The 7 Deadly Sins of Traditional Data Loss Prevention (DLP) in the New World of Shadow IT

Data Loss Prevention (DLP), sometimes referred to as Data Leakage Prevention, encompasses technologies designed to manage and prevent sensitive data exfiltration from an organization. Whether data is considered sensitive or not varies significantly across industries, but can include intellectual property, patent documents, Payment Card Industry (PCI) information, patient Protected Health Information (PHI), source code, financial statements, design documents, and any other data that is critical or valuable to the organization.

Traditional DLP solutions are deployed either within the organization's network (e.g., as an appliance or virtual appliance) or are deployed on the endpoint. In general, they provide coverage for one of three use cases: data in use (e.g., on an endpoint); data at rest (e.g., endpoint or data center); and data in motion (e.g., traversing over the enterprise network).

Unfortunately, these traditional DLP solutions have significant shortcomings when deployed in more modern enterprise environments, which often employ a hybrid of on-premise and SaaS applications. In particular, as organizations move more of their sensitive data onto SaaS providers' servers, traditional DLP solutions are unable to provide sufficient visibility into the SaaS environment, or worse yet, are not able to operate in the "as a service" environments at all.

Let's quickly take a look at some of the most significant challenges of providing DLP for SaaS file sharing applications (e.g., Box, Dropbox, Google Drive, One Drive, Syncplicity, and many others).

Traditional DLP solutions fall victim to 7 deadly sins when applied to SaaS applications:

Sin #1

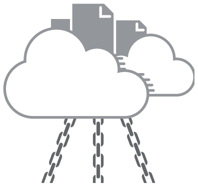
Lacking basic visibility into SaaS applications

The most obvious shortcoming of traditional DLP is that it can only monitor traffic on enterprise-controlled assets (e.g., networks/endpoints). However, traffic to and from a SaaS application might not go over an enterprise network at all. It could be generated, for example, by a mobile user, through a native mobile application, over a mobile network. This trifecta is outside the purview of traditional on-premises enterprise solutions, and as such, is fundamentally beyond the scope of what classic DLP solutions were designed to handle.

Sin #2

Failing to interpret encrypted traffic

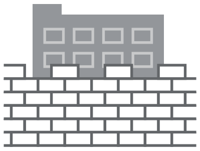
Traffic to and from SaaS applications is typically encrypted (e.g., transmitted over SSL/TLS). Therefore, even if a traditional DLP solution managed to gain network-level visibility into the traffic, it might not be able to interpret the underlying content. Again, without basic visibility, there is little that can be accomplished by traditional solutions. Some traditional DLP solutions have started to address this limitation, but very few are able to adapt to the greater demands of the SaaS model.



Sin #3

Interpreting links versus raw data

Traditional DLP solutions are predicated on processing raw data directly. However, when you think about it, data is never being directly shared in SaaS file sharing applications. Instead, what is being shared is some type of link (e.g., a URL) to the content. The link itself reveals little to no useful information about the content being shared. What must be done, therefore, is to analyze the content being pointed to by the link – which is not something that traditional DLP solutions can do.



Sin #4

Using “perimeter defense” sharing semantics

In the context of traditional enterprise environments, data loss or leakage has a well-defined meaning -- namely the crossing of data across the enterprise perimeter. For SaaS file sharing applications, however, the definition of leakage or loss is fundamentally different for two reasons. First, once data is hosted with a SaaS provider, it already resides outside the enterprise network. Moreover, it can be shared with third parties who are also outside the network. Beyond a certain point, sharing no longer involves data migrating over enterprise assets. Second, data in a SaaS application is shared on a per-user basis. For example, if you want to share a file with someone else, you can typically do so by simply entering that person’s email address or the username they use for the SaaS application. Traditional DLP solutions do not understand these sharing semantics, and cannot assess if data is being “lost” or leaked.

Sin #5

Applying algorithms not designed for file-based data

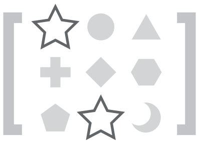
Traditional DLP technologies might make different assumptions regarding the data they have to process. For example, they may assume that data is transmitted in a stream and has to be processed as such. When dealing with SaaS-based file sharing applications, the data model generally involves being able to access entire files containing sensitive data. Algorithms that are designed for streaming data might not perform well on file-based data (and vice versa). As a result, to achieve optimal performance for SaaS-based enterprise file sharing applications, it is important to develop algorithms that were designed to take advantage of full-file content.

Sin #6

Viewing content myopically, while ignoring broader context

Traditional DLP solutions might examine a piece of content in isolation and use that as the sole basis for determining whether or not the transmission of that content represents a violation of a pre-defined policy. For DLP in SaaS applications, we have access to much richer context about a particular file. For example, what type of exposure are we dealing with? Has the file only been shared internally or is it being shared externally? Even worse, is it being shared publicly? With whom is the file being shared? Who originated the sharing of the file? Was it an external party or did it come from the inside?

Considering context is not just important for determining whether a policy is violated, but it is also important when remediating issues. For example, you might be fine with a SaaS application hosting a file containing specific content, as long as the users with whom that file is shared are internal to the organization. If an attempt is made to share the content with an unauthorized third party (either intentionally or due to user error), then you might want to block only this type of access. Because traditional DLP solutions are not privy to the mechanics of SaaS file sharing applications, they are unable to provide enforcement capabilities that are consistent with the way these applications work.



Sin #7

Depending on regular expression and pattern matching

Traditional enterprise DLP technologies rely primarily on basic pattern matching and regular expressions for identifying sensitive content. For example, to identify a credit card number, the DLP solution might look for sixteen digit numbers formatted in the appropriate way that pass the Luhn test. While this approach will be highly sensitive to finding credit card numbers, it will have poor specificity. In particular, there may be many instances of files containing numerous digits that can be misconstrued as credit card numbers. To address this concern, it is important to apply techniques from natural language processing and machine learning. These approaches go beyond simply trying to understand the raw content, and instead focus on being able to understand the underlying context. For example, the presence of a sixteen digit number is itself vague. If, however, in proximity to that number we see what appears to be a name, an address, and a date – then we can have more confidence that we are dealing with sensitive financial information. This example is fairly basic, but it hopefully imparts a flavor for what more complex variations exist.

Along similar lines, imagine that you are trying to see if a file contains source code. In this case, attempting to do basic string and regular expression mapping can easily break down. On the other hand, there are more elaborate methods you can use from areas like machine learning. These methods involve understanding generic statistical patterns, structural attributes, and the like. They can be used to identify that a file contains source code, but without looking for a very specific set of characters.

Data loss prevention in the context of SaaS applications is starkly different from what needs to be done for traditional on-premises enterprise applications. Clearly a new approach to DLP is needed by organizations that are now relying on a hybrid of on-premise and SaaS applications.

To find out how your enterprise can prevent data loss in both on-premise and SaaS environments, visit us at www.elastica.net.