EXTENDING DATA LOSS PREVENTION (DLP) TO THE CLOUD

UNDERSTANDING AVAILABLE OPTIONS AND KEYS TO PROGRAM DESIGN

The Challenge

A 2016 study of cloud usage for more than 30 million employees at more than 600 enterprises worldwide found that the average employee actively uses 36 cloud services at work, including nine collaboration services, six file sharing services and five content sharing services. Additionally, 18.1% of all documents uploaded to the cloud contain sensitive information and 9.3% of documents shared externally contain sensitive content.[1] While IT is generally aware of the cloud-based solutions used by departments such as human resources and accounting for things like employee benefits, payroll and billing, they are often unaware of employees who may be using services such as Box® and Dropbox® to collaborate with internal or external peers. Merger and acquisition teams may be collaborating with potential partners by sharing highly sensitive information, or research teams may be sharing early product design and cost information with external resources in other locations. Often, employees have both personal and business accounts for the same service and could potentially upload a file with intellectual property to a personal repository. The result of any of these actions is increased risk and liability to the organization in case of an incident.

CONTENTS

SANS conducted a survey regarding the types of data being placed in the cloud and found that 40% of respondents affirmed that they process or store sensitive data in the cloud. The most common types of data stored are business intelligence (52%), financial accounting (52%), employee records (48%) and customers' personal information (40%). Thirteen percent of organizations indicated that they do not know whether they even have sensitive data in the cloud.[2]

Organizations continue to embrace the cloud due to the economic benefits it provides. A study by analyst firm Vanson Bourne found that organizations that embrace the cloud grow 19.6% faster than those that do not.[3] The cloud has gone from being a place where developers can spin up a quick server for testing purposes, to now powering and delivering business-critical applications for the Fortune 500.

Cyber espionage netted an estimated $450B in profits in 2015, with more than two billion records lost, and the market is expected to grow to $2 trillion by 2019.[4] In the US alone, 1,093 data breaches were tracked in 2016, an average of almost three a day.[5] The cloud creates an additional exfiltration point for threat actors to exploit and well-meaning employees to accidentally expose data. This paper discusses how organizations can protect their data when it is stored in the cloud through the use of Data Loss Prevention (DLP) and the basic steps required to create a strong data protection program built to support cloud environments.

---

[1]Skyhigh Networks Cloud Adoption and Risk Report, 2016
[2]SANS, Orchestrating Security in the Cloud, A SANS Survey Written by Dave Shackleford September 2015
[3]"The Business Impact of Cloud", Vanson Bourne, 2016
[4]Caleb Barlow, IBM, Ted Talk November, 2016
[5]January 2017, Identity Theft Resource Center (ITRC) and CyberScout (formerly IDT911) Study.

Data Loss Prevention Options

As with everything related to security, there is a trade-off between the ease of doing business and minimizing risk. DLP technologies, once only available through on-premises solutions are expanding to include capabilities to help organizations maintain their productivity while decreasing the risk associated with expanding the IT perimeter to the cloud.

Concerns about protecting assets in the cloud are now paramount, as organizations increasingly have either already embraced the cloud in one form or another – Infrastructure as a Services (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS) – or have plans to do so in the future. A 2016 global study conducted by McAfee indicates that cloud adoption in the enterprise is rapidly approaching a tipping point, with organizations stating that 80% of their IT budget will be taken up by cloud in 16 months or less.[6]

Premises-based, purpose-built DLP solutions provide tremendous value for organizations in protecting the assets that drive their businesses and brands, but have been limited in the ability to address data that traverses to, or is stored in, the cloud. Solutions have emerged to address this gap as the cloud is increasingly being used for business-critical functions.

Manufacturers have also been adding and enhancing functionality to create integrated DLP solutions. Additionally, purpose-built DLP solution vendors have been partnering with or acquiring Cloud Access Security Brokers (CASBs). Examples include Symantec/Blue Coat's purchase of Elastica and Forcepoint's acquisition of Skyfence. The available options for organizations now include:

CLOUD MISCONCEPTION #1 TRANSFER OF RISK

When data is stored in the cloud, the risk associated with the loss or theft of that data remains with the owner. It does not transfer to the cloud provider. Most contracts limit provider liability to a negligible amount with respect to the data voluntarily stored with them. Most of the time, the data organizations are concerned about actually belongs to someone else in the form of customer information such as Personally Identifiable Information (PII) or health information protected by the Health Insurance Portability and Accountability Act (HIPAA), financial information, etc. You cannot reassign that accountability to a third party without each individual customer, as well as the third party, agreeing to that transfer of accountability.

1. Stand-alone CASBs
2. Enterprise-class DLP combined with a CASB
3. Integrated solutions as part of another technology such as firewalls, proxies or SaaS

Protection Through Cloud Access Security Brokers (CASBs)

Over the past few years, CASBs have emerged as the stewards of protecting all services connected to the cloud. A CASB is designed to secure and broker access to cloud services by leveraging points of presence in the cloud that function as gateways. "By 2020, 85% of large enterprises will use a cloud access security broker platform for their cloud services, which is up from less than 5% today."[7] CASB functions include protecting data in applications, detecting malware in cloud uploads/downloads, detecting anomalous behavior that could signify that a user's credentials have been stolen, and monitoring usage of shadow IT applications. CASB solutions such as Elastica (Symantec), Skyfence (Forcepoint), Skyhigh Networks, Netskope, Bitglass and others, continue to enhance their functionality and reach to protect data traversing to/from and resident in the cloud.

A CASB's ability to detect and protect data to/from and within a cloud service is dependent on the methodology of operation employed, with most CASBs leveraging a combination of the following:

- Application Program Interfaces (APIs) – users access the cloud directly, utilizing cloud service APIs to monitor and control user access and enforce policies. Cloud service API licenses are needed with this approach, adding to overall costs.
- Forward proxy – users are routed to a CASB gateway that brokers access to the application.

[6]"Blue Skies Ahead? The state of cloud adoption", McAfee, 2016
[7]Gartner, "Market Guide for Cloud Access Security Brokers", October 24, 2016

- Reverse proxy – user requests go directly to the cloud service, but the cloud service is configured to return all requests to the CASB.
- Hybrid approach of proxy and API modes.

Deployment Option	Pros	Cons
API	Inspects content in sanctioned applications for sensitive content – Data At Rest (DAR) scanning	No Data In Motion (DIM) functionality, no support for unsanctioned applications
	Inspects content in sanctioned applications for malware	Limited to data available via SaaS vendor's API
	Policy enforcement for sanctioned applications – restrict, quarantine, encrypt, alert	Near real-time policy enforcement since it is out of band
Forward Proxy	Inspects content going to/from sanctioned/ unsanctioned apps for sensitive content and malware – DIM	No DAR scanning functionality
	Quarantine malware; block data exfiltration from sanctioned to unsanctioned cloud application	Requires an agent or Proxy Auto Configuration (PAC) file on end user machines to direct traffic to CASB
	Granular access control for managed and unmanaged devices	
Reverse Proxy	Inspects content going to/from sanctioned applications for sensitive content and malware – DIM End user privacy – not scanning unsanctioned application traffic	No DAR scanning functionality
	Quarantines malware; blocks data exfiltration	No unsanctioned application support
	Real time granular enforcement in context of user, device, location, activity, content for sanctioned applications	
	Real time visibility and control of sanctioned application cloud usage by users outside of the company that are using the company's cloud applications to collaborate	
	Doesn't require an agent or PAC file on end user's machine – easy to deploy Good solution for mobile device support since agent on device not needed.	

Table 1 - Cloud Access Security Broker (CASB) Methodologies

CASBs + Enterprise Class DLP

CASBs typically have some form of basic content based DLP available, however, CASBs are increasingly partnering with, or are being acquired by, enterprise DLP providers, strengthening their data detection methods. The end user benefits from having a single vendor to address both premises and in-scope cloud services, with the ability to leverage a single portal for policy configuration, event triage, incident management, and reporting across the hybrid environment. Additionally, the organization can take advantage of best-in-class detection methods available through purpose-built DLP solutions now extended into the cloud.

Integrated Versus Enterprise Class DLP

Many firewall, email and web proxy, endpoint, and SaaS vendors have some level of DLP functionality integrated into their products via an additional license, which can be cost effective and adequate in some cases. Functionality varies considerably by solution and vendor, so it is important to understand the implications of each approach. If an organization is looking for a compliance-based answer, integrated DLP solutions may be an appropriate balance between meeting a "check box" need and an attractive price point. Basic functionality can be met in terms of tracking PII, PHI, credit card numbers, social security numbers (SSNs), and other structured data formats leveraging modifiable vendor-supplied policy templates for ease of setup. Integrated DLP solutions, however, focus primarily on detecting content, not necessarily understanding context or unstructured data.

Enterprise-class DLP solutions provide a risk-based approach that requires greater functionality to be able to address an expansion in the scope and complexity of assets that will need to be protected, such as intellectual property. Understanding an asset's context related to what it is, when it is accessible and how it can move within and outside an organization affects policies and event triage work flows. Additionally, the security program has to evolve as critical assets change, requiring that policies are tuned or created over time. In order to increase the likelihood of effective protection for more complicated and unstructured data, the layered detection methods within enterprise-class DLP will increase in importance, justifying their greater expense.

Some cloud services, such as Microsoft Office 365®, already have some form of DLP that can be licensed, providing a cost-effective means to achieve basic functionality. This may be all an organization needs if compliance is its main driver and it is trying to detect largely structured data such as Social Security Numbers, credit card numbers, and PHI within the confines of the Office 365 application suite. However, if your critical assets are more complex, or you want to have a cohesive set of centrally managed policies across a wide variety of cloud applications, an enterprise-class DLP solution is likely a better choice.

As an example, Office 365's DLP functionality works across Exchange Online, SharePoint® Online and OneDrive® for Business. Content analysis is performed through keyword matches, dictionary matches, text pattern matches through regular expressions and document fingerprinting, which detects sensitive information in standard forms.

CLOUD MISCONCEPTION #2 DATA REGULATIONS DO NOT APPLY TO THE CLOUD

Many people assume that data protection regulations do not apply to the cloud. This is not true. If you are subject to a regulation that mandates you store your information only within the terrestrial borders of Germany, for example, and you contract with an international cloud service or storage provider, they must have an offering that guarantees none of the data leaves Germany for you to remain compliant. This is true even if the primary data centers reside in Germany, but the backups are housed in another country. Most cloud providers have fail over capabilities and redundancy capabilities that make it very difficult for them to make such guarantees.

Enterprise-class DLP systems also leverage those content analysis functions, but layer additional machine learning and optical character recognition functionality into overall detection, leading to better accuracy rates, while also being able to detect unstructured information such as diagrams, source code and formulas. Events can be missed without this level of sophistication and too many false positives can occur, leading to analyst fatigue. Therefore, an understanding of what critical assets you are trying to protect is key to determining the best solution. Leveraging integrated solutions means having to manage multiple portals across vendors for policy configuration and management, as well as event triage and incident management, creating more work for your analyst and engineering staff. The goal is to find a solution that achieves a high degree of accuracy and is not overly burdensome to manage at a price point in line with your budget.

The Need for a Programmatic Approach to Cybersecurity

Organizations need to understand that technology alone will not decrease their risk of data loss whether in the cloud or elsewhere. Many DLP deployments fail due to organizations not properly addressing the following key steps required for successful security programs. Like the security program itself, the outcomes of these steps are not static and should continue to evolve over time.

Step 1: Understand Your Critical Assets

The first step to creating a security program is to understand what you are trying to protect and your business drivers. Are you compliance driven or have you moved to a risk based approach that will also meet your compliance needs? Under what circumstances, and by whom, is the sensitive data in your organization accessible? Who creates this data and where is it allowed to reside?

An organization's most critical assets are those assets that, if lost or otherwise compromised, will cause severe financial and reputational impacts to the business. Once those have been identified, organizations need to understand the Content, Community and Channel of those assets.

- Content refers to the actual attributes of the asset; the kind of information that asset holds. Things like personally identifiable information, product development schema, pricing information, etc.
- Community refers to who should have access to the critical asset, whether internal or external.
- Channel refers to how the critical asset can move into and out of the organization. Can it be uploaded to the cloud and shared with partners, downloaded to a thumb drive, emailed to outside vendors, etc.?

For example, a healthcare employee may be allowed to email his social security number to HR, but is not allowed to email patient social security numbers to anyone either internally or externally. Multiple business units within an organization, such as R&D, HR, legal, finance and others, need to be involved in determining critical assets and their acceptable use. A data discovery session is recommended to understand where an organization's critical assets already reside – on premises or in the cloud - including an understanding of existing shadow IT to help manage risk.

Step 2: Understand the risk that cloud presents to your organization

Conduct a discovery of your organization's cloud usage for visibility into what cloud services are currently being used, by whom, and the associated risks. This information can then be used to define requirements for your future technical solution, as well as help you understand what policies should be configured to protect your data while maintaining user productivity.

Many vendors provide free proof of concept services that leverage firewall and web proxy logs, to identify who is using cloud applications, with the ability to determine a risk score by application. This gives IT an idea of the Shadow IT usage that exists in their environment, so that they can then set the appropriate policies going forward.

Step 3: Identify Governance and Working Groups to drive the program and be responsible for its maintenance

A Governance Group consists of key decision makers needed in order to sign off on policy additions, changes, and escalations. Governance group membership should represent a cross-functional team of multiple business areas. Having a diverse group of individuals helps organizations build and maintain a program that meets the needs of the organization, especially as it continues to evolve.

The Working Group consists of personnel involved in ensuring the successful steady state operation of the technology, overall program and incident management activities. This is a more hands on group, but one that still contains members from various business units or organizational areas.

Step 4: Define Policy Governance

Ensuring the Governance and Working Groups understand their roles and responsibilities is key for success. The Governance Group sets and maintains the strategy and oversight for policy governance with the Working Group involved in the implementation, tuning and triage of policies. Properly defining policies and processes up front alleviates destabilization of a program over time, as well as internal conflicts. Processes need to be defined regarding the addition of new policies and making changes to existing ones in order to evolve the program as business needs change. Expectations need to be set in terms of policy accuracy goals. Additionally, policies should be reviewed at regular intervals to make sure they continue to meet the needs of the organization.

Step 5: Event Triage and Incident Management

The Working Group should have well-defined policies regarding remediation processes in case certain situations such as a breach occur. These policies should include the reporting of any incidents to not only the Governance Group, but also to the appropriate departments within the organization.

Step 6: Define Success Criteria and Reporting Against that Criteria

Many security programs become stagnant due to poorly, or a complete lack of, defined and measurable goals. Executive teams and organizational boards are increasingly seeking more specific answers to how well security programs are performing. By defining success criteria at the onset of a program, or refining criteria for an existing program, the appropriate key performance indicators (KPIs) will be identified and the proper reports for all levels of an organization can be created to highlight the justification of the technology and overall security program.

Conclusion

As cloud adoption continues, it will become increasingly important for organizations of all sizes to adopt some form of data loss prevention. It is critical to assess overall organizational goals, tolerance for risk, staffing, and budget before selecting the right DLP technology. There is no “one size fits all” approach. Consider short term and long term goals in the decision. Engage in proof of concepts with vendors in order to truly understand what the capabilities of a solution are before you commit. The technology landscape is ever changing as new CASBs enter the market and existing CASBs get acquired or forge partnerships with enterprise DLP vendors to leverage the best of both worlds. Other solution vendors (SaaS, FW, proxy) will continue to expand into economical DLP “lite” offerings that target more compliance-oriented organizations. Take into consideration that security technologies you might want to add in the future such as Identity and Access Management (IAM) and encryption are reliant on having a good, solid DLP baseline leveraging content analytics in order to determine what is sensitive. No matter which solution is selected, don’t forget the need for a comprehensive program to complement the technology investment so your organization does not end up on the Identity Theft Resource Center (ITRC) breach report.