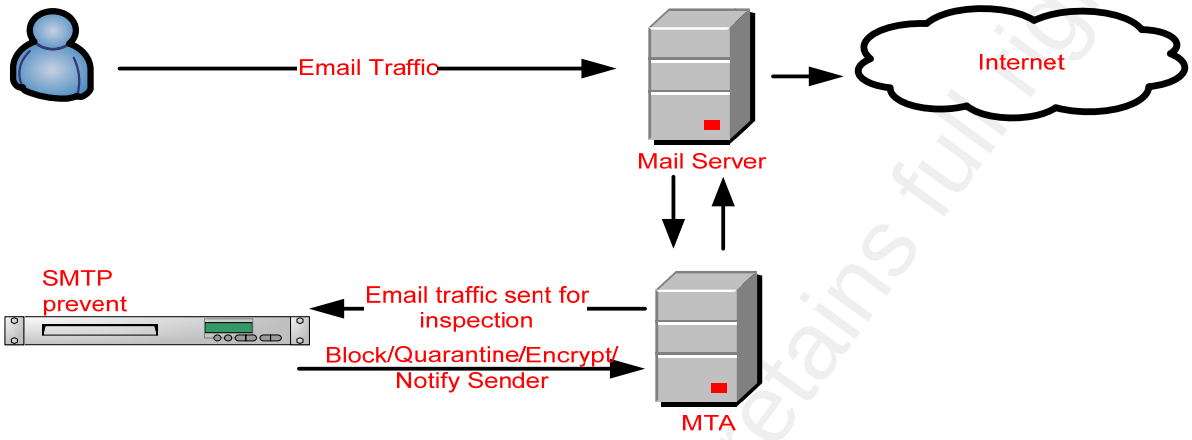
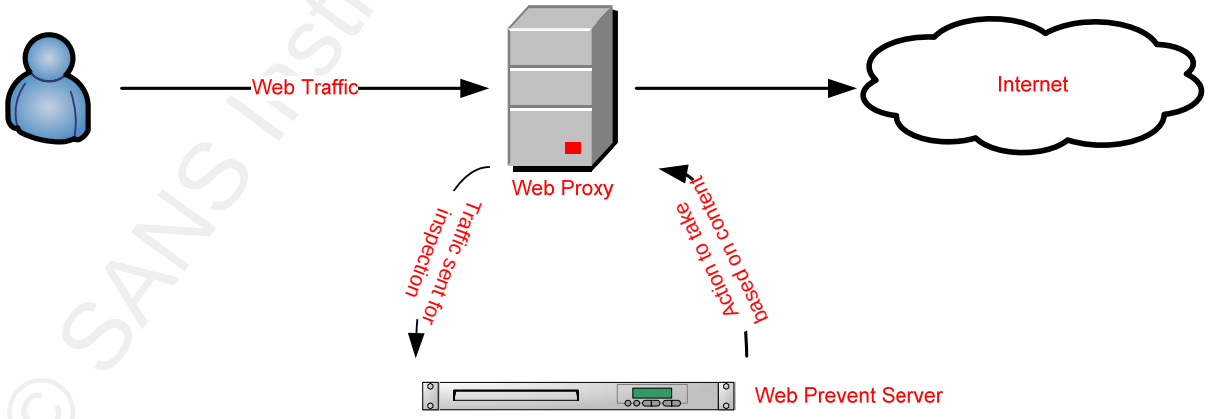


Figure 4.

HTTP and HTTPS blocking are enabled by integrating with HTTP proxy server and HTTPS proxy server respectively. This is very similar to WebSense URL blocking feature, except HTTP prevent servers allow or block request based on content. This feature prevents sensitive information leaving a company via web mail, any external blogging sites, news groups, etc.



© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

Figure 5.

FTP prevention server functions the same way as well. It integrates with FTP proxy servers to enable blocking. There is currently no DLP vendor that can block P2P or IM activity. P2P and IM are growing concerns to large organizations. While there is a challenge in blocking sensitive content alone without disrupting the session, I would like to see this being incorporated into DLP, as it matures over the next year or so.

7. Proof of Concept

Since the technology is still in its early stages, it has been a common trend amongst vendors to demonstrate their solution with a free trail offer. Almost all vendors make this offer to crawl their way into winning customer's business. Some vendors are even willing to take up the challenge for a head to head comparison with their competitors. This is certainly a great offer for serious customers. Vendors usually ship out one test device for proof of concept. I fully recommend making use of this free offer and perform a conclusive test. Everything that was outlined in the scope of the project should be thoroughly tested. Often times, if the test was not conclusive, only after the DLP has been fully implemented, will we become aware of any shortfall of the product. For example, if the data discovery feature was run on a windows based machine, we will not be fully aware of its requirements for another operating platform.

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

This was the case with IBM AIX machines with a particular vendor. Secondly, the bandwidth consumption should be factored into the equation in the case of data at rest. This number gets even higher once multiple such appliances are sitting on the network performing discovery scanning. This can cause latency/disruption to day to day activities. Third, if the test included a handful of servers, we will not be fully aware of the performance hit that centralized management console might endure, once it goes into production. Finally, there is no harm in testing the technology by placing sensitive data in random test machines. This will speak for the quality of the policy. Overall, the trial offer should be treated as though it was in production. This will not only assist in understanding the product better, but discover any shortcomings as well.

Before contacting a vendor, a clear definition of the need for DLP should be in place. Vendors will try and force their way in with their entire suite of DLP solution. I highly recommend taking a phased approach. Most organizations start out with data in motion and move slowly into other solutions, as they understand the technology better. Needless to say, preventive mode should be out of the equation at this stage, for it can disrupt the business. Finally, though it demands much resource and time, I highly recommend that this free offer be made use of with as many vendors as possible before making a final decision.

8. Alternative to vendor solutions

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

Everything we have talked about so far focuses on a vendor solution, which comes with a heavy price tag. Over the years as we have seen alternative free solutions to each and every new technology, that has gone far and beyond what commercial solutions can offer. Such free solutions grow and mature overtime with the contribution by users themselves. A few examples to quote are Snort for IDS and ipchains for firewalls, etc.

Similarly, we can utilize existing free resources to duplicate some of DLP's functionality. For example the primary goal of data discovery feature of DLP is identifying sensitive contents residing on non-secure servers. In essence we are concerned about data leakage occurring via such non-secure servers, such as webservers. This type of data leak can be identified using google alerts as well. Google alert is an automated search solution, which can be setup to automatically notify the concerned parties of interesting data as they appear.[7] It will bring to light any sensitive data that is exposed to the internet. While the basic notification is free, Google does expand this offering from a free trial to platinum service. The fee ranges from $4.95 to $39.95 for this service.

These services have the capability to configure a few items per your preference. Some key configurable items to note are search category, choice of country/language, frequency of notification, filters for certain matches, etc. This includes technical support as well. This very google alert solution can be used to identify any data being posted on forums, blogs, etc. This

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

is what data in motion of the DLP solution accomplishes. For now, this can only function as a detective mechanism. DLP goes a step further and has the capability to block such content before it can be exposed on the internet.

Those who have already invested in content aware technology such as secure email products and websense, might want to consider expanding on them. This in combination with google alerts might accomplish much of what is in scope. This does not guarantee a full fledged DLP solution yet. Since DLP is a fairly new technology, such full fledged open source solutions have not made their way. Perhaps the need for an organization to deploy DLP resides well within what can be accomplished with the tools above. Hence, such huge investment on DLP will be overkill.

One key area, where not much of this can be duplicated today is the overall workflow feature built in a commercial DLP. For large organizations, bits and pieces of DLP like functionality will add a lot of overhead. To avoid duplicate efforts on remediation, all violations reported by secure email products and websense will need to be congregated with google alerts. Myriads of actions might be needed in remediation efforts alone. This will require a lot of documentation. Hence tracking each remediation effort can get very troublesome. With a commercial DLP, much of this can be accomplished with relative ease.

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

9. Glossary

Bit torrent - BitTorrent is an open protocol for sharing large files and filesets.

Data Loss Prevention (DLP) – A security countermeasure used to inspect data on networks for sensitive information. It is capable of running in both monitoring and preventive mode.

Graham Leach Bliley Act (GLBA) - The Gramm-Leach-Bliley Act (GLBA), which is also known as the Financial Services Modernization Act of 1999, provides limited privacy protections against the sale of your private financial information. Additionally, the GLBA codifies protections against pretexting, the practice of obtaining personal information through false pretenses.

Health Information Portability and Accountability Act (HIPAA) - HIPAA, the Health Insurance Portability and Accountability Act, was enacted by the U.S. Congress in 1996, and became effective July 1, 1997. This act is a grouping of regulations that work to combat waste, fraud, and abuse in health care delivery and health insurance.

Lightweight Directory Access Protocol (LDAP) - is an internet protocol used to lookup information.

Personally Identifiable Information (PII) - In information security and privacy, PII is any piece

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

of information which can potentially be used to uniquely identify, contact, or locate a single person.

Point to point (P2P) – is a network of users of sharing files amongst them.

Protected Health Information (PHI) - Under the US Health Insurance Portability and Accountability Act (HIPAA), is any information about health status, provision of health care, or payment for health care that can be linked to an individual. This is interpreted rather broadly and includes any part of a patient's medical record or payment history.

Sarbanes Oxley (SOX) - Sarbanes-Oxley Act (SOX), which targets to prevent misconduct and improve corporate governance practices. It applies to all companies, whose shares are listed on the stock exchanges under the jurisdiction of the U.S. Securities and Exchange Commission (SEC).

Security Information and Event Management (SIEM) – A tool used to manage logs from multiple security devices and to respond to incidents.

Simple Mail Transfer Protocol (SMTP) - Is the de facto standard for email transmissions across the Internet. It is a relatively simple, text-based protocol, in which one or more recipients of a message are specified (and in most cases verified to exist) along with the

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

message text and possibly other encoded objects.

Technology Steering Board (TSB) – is responsible for approving all technical standards and exceptions for an organization.

10. References

[1] Evers J. (2007, January 18). T.J.Maxx Hack Exposes Consumer Data. Cnet. Retrieved

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

from http://news.cnet.com/T.J.-Maxx-hack-exposes-consumer-data/2100-1029_3-6151017.html

[2] Tim W. (2008, April 11). DLP Outlook Hopeful, But No Silver Bullet. DarkReading. Retrieved July 22, 2008, from http://www.darkreading.com/document.asp?Docid =150844

[3] Social Security Numbers of Prominent Figures Leaked over LimeWire Network. (2008, July 9). Los Angeles Times. Retrieved from http://www.latimes.com/technology/la-fi-privacy9-2008jul09,0,3505635.story?track%20=rss

[4] Rick M. (n.d) Understanding and Selecting a Data Loss Prevention Solution. Securosis. Retrieved July 12, 2008, from http://securosis.com/publications/DLP-Whitepaper.pdf

[5] Simple Mail Transfer Protocol. (2008, July 21). Wikipedia. Retrieved July 18, 2008, from http://en.wikipedia.org/wiki/Simple_Mail_Transfer_Protocol

[6] Bit Torrent. (2007, Oct 30). Etree. Retrieved from http://wiki.etree.org/index.php?page=BitTorrent

[7] Personally Identifiable Information. (2008, July 4). Wikipedia. Retrieved July 18, 2008, from http://en.wikipedia.org/wiki/Personally_identifiable_information

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

[8] Protected Health Information. (2008, July 5). Wikipedia. Retrieved July 18, 2008, from http://en.wikipedia.org/wiki/Protected_health_information

[9] The Gramm-leach-Bliley Act. (2005, Jan). Electronic Privacy Information Center. Retrieved July 18, 2008, from http://epic.org/privacy/glba/

[10] What is Sox?. (2006, Mar 10). Metso. Retrieved July 21, 2008, from http://www.metso.com/corporation/home_eng.nsf/FR?ReadForm&ATL=/corporation/articles_eng.nsf/WebWID/WTB-050704-2256F-A1200

[11] What is HIPAA?. (n.d) TechFAQ. Retrieved July 21, 2008, from http://www.tech-faq.com/hipaa.shtml

[12] Proctor P.E, Mogull R. Ouellet E. (2007). Magic Quadrant for Content Monitoring and Filtering and Data Loss Prevention. Gartner. Retrieved from http://www.gartner.com/DisplayDocument?id=503457

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

[13] Raschke T., Penn J. (2008, June 6). The Forrester Wave™: Data Leak Prevention, Q2 2008. Forrester. Retrieved from http://www.forrester.com/Research/Document/Excerpt/0,7211,45542,00.html

© SANS Institute 2008, Author retains full rights.

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

11. Footnotes

[1] http://securosis.com/publications/DLP-Whitepaper.pdf

[2] http://www.latimes.com/technology/la-fi-privacy9-2008jul09,0,3505635.story?track =rss.

[3] http://news.zdnet.co.uk/security/0,1000000189,39448533,00.htm

[4] http://findarticles.com/p/articles/mi_pwwi/is_200705/ai_n19162742

[5] http://www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9109578&source=NLT_SEC&nlid=38

[6] http://www.forbes.com/feeds/ap/2008/04/08/ap4868785.html

[7] http://www.googlealert.com/faqs.php

© SANS Institute 2008, As part of the Information Security Reading Room Author retains full rights.

Last Updated: May 25th, 2013

Upcoming SANS Training

Click Here for a full list of all Upcoming SANS Events by Location

SANS Malaysia @ MCMC 2013	Cyberjaya, MY	Jun 03, 2013 - Jun 08, 2013	Live Event
SANS Pen Test Berlin 2013	Berlin, DE	Jun 03, 2013 - Jun 08, 2013	Live Event
Industial Control Systems Security Training - Houston	Houston, TXUS	Jun 10, 2013 - Jun 15, 2013	Live Event
Security Impact of IPv6 Summit 2013	Washington, DCUS	Jun 14, 2013 - Jun 16, 2013	Live Event
SANSFIRE 2013	Washington, DCUS	Jun 14, 2013 - Jun 22, 2013	Live Event
SANS Canberra 2013	Canberra, AU	Jul 01, 2013 - Jul 13, 2013	Live Event
Digital Forensics & Incident Response Summit 2013	Austin, TXUS	Jul 09, 2013 - Jul 16, 2013	Live Event
SANS London Summer 2013	London, GB	Jul 09, 2013 - Jul 16, 2013	Live Event
SANS Rocky Mountain 2013	Denver, COUS	Jul 14, 2013 - Jul 20, 2013	Live Event
SANS Mumbai 2013	Mumbai, IN	Jul 22, 2013 - Jul 27, 2013	Live Event
SEC528 SANS Training Program for the CompTIA® New Advanced Security Practitioner (CASP) Certification	Washington, DCUS	Jul 22, 2013 - Jul 26, 2013	Live Event
SANS San Francisco 2013	San Francisco, CAUS	Jul 29, 2013 - Aug 03, 2013	Live Event
SANS SEC 560: Network Penetration Testing @ Bangalore 2013	Bangalore, IN	Aug 05, 2013 - Aug 10, 2013	Live Event
SANS Boston 2013	Boston, MAUS	Aug 05, 2013 - Aug 10, 2013	Live Event
Critical Security Controls Summit	Washington, DCUS	Aug 12, 2013 - Aug 18, 2013	Live Event
Industrial Control Systems Security Training - DC	Washington, DCUS	Aug 12, 2013 - Aug 16, 2013	Live Event
SANS Thailand 2013	Bangkok, TH	Aug 19, 2013 - Aug 31, 2013	Live Event
SANS Virginia Beach 2013	Virginia Beach, VAUS	Aug 19, 2013 - Aug 30, 2013	Live Event
Mobile Device Security Summit 2013	OnlineCAUS	May 30, 2013 - Jun 06, 2013	Live Event
SANS OnDemand	Books & MP3s OnlyUS	Anytime	Self Paced