Shadow Data Exposed

Analysis of files shared by leading organizations sheds light on the growing risk to enterprise data.

While it is broadly understood that cloud services are sweeping the IT landscape, the implications of such swift adoption are still relatively unexplored. IT security teams are struggling to keep up and often lack even the basic tools to gain visibility into adoption and use of such services. Of the various categories of cloud technologies available, the majority are Software-as-a-Service (SaaS).

One challenge organizations face is getting a handle on the potential risks of “Shadow IT”, where employees are adopting SaaS apps on their own without the knowledge of the IT organization. Beyond Shadow IT, however, the massive adoption of mainstream file sharing services demand a deeper analysis into the risks of “Shadow Data”, or what type of data and content is being shared, even when this data is residing in IT-approved cloud apps. Given their swift adoption (total data is doubling every 18 months according to IBM[1]), file sharing technologies such as Box, DropBox, Google Drive and Microsoft OneDrive demand particular scrutiny from IT security teams.

At a minimum, IT security teams should seek to understand the nature and magnitude of the risk associated with file sharing in the cloud. Fortunately, extensive real-world data is available from Elastica’s analysis of numerous organizations’ usage of file sharing. An examination of this data can reveal patterns that likely apply to many organizations, including yours.

The Problem: Ignorance

A new source of information that sheds light on file sharing use and associated risk is required because relying on legacy IT security tools (e.g., antivirus, firewalls, or data loss prevention) leaves organizations in a state of ignorance with respect to the magnitude and nature of file sharing use. Such tools either lack integration with the cloud, and thus are largely irrelevant, or lack the ability to understand the nature and context of the data, and thus are ineffective. Or both.

When relying on legacy IT security tools alone, organizations don’t know which cloud file sharing services end-users have elected to adopt on their own. Organizations don’t know what files and data have been shared, broadly shared, or have been outright compromised. Organizations are blind to many inbound threats, and don’t know who (i.e., what user) is doing what. Of course, a lack of visibility precludes any ability

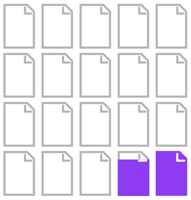
[1] http://www-01.ibm.com/software/data/demystifying-big-data/

to control or remediate the problem. Before launching an investigation of what is occurring on your network, with your data, an analysis of activity uncovered on other real-world enterprise networks may prove valuable.

The Measured Reality – Risks Exposed

Elastica has analyzed over 100 million files it has under management on behalf of the enterprises it serves. These organizations span the spectrum of industries – from those that are highly regulated (e.g., financial services and healthcare) to those that are often targeted due to the inherent value of their data (e.g., retail, high tech). This data has been anonymized and aggregated, and what it reveals can shed light on typical file sharing behaviors, the nature of the data in question, and the possible consequences of such activity for an organization like yours.

To the surprise of many organizations, initial scans of their file sharing applications revealed numerous exposure risks that warranted remediation. This was due to end users' inadvertent or deliberate sharing of sensitive compliance-related data. Furthermore the majority of the exposure risks were concentrated within a relatively small portion of the employee population, enabling targeted remediation strategies.



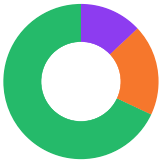
Risk #1

Volume of content in file sharing apps is on the rise

The first set of questions to ask is: how much file sharing is taking place in the organization, and how many files are being shared? Knowledge workers are pushed to do more with less, and have learned through the adoption of consumer mobile and cloud technologies to expect more flexibility in their tools. Mobile devices, remote work, adoption of cloud technologies, and the overall consumerization of IT are trends that have taken root in virtually all workplaces. As a result, different users and groups use different file sharing platforms based on preference and convenience, or as preferred by the outside entities with whom they collaborate. It is no surprise to learn that in most organizations, extensive file sharing is taking place.

The Elastica research reveals that, on average, **2,037 files per user are resident on cloud file sharing solutions**. This is a consequence both of the ease of sharing via such platforms, as well as the convenience and productivity that ensues. Perhaps more interesting is that of these files, an average of **185 files per user are broadly shared** – across the entire organization or with users outside the enterprise. Risk to the organization is impossible to estimate or mitigate without knowing how many files are shared internally, externally, or even worse – made fully public.

Elastica's research found that broadly shared files break down into the following categories:



13% made fully public with no controls whatsoever

19% shared externally

68% shared company wide

This volume of sharing is not in and of itself a problem, but the sheer scope of sharing certainly creates the possibility that among the broadly shared files exist compliance related or sensitive data that demand tighter controls. This leads to the next critical set of questions to examine: what is contained in these files?

Risk #2

Up to 20% of broadly shared files contain compliance-related data

An understanding of the nature of the files shared is required in order to assess whether an organization is at risk. First on the list of most organizations' concerns is compliance with privacy regulations, in particular those associated with PII (personally identifiable information), PHI (protected health information), or PCI (Payment Card Industry). In broad strokes, these regulations require that organizations safeguard the privacy of the individually identifiable information they hold.

Even when organizations pass an audit, it is quite likely that their compliance program is only assessing the known or approved usage of regulated information and is failing to examine whether regulated data is resident in file sharing applications. Beyond simply passing the checks associated with authorized IT systems, organizations have a requirement to ensure that regulated data doesn't leave those systems and enter the unmanaged realm of Shadow IT. More importantly, they must prevent such information from being broadly shared via such systems with users or entities for whom such access is not authorized.

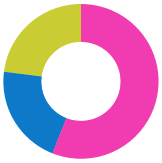
PII is of concern no matter where it resides, because organizations generally face breach or disclosure notification requirements if any such information under their care is inappropriately accessed. PCI data is regulated by the Payment Card Industry Data Security Standard (PCI DSS), and requires that cardholder information be protected both in transit and at rest. In fact, to a large degree PCI compliance entails strictly limiting and delineating the systems on which such data resides, or through which it passes, and then to establish adequate controls around such systems.

PHI relevant data is regulated by HIPPA, and similarly must be carefully managed. While PHI is also intended to ensure the portability of health information as well as its privacy, this portability is through approved mechanisms only, and certainly not intended to lead to broadly shared information with limited access controls.

The Elastica research shows that as **much as 20% of broadly shared files contain compliance-relevant data**. It is worth noting that this is the level uncovered across organizations upon initial scans (i.e., before any remediation has occurred). After remediation, many of these files were removed from file sharing, or had their permissions significantly adjusted. Without such visibility into what is shared, however, organizations cannot satisfy compliance requirements.

Organizations must determine whether compliance-related data is in fact exposed via cloud file sharing, and Elastica's research sheds light on this critical issue.

The research reveals that upon initial scan:



56% of compliance relevant files contained PII (e.g., Social Security numbers)

29% included protected health information (PHI)

15% included PCI-related content (i.e., credit card information)

The compromise of such compliance-related data could trigger expensive and brand-damaging notification requirements. An understanding of the content of the files, and categorization by compliance category, is required to assess risk of non-compliance and institute appropriate policies and controls to prevent inappropriate exposure for any organization subject to compliance regulations or breach disclosure notification mandates.

Risk #3

Sensitive and valuable data is often at risk

Beyond information relevant to compliance with privacy regulations, organizations often are custodians for sensitive and valuable internal and partner information that should not fall into the wrong hands. As such, broadly shared files can expose not just compliance-relevant data, but can also result in exposure of sensitive data critical to the organization such as source code and sales figures.

Are engineering files over-exposed internally, made accessible to temp workers, or other departments unnecessarily? Is source code being posted to GitHub or similar platforms with insufficient controls over access? Are marketing documents or sales projections exposed to possible access by competitors? It is important to ensure that the fluidity of collaboration and access that enhances user productivity doesn't prove to be a double-edged sword that similarly empowers competitors or outsiders to gain unfair advantage. This requires extending the ability to monitor and control file sharing apps beyond compliance related-data, to include detecting a wide range of sensitive data types such as source code, legal files and financial information.

Risk #4

Inbound sharing can create liability and risk for your organization

Just as your compliance-related data can be broadly shared, including with outsiders inappropriately, it is equally possible that your organization is unknowingly receiving and storing sensitive or risky data from a third party. Such inbound sharing can occur incredibly easily, with no notice to you or confirmation on your part.

With a click of a button, a partner could share the files in their instance of a file sharing app with you on yours. This data won't pass through your perimeter per se, but could over time be downloaded or synced to a local cache and end up stored on your systems. Has customer support received live customer data, including PII, PHI, or PCI during the process of troubleshooting an issue? What kind of liability does this create? Could malware circumvent traditional protections and enter through cloud apps?

The convenience and power of file sharing can as easily provide your organization with access to gigabytes of data you would rather not be responsible for as it can expose your data to others. Remaining in the dark about such activity benefits no one.

Risk #5

The worst offenders are often concentrated to just a few

After gaining an understanding of the kinds of information that is flowing outbound and inbound, organizations should seek to understand who is engaging in file sharing activities. User training, and even disciplinary action in the case of egregious or malicious activity, needs to play a role in any security program. Regaining control of cloud file sharing services is no different. Most of all, it is important to understand how to best apply such efforts efficiently, rather than wasting time on user training for departments that have little or no file sharing activity.

As is often the case, the analysis reveals the Pareto Principle (i.e., the 80/20 rule) is in effect. In fact, the Elastica research reveals an even more concentrated example than usual, with **just 5% of users responsible for 85% of the total risk exposure**. Identifying these over-active few is very powerful. In the case of malicious activity, the bad actors can be quickly identified and terminated. In the more common case of inadvertent exposure, the ability to concentrate education and remediation efforts on a small subset of the total users, but affect such a large set of shared data, proves a very efficient use of resources.

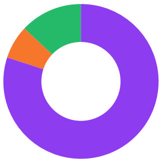
Risk #6

Passwords and encryption are not enough

Many organizations feel that the use of basic transport encryption protocols such as SSL, in combination with strong password requirements, is sufficient to ensure protection of information and systems. Unfortunately, this is often not the case. Whether accidental or deliberate, data exposure often occurs in ways that circumvent the protections that passwords and encryption afford.

In the case of hacker activity, this often entails defeating the password recovery mechanisms if not the passwords themselves, or obtaining password with phishing attacks. Hackers can also circumvent password and encryption completely through installed malware that hijacks legitimate sessions. And of course insiders already know the passwords and have the access they need to cause damage – whether they mean to or not.

Typical industry breakdowns of breaches indicate that[2]:



80% of exposures are inadvertent

7% are the result of disgruntled employee action

12% are the result of accounts having been taken over by hackers or other external threat actors

Unfortunately, cloud file sharing services are just as available worldwide to hacker activity as they are to legitimate use. While most vendors do a good job securing their infrastructure, passwords remain a weak link. As the recent high-profile celebrity iCloud account compromises have shown, that link is often exploited. How can your organization address threats if it has no visibility into activity in cloud file sharing services and no ability to detect suspicious patterns of behavior? If almost 20% of risky activity is the result of malicious internal or external users, organization must pursue a course of gaining visibility and control over such platforms in order to counter that significant threat.

Risk #7

Efficient remediation can save days of effort per user

Visibility is important, but ultimately the ability to take action to reduce detected risk is critical. After understanding what information is being shared, how it is being shared and who is involved, organizations will need to understand whether remediation will be required – and whether the scale of file sharing activity demands automated tools or whether manual efforts will prove adequate.

[2] Aberdeen Study 2014. http://www.bitpipe.com/detail/RES/1397504291_388.html

Elastica's research found that the average time to remediate sharing violations was about 67 minutes per user, however leveraging automation this time dropped to a mere 16 seconds. For a single user, this may not seem significant, but for any reasonable sized organization – the cost savings of automated remediation is significant. Based on these averages, an 8,000-person organization could spend **over a year trying to manually remediate all of its sharing violations**. In addition, the landscape is never static, as users are always uploading new content, so the manual approach quickly becomes untenable even in small organizations.

Perhaps more important than the overall time to remediate is the lag time in identifying the sharing violations in the first place – which can unknowingly expose companies to risk. Clearly automated tools are needed to both identify AND remediate risky exposures if file sharing applications are to be broadly adopted as mainstream tools for enterprises.

Legacy Security Technologies

How are such results possible given the investment over the years in IT security tools and education? The answer is quite simple: new users and old tools. Users have entered the workforce accustomed to the convenience of always-on mobile devices and easy-to-adopt cloud services. Having made extensive usage of such services in their personal lives, they cannot imagine forgoing such tools at work.

Meanwhile, the legacy security solutions of antivirus, network defenses, and data loss prevention are simply not designed to address data stored outside the enterprise boundary. This new cloud model, where organizations are "renting" the network and "renting" the application, has had a profound impact on traditional security approaches, requiring new tools to gain the necessary visibility and controls to which enterprise organizations are accustomed. Data loss prevention technologies, with their combination of network and endpoint based inspection, perhaps come the closest to addressing the concerns around file sharing. But, a significant amount of traffic to and from cloud file sharing services may never traverse the enterprise network, or reach enterprise-controlled endpoints. This, in combination with a number of other shortcomings (see "7 Deadly Sins of Traditional Data Loss Prevention in the New World of Shadow IT" white paper for more information) shows that DLP, too, is an insufficient tool when applied to cloud file sharing services. The result is that most organizations exist in a state of ignorance regarding the magnitude and nature of risk posed by the use of file sharing.

Consequence – Intolerable

The consequence of persisting in this state of ignorance for most organizations is simply intolerable. While some will attempt to limit or prevent such file sharing behavior through a combination of written acceptable use policies (easily ignored) or perimeter blocking rules (easily circumvented), the only robust path forward is through an understanding of current user behavior with respect to file sharing and adoption of tools that can continuously monitor and manage associated risk.

Elastica's research data from numerous real-world organizations helps highlight what is typical: a large number of broadly shared files, 20% of which contain compliance-relevant data. Given this, it behooves most organizations to understand the specifics of their environment. The alternative, persisting in ignorance while resisting the technologies that users require to remain effective and productive, will simply alienate the user base and cause a growing, but unknown, exposure to the risk of data loss. The better option is for IT security organizations to adopt technologies that inform which users are using cloud file sharing services today, what data they are sharing, the sensitivity and compliance implications of that data, and of course, with whom that data is being shared.

Architecture of a Solution

A solution meant to provide visibility and control of cloud file sharing, and ultimately remediation in cases of broad and possibly inappropriate sharing, must be deeply embedded in the cloud rather than mired in legacy perimeter and on-premises tools. This ensures it has complete visibility into file sharing activity while minimizing latency and maximizing performance. Further, the solution should support multiple control points. It must monitor files as they are uploaded to cloud services. In addition, via integration with cloud file sharing services themselves, the solution should scan cloud-stored files and folders and offer direct manipulation of associated file sharing controls and policies for remediation. Finally, such a solution must be formulated to take advantage of the latest breakthroughs in data science and machine learning in order to scale to the vast array of cloud file sharing services on offer while unlocking a robust understanding of what information actually resides in these files, along with the context of the sharing taking place. Only with such a combination can a solution hope to provide visibility across all the major forms of sharing that users undertake, and ultimately appropriate controls and remediation as well.

Conclusion

As we all know, the cloud services train has left the station. Users are onboard and corporate data is along for the ride. The only true choice IT organizations face is whether to alienate end-users by attempting to prevent this productive activity, allow risk to grow unknown and unbounded, or to seek out new solutions that are appropriately architected to address this growing risk. An exhaustive analysis of the real-world cases of data resident in file sharing applications and broadly shared by enterprise organizations reveals that vast quantities of files, containing compliance-relevant and sensitive information, are exposed. A combination of inadvertent and malicious behavior has already put the organization at risk, and that risk is only growing. The Elastica research and analysis of this "Shadow Data" demonstrates the reality of this growing risk across a number of organizations, and is representative of common patterns of behavior likely occurring in your environment as well. Ignoring this reality is dangerous.