The Next Generation of Network Data Leakage Protection

A Spire Research Report

Executive Summary

For a security professional to thrive in an environment that seems designed to thwart all control mechanisms requires a forward-thinking, aggressive yet thoughtful approach to information control. Over the past few years, network data leakage protection (DLP) has been adopted by many organizations in pursuit of a solution for protecting the sensitive information that saturates every organization. The goal is to encourage usage for all legitimate circumstances while singling out those cases where illegitimate use creates increased risk and likely compromise.

At this stage of advancement, it is worth looking at DLP to understand usage scenarios for leaders and laggards in the information protection arena. This paper reviews these positions in three key areas: network coverage, identification techniques, and response methods.

With leadership positions identified, the paper distills out the key requirements for any DLP program and discusses them in context. These requirements – performance, resilience, accuracy, full coverage, pervasiveness, context, and integration – make up the backbone of the next generation DLP program.

Finally, the paper looks at the future of DLP and what security professionals can look forward to as they build out their real-time information risk management programs.

About Spire Security

Spire Security, LLC conducts market research and analysis of information security issues. Spire provides clarity and practical security advice based on its "Four Disciplines of Security Management," a security reference model that incorporates and relates the functions of identity management, trust management, threat management, and vulnerability management. Spire's objective is to help refine enterprise security strategies by determining the best way to deploy policies, people, process, and platforms in support of an enterprise security management solution.

This white paper was commissioned by Fidelis Security Systems. All content and assertions are the independent work and opinions of Spire Security, reflecting its history of research in security audit, design, and consulting activities.

THE NEXT GENERATION OF NETWORK DLP

Table of Contents

Introduction

The proverbial "headlines in the Wall Street Journal" that make security professionals shudder are popping up these days like pimples on a teenager. Most people remember the incidents at Choicepoint, the Veteran's Administration, Heartland, and T.J. Maxx, but who is paying attention to the Oklahoma Dept of Human Services (1 million records), Chilean Ministry of Education (6 million records), or C-W Group (3.2 million records)? The Open Security Foundation's DataLossdb (www.datalossdb.org) chronicles hundreds of incidents covering millions of identities. Identities are being "stolen" by professional thieves. Bank accounts are being ravaged. Credit cards are being maxed out by others (not just our spouses).

Privacy and identity fraud are prominent social issues and organizations are feeling the heat to protect sensitive personally identifiable information. But privacy is one aspect of a broader confidentiality issue. From an organizational perspective, protection of trade secrets, intellectual property, classified data, and other sensitive information falls into this broader concern. The U.S. Department of Justice Computer Crimes and Intellectual Property Section (http://www.usdoj.gov/criminal/cybercrime/ip.html) catalogs cases such as a conspiracy to steal trade secrets from Goodyear Tire and Rubber Company and an employee at Metadyne's attempt to benefit a Chinese competitor with confidential business information.

Regardless of whether sensitive information is private to individuals or confidential to enterprises, it is clear that leaks and other exposures pose a significant problem to companies. And let's face it – our technical infrastructures are so complex that it isn't hard to lose data throughout the zones and perimeters and networks in any organization.

DLP, however, is not without its weaknesses, and the prudent organization must understand and assess the individual capabilities of solutions to ensure an optimal solution is provided. Initial offerings on DLP quickly demonstrated value and provided key indicators to the future benefits that may be gained.

The Next Generation of Network DLP

For DLP, the network is the likely place to start. While laptops and other physical devices can be stolen in an ad hoc fashion, the network typically manifests the highest level of risk of data loss, especially when factoring targeted, malicious compromises.

Network Data Leakage Protection (DLP) solutions provide a way to scale with the systems that are aggregating and distributing confidential information with maximum throughput. In fact, automated systems are the only hope of the security professional intent on providing identification, detection, and prevention capabilities to an organization.

This section describes where the promise of DLP meets reality. It highlights the path to protection for DLP solutions in three key areas - coverage, identification, and response. For each of these areas, there is a minimum acceptable level of protection for security professionals with high risk tolerance and a master level for those professionals with extremely sensitive data, distributed organizations, and a low risk tolerance.

Coverage

Coverage is a function of completeness. By definition, network DLP looks for leakages at network points, but the level of coverage spans the extent of the network and the breadth of network traffic understood, including both ports and applications.

Minimum Level

To provide the most protection against data leaks, it is common sense to place DLP sensors where the highest risk of leakage exists. Without specific information to the contrary, then, it makes sense to place a solution at the highest volume exit points in a network. An exit point is a trust boundary where trust levels change from one level of trust to another. At the Internet perimeter, the most obvious trust boundary, data leaks occur through email and Web browsing.

Web browsing and email (or more likely, ports 80 and 25) provide necessary but insufficient coverage for data leakage. Not only do Web and email applications often take advantage of the presence of 65,535 different ports, but also data leaks through a myriad of other avenues. Therefore, the next step in maturity for placement of DLP sensors is to look at trust boundaries in a broader way – identifying extranet networks, WAN links, and any other connections where that change in trust levels exists.

Master Level

The trick to getting to this level of maturity is to recognize that the more data is used legitimately in an organization, the higher its value. And in fact the goal of IT is to leverage information as frequently as possible. Coinciding with that usage, however, is the increased risk associated with usage. As value increases, so does risk. It is the security professional's responsibility to provide an appropriate control infrastructure to meet the needs of increased usage.

At the master level, organizations recognize the need for information flow management. This is where DLP involves a change of perspective and not just a technical solution. The notion of "leaks" becomes less black and white and more like the real world – where data is shared and people collaborate and leverage information all the time. At this "gray" level, it is obvious that DLP sensors should be placed across the different trust boundaries on the network – covering all network traffic as it is transmitted between sources and destinations, whether from inside to outside, between divisions, or other internal network control points. It is slightly less

obvious that coverage should include all 65,535 ports and not just the most popular ones.

Identification Techniques

It is safe to say that the most important function of a DLP solution is its ability to properly identify sensitive data. This data may be very structured and specific like credit card numbers or social security numbers or it may be unstructured as in source code or financial data. The first thing the administrator must do is configure a DLP solution to capture and identify sensitive data.

Minimum Level

By far, the most common way to identify sensitive data is using keywords and regular expression patterns. Key words help with data at its most unstructured – the free-form documents describing secret projects and financial highlights and new products that make up the most sensitive information an enterprise owns. Pattern matching using regular expressions is particularly useful for predictable structured data that is often related to privacy – account numbers and personally identifiable information (PII).

After the "low hanging fruit" basics, identification quickly turns into a dark art of mathematical mayhem. The fundamental approach is to "fingerprint" known sensitive documents by taking overlapping micro-slices of the content and hashing them. Various techniques may be applied in the ultimate comparisons of the known hashes to those created on-the-fly.

There are two inherent drawbacks to the fingerprinting approach. First, it is extremely difficult to determine which vendor's technique will be most successful without extensive testing in production and under an assumption that content is going to remain consistent. Second, it requires identification of the sensitive information and documents in advance. While useful, it creates an opportunity for obfuscation and deception.

Master Level

Comparatively speaking, the final level of maturity in identification involves context. Not strictly identification, the context adds value that is necessary for final disposition. The challenge here is to identify irregular behavior amidst the legitimate traffic. Capturing the dates, times, sources, destinations, users, and other information provides an opportunity to perform contextual analysis to benefit the detection process.

The benefits of context are even more apparent with the change of perspective referenced earlier. With a move towards understanding information flow, contextual information demonstrates the patterns of legitimate usage.

Response Methods

The most costly part of the DLP effort involves response. Because the information involved is sensitive, a level of prudence and scrutiny befitting the circumstances is necessary. Couple this with the effort required to evaluate each circumstance, and the costs increase significantly.

Minimum Level

After sensitive content is detected, and its inappropriate use has been ascertained, action is required to minimize the possibility of damage. The basic DLP response is detection and notification – a simple email message, console alert, snmp message, or other notification to the administrator. Notification provides a simple way to assess the efficacy of the control without interrupting processing.

With basic detection and notification, scalability and accuracy often coincide as important factors. Scalability limitations of the solution typically hamper a notification-only response, but low accuracy staunches the willingness to automate. While notification often uses a broad approach to defining inappropriate communications, tuning is required to move to prevention mode. At the prevention level, organizations understand their data flow and must be more certain about the nature of a leakage event.

Focus on Prevention

The move from notification to prevention is the most significant one an organization can make. It requires a level of precision and accuracy before the response can be automated. But the benefits are even clearer.

Prevention focuses participants on the objective of a DLP program – stopping breaches and securing the information. It is one thing to be "compliant" for regulatory purposes, but a different problem when dealing with active compromises seeking to steal data.

Total cost of ownership (TCO) is always a concern with monitoring programs. Having to research every event is expensive. Notification only automates the initiation of an expensive incident response and remediation processes. On the other hand, prevention stops the incident before it occurs and records

Master Level

The benefits of prevention mode are clear – it is a timely response to a serious problem. There are some enterprises leveraging prevention today and this number is growing as security professionals grow comfortable with the quality of the DLP solution. Of course, when dealing with sensitive information, mistakes can be costly.

With DLP, there is a new response method that doesn't require dropping packets or connections like traditional prevention. Because there is a focus on sensitive data, a solution can also employ encryption to provide an extra level of security. So, when a sensitive document is identified, it can be intercepted by the DLP solution and encrypted. Then, a message can be sent to the intended recipient that provides information for retrieving and decrypting the document. This way, legitimate data

that gets caught in a DLP solution can still be recovered, while illegitimate usage will be identified.

Network DLP Requirements

The minimum and master level requirements for the three key areas of coverage, identification, and response point to a set of requirements that will look familiar. We've seen the need for a similar set with intrusion detection solutions. It is worth understanding how they impact network DLP solutions.

Performance

The cardinal sin for any network security solution is to hamper performance in some way. A security device cannot create a bottleneck in the network as packets are processed or they can slow down the network and/or drop packets.

This requirement is crucial as a solution is migrated from detection mode to prevention mode not only because the solution typically becomes an inline "bump on the wire" but also because the business process for managing the devices changes. Processes should be streamlined as confidence in the solution increases.

Resilience

Breaking the network will leave an indelible mark on the security program for years to come. For that reason, it is necessary to have a failover or bypass mode for a security solution in case of performance problems. The sensors and the entire solution must be engineered with the network in mind – to perform at loads that are typical and also to be resilient in the face of a problem. Resilience of the devices in question drives the future capabilities of a data leakage program.

Accuracy

Inline security solutions apply real-time tests to the network traffic. In the case of DLP, the tests look for sensitive data being inappropriately transmitted. There are four potential outcomes to this test – a true positive where information is identified as being sent inappropriately, a true negative where legitimate use is determined, a false positive mistakenly identifies legitimate traffic as inappropriate, and a false negative mistakenly allows inappropriate traffic through.

While the first two requirements covered network-related issues, this is the first requirement that identifies the security requirements. False positives are a bane to productivity and false negatives indicate failure in the security program's mission to protect its sensitive data. Accuracy drives the effectiveness of the program.

Full Coverage

Email and Web are the most common communication means and therefore it makes sense to focus on those channels when beginning a data leakage program. However, remember that this is clear to intruders as well, and they have sixty-five thousand

other ports to choose from. As the DLP program develops, it makes sense to pursue those craftier attackers that will sneak the data out of the enterprise using stealthier (or at least lesser used) programs and protocols.

Pervasiveness

When more programs and protocols are considered, enterprises must assess (or reassess) the location of their DLP sensors. There are often multiple exit points that constitute trust boundaries throughout an IT infrastructure and each one of these should be considered a leakage point as well.

As perspective changes from point solutions looking for leaks to a full-fledged information risk management program, locations become even more important. At this stage, it is important to follow information through data centers, internal work groups, file servers, and other internal locations.

Context

Because the data is at its highest value when it is being used legitimately, there should be a high volume of activity in transmitting it throughout the network. That means that just identifying an account number on the wire does nothing to aid in the determination of whether that flow is legitimate or not.

We can't ascertain what is in the heart and mind of the user, so we define irregularities - circumstances at the outer edges of normal behavior – using whatever contextual information is available. The most common attributes that are evaluated are the source and destination involved in the transmission, since they are always available. These two points might be known as IP addresses, user names, or grouped entities (e.g. ascertained through domain name analysis).

Source and destination are often not enough information to make the call on whether data is being leaked. Therefore, looking at other attributes such as the time of day, programs and protocols in use, and patterns of past behaviors are incorporated for better insight.

Integration

Taking context to the next level, the many sensors deployed throughout the network should inform each other in making decisions. This sharing of information assists in truly identifying anomalies and in identifying patterns throughout an environment. The integration of sensors and management units create a more robust system.

In addition to an integrated DLP solution, this information can integrate with security event management solutions, forensics programs, and encryption/rights management solutions.

The Future of Network DLP

DLP is mature enough to take a step back and evaluate its strengths and weaknesses. The value proposition is clear, but there are still opportunities to fulfill the promise of truly identifying sensitive data being used inappropriately or maliciously.

From negligence to malice

The value of many data leakage programs today revolves around negligent activities of naïve users. These circumstances have no true malicious actor involved, only users that are cutting corners trying to get their work done. But this is only the tip of the iceberg. While it makes sense to keep expectations low when starting a new program, it is crucial to understand the threat associated with intentional compromises.

The future of DLP requires a more assertive approach to identifying malicious activity. It is okay to minimize the expectations of others, but any program leader must recognize the need for higher value protection. The only way to get to this is by honing in on the attributes described herein and leveraging all of the capabilities of a solution.

Understanding information flow

Any program leader thinking strategically recognizes the inherent challenges of a reactive program. Looking for anomalies is by definition reactive when there is no strong understanding of legitimate behavior.

One of the more exciting aspects of data leakage programs is their ability to actually become information flow monitoring solutions. This moves the program from one with tactical advantage to a strategic asset in the company.

The organization that understands how its information is flowing throughout its infrastructure has a much better understanding of its IT-related risk.

Thinking about value and risk

Building on increased knowledge, the DLP program provides historical reference points for identifying those areas that are higher volume with information of higher sensitivity. From this point, frequencies of unwanted use (risk in disguise) can be ascertained and addressed more closely.

Spire Viewpoint

Data, content, information assets – whatever you want to call it – are the crown jewels of information and IT-related security. In most cases, protecting these jewels is the reason we exist. DLP solutions are the last line of defense for keeping this data protected.

The DLP program leader must identify ways to further develop a program that is typically in an early stage of maturity today. It is not enough to just address the problem of negligent users. It is not enough to look at the outer edges in a sea of data. It is not enough to rely on human intervention in a highly-scaled technical environment.

The true value of the DLP program comes when the technical solutions are fully integrated throughout the IT environment and the resources are all integrated into a comprehensive risk management plan. At this strategic level, the enterprise information assets are provided the optimal level of protection and customer trust is preserved.

Contact Spire Security

To comment about this white paper or contact Spire Security, LLC about other security topics, please visit our website at www.spiresecurity.com.

This white paper was commissioned by Fidelis Security Systems. All content and assertions are the independent work and opinions of Spire Security, reflecting its history of research in security audit, design, and consulting activities.